An investigation into table grape risk factors that affect quality along the export supply chain

By

Christopher Guillaume Rossouw

Thesis presented in fulfilment of the requirements for the degree of Master of Commerce (Logistics Management) in the Faculty of Economics and Management Sciences at Stellenbosch University

Supervisor: Prof L.L. Goedhals-Gerber, Stellenbosch University, South Africa

Co-supervisor: Prof T. Mantadelis, Ledra College, Cyprus

December 2022

Plagiarism Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

X

Signature

Christopher Guillaume Rossouw

Name in Full

Date

Acknowledgements

I would like to thank the following people for their contributions and involvement toward my completion of this Masters thesis:

Firstly, to **my mom**, thank you for the support and words of encouragement along this journey. Thank you, **dad**, for your technical advice on all things table grape related, the many conversations over the course of the study fuelled the various avenues of enquiry. Thank you to **my siblings**, especially my brother who gave me advice on the more technical aspects of the machine learning, pointing me in the right direction.

To my supervisors **Prof Leila Goedhals-Gerber** and **Prof Theo Mantadelis**, thank you for this journey and your time, I think we all learnt a lot and I really enjoyed the time spent working with you both.

Thank you to **Company X** and **Werner Naudé** for granting me access to the data, without which this study would not have been possible. Thank you to **my colleagues** at work for covering for me when I had thesis related obligations. Your patience is much appreciated.

Lastly, thank you **Jessica** for your daily love, support, and reassurance which helped me see this project through.

“The sun, with all those planets revolving around it and dependant on it, can still ripen a bunch of grapes as if it had nothing else in the universe to do.” – Galileo Galilei

Abstract

Table grapes are a highly perishable product, where a large proportion of grapes produced for export to Europe arrive in a substandard condition. Fruit in this condition requires repacking to remove the rotten food parts, or in extreme cases, the entire shipment is dumped resulting in a total loss. Both outcomes' result in a potential loss of income for stakeholders and food waste, which could be avoided if proper upstream intervention had been taken. This ongoing occurrence prompted the investigation into what the factors are that cause the poor arrival quality of table grapes. The study also applied machine learning techniques to predict the probable arrival scores (green, amber, and red) based on input variables gathered throughout the supply chain.

The data analysed was obtained from five diverse secondary sources consisting of intake quality shed reports, arrival quality reports, logistical nominal data, recorder temperature data, and climate data. The eventual dataset consisted of 467 observations. The analysis process applied consisted of descriptive and inferential statistics to explain the relationship between the upstream variables and the downstream resultant quality scores as well as how the upstream variables interact with one another. The results from the preliminary analysis aided in feature selection for the model building process.

Four classification models, consisting of Logistic Regression, k-Nearest Neighbours, Decision Trees, and Random Forests (RF), were trained, and evaluated. The RF classifier demonstrated the best cross-validation score on the training data and was retained for further evaluation. The RF classifier's accuracy score was 0.63 for the unseen test set and performed best when predicting red class-labels but struggled on green and performed worst for amber class predictions. Variables that had the largest impact on the arrival quality scores consisted of the climactic variables two weeks prior to harvest, the specific variety and ˚Brix at harvest, the number of decayed berries found in the packhouse as well as the overall packhouse quality score, and the type of packaging used (either punnets or loose pack). The effect of the supply chain was also evaluated but did not have any effect for the 2020 season. The attributes of poor quality were also identified in relation to the most important variables, so that upstream mitigation strategies could be determined to reduce financial claims and food waste.

The potential upside of accurate arrival quality predictions prior to shipping would allow for improved allocation decisions leading to profit maximisation through loss reduction and cost savings. From an environmental perspective, assured sound arrival quality would reduce end of chain food waste and would increase product shelf life for consumers.

Keywords:

Botrytis, Cold chain, fresh fruit exports, fruit quality, machine learning, South Africa.

Opsomming

Tafeldruiwe is ‘n hoogs bederfbare produk en hierdie karaktereienskap het tot gevolg dat ‘n groot hoeveelheid tafeldruiwe wat vir die uitvoermark geproduseer word in ‘n onvoldoende toestand in Europa aankom. Druiwe in dié toestand moet gewoonlik herverpak word om die vrot gedeeltes te verwyder en in sommige gevalle moet die hele besending weggegooi word. In beide gevalle is daar ‘n verlies aan inkomste vir belanghebbendes sowel as voedsel vermorsing wat vermy kon word indien daar voldoende stroomop intervensies toegepas is. Hierdie voortdurende gebeurtenis het die ondersoek aangespoor na wat die faktore is wat die swak aankoms gehalte van tafeldruiwe veroorsaak. Hierdie studie het ook masjienleer tegnieke toegepas om die waarskynlike aankoms graderings (groen, geel en rooi) gebaseer op inset veranderlikes wat regdeur die voorsieningsketting versamel is, te voorspel.

Die data wat ontleed is, is verkry uit vyf diverse sekondêre bronne wat bestaan uit inname kwaliteit pakhuis verslae, aankoms kwaliteit verslae, logistieke nominale data, koue ketting temperatuur data en klimaat data. Die uiteindelike datastel het uit 467 waarnemings bestaan. Die toegepaste analitiese proses het uit beskrywende en inferensiële statistiek bestaan om die verhouding tussen die stroomop veranderlikes en die stroomaf resulterende kwaliteit graderings te verduidelik asook hoe die stroomop veranderlikes met mekaar in wisselwerking tree. Die resultate van die voorafgaande analise het gehelp met die kenmerk keuse vir die model-bouproses.

Vier klassifikasie modelle, bestaande uit Logistiese Regressie, *k-Nearest Neighbours, Decision Trees,* en *Random Forests (RF),* is opgelei en geëvalueer. Die *RF*-klassifiseerder het die beste kruis-validasie gradering op die opleidingsdata getoon en is vir verdere evaluering behou. Vir die onsigbare toetsstel was die *RF*-klassifiseerder se akkuraatheid telling 0.63 en het die beste presteer wanneer rooi klas etikette voorspel word, maar het op groen gesukkel en die swakste gevaar vir geel klasvoorspellings. Veranderlikes wat die grootste impak op die aankomskwaliteit gradering gehad het, het bestaan uit die klimaat veranderlikes twee weke voor oes, die spesifieke variëteit en $^{\circ}$*Brix* tydens oes, die aantal verrotte druiwe korrels in die pakhuis gevind, asook die algehele pakhuis kwaliteit gradering en die tipe verpakking wat gebruik word (in “punnets” of los gepak). Die effek van die koue ketting op tafeldruiwe is ook geëvalueer, maar het geen effek vir die 2020-seisoen gehad nie. Die eienskappe van swak gehalte is ook geïdentifiseer in verband met die belangrikste veranderlikes sodat stroomop mitigeringstrategieë bepaal kon word om finansiële eise en voedselvermorsing te verminder.

Die potensiële voordeel van akkurate aankomskwaliteit-voorspellings voor versending sal verbeterde allokasie besluite moontlik maak wat gevolglik tot wins maksimering deur

verliesvermindering en kostebesparings sal lei. Vanuit 'n omgewingsperspektief sal goeie aankomskwaliteit voedselvermorsing aan die einde van die koue ketting verminder en produk raklewe vir verbruikers verleng.

Sleutelwoorde:

Botrytis, Koue ketting, masjien leer, Suid-Afrika, vars vrugte uitvoere, vrugte kwaliteit.

Table of Content

Table of Figures

Table of Tables

Table of Equations

List of Acronyms

QC -	Quality Control
PH -	Postharvest
PHL(s) -	Postharvest loss(es)
FFV -	Fresh fruit and vegetables
FLW -	Food losses and waste
FSC -	Food supply chain
BBR -	*Botrytis* bunch rot
Reefer -	Refrigerated container
FAC -	Forced Air Cooling
TSS -	Total Soluble Solids
VPD -	Vapour Pressure Deficit
RH -	Relative Humidity
PPECB -	Perishable Product Export Control Board
CV -	Cross-validation
RF -	Random Forest

Chapter 1: Introduction

According to the European Commission’s Knowledge Centre for Bioeconomy (2020: 5), 129 metric tonnes of food is wasted in the European Union annually, roughly equating to 20% of all food produced for the region. Fresh fruit is the second largest contributor to waste in absolute terms but when considering the ratio of fruit supplied to what is wasted, the proportion is roughly 41%. Beyond the obvious economic losses that are associated with this waste, there are currently around two billion people globally who suffer from some sort of a vitamin or mineral deficiency and about 45% of deaths amongst children under five are linked to malnutrition (WFP, 2018). These points indicate that there is currently a rift separating the areas where food is plentiful and easily accessible from those that are in desperately short supply.

Due to globalisation, food supply chains have become progressively more complex, bolstered by ever changing consumer expectations regarding variety, freshness, and quality. This results in growing distances between the point of production and consumption, with more intermediaries resulting in an increased risk of loss (Priefer, Jörissen & Bräutigam, 2016).

Of South Africa’s total table grape production, 94% is exported resulting in 77,7 million 4.5kg equivalent cartons moving through the export orientated food chain (FAO, 2016; SATI, 2022). Europe receives over half of all table grapes produced in South Africa annually, equating to 39.9 million equivalent cartons for the 2021/ 2022 season (SATI, 2022: 14). The reason the export market orientation stems from South Africa’s labour supply.

South Africa’s average unemployment rate was 43.2% from 2000 to 2019. This over supply of labour in turn forces wages to remain low, especially for unskilled labour (Trading Economics, 2019). The motivation for South African producers having a foreign market positioning stems from globalisation and the effect of what is known as global labour arbitrage. This is arbitrage of international wage differences without the physical movement of workers due to the removal and reduction of international trade barriers allowing for offshore outsourcing of production. This in effect allows farmers to sell high and produce cheaply, allowing them to increase their profit margin beyond those of their high-income region production competitors (Bottini, Ernst & Luelbker, 2007).

The agricultural sector, with a specific focus on the table grape industry, employed a total of 78 670 people for the 2019/2020 season. The labour force is split into 17% permanent and 83% seasonal workers (SATI, 2020). Table grapes are produced in three South African provinces, which are the Western Cape province, Northern province, and the Limpopo province. The proportion of the rural population for three provinces mentioned which the table grape industry employs, can be seen in *Table 1.1*.

Table 1.1: The table grape industry's contribution to employment in rural areas

Province	provincial population	rural split	Rural Population	table grape labour 2019/2020	proportion employed
Western Cape Province	6 844 272	9.6%	657 050	42 944	6.54%
Northern Province	1 263 875	17.3%	218 650	22 888	10.47%
Limpopo Province	5 982 584	86.7%	5 186 900	12 838	0.25%
					17.25%

Source: Adapted from Civilian Secretariat for Police Service, 2016; SATI, 2020

For the Western Cape and Northern provinces, the table grape industry significantly contributes to the employment of rural people (6.5% and 10.5% respectively) and is of great importance due to the multiplier effect. The multiplier effect is concerned with the impact of forward and backward linkages along the supply chain and the resultant economic impact. From an employment perspective, according to the multiplier effect for each additional R1 million increases in horticultural product sold, an additional 92.8 new jobs are created overall. In comparison, only 29.4 jobs are created in the non-agricultural sector with the same increase in revenue (Fresh Produce Exporters' Forum, 2016: 7).

To ensure that fresh produce arrives at the various global destinations, the fruit is palletised and loaded into refrigerated freight containers (reefers), which are then placed on container ships ready to undergo the two week plus overseas voyage that awaits. Cooling is the most useful method currently implemented to control postharvest ripening and senescence. Postharvest fresh fruit deteriorates as it is living, and therefore, respires to produce energy. Over time, the fruits' energy reserves are depleted, as they are no longer attached to the plant, resulting in senescence and the deterioration of the fruits' quality. Maintenance of low fruit temperature is one of the key factors in fresh fruit quality preservation as it extends the fruits shelf life by slowing the fruits' natural rate of respiration and transpiration, reducing moisture loss. Biochemical reactions are also retarded at low temperatures and most microorganisms that cause decay cannot grow at low temperatures. Therefore, through good cold chain temperature management, fruit quality is maintained.

Good cold chain management results in good quality, but in practise there are many claims linked to poor quality, which are instituted upon table grape shipments departing from South Africa, exported globally. These claims result in a reduced sales price, a loss in shelf life for the end consumer, a reduced return to the grower, and food loss.

Due to the delicate nature of a bunch of table grapes, physiological quality related issues often occur. The cold chain between South African and Europe can last up to 30 days post-harvest to retailer, where the fruit spends 14 to 20 days on a ship at sea. Due to the limited cold chain visibility and the delicate nature of the commodity, fruit that departs the producers' packhouses in a perceived good condition can arrive further along the supply chain in poor condition. Fruit that arrives in an unsatisfactory condition will either be reworked on a carton-by-carton basis or if the quality is too poor the entire pallet can be dumped – both outcomes result in food waste and financial loss due to poor allocation decisions. Without proper accountability, cost is invariably shifted back to the producers at the source without concluding what the causes of the quality deterioration/food waste are.

1.1 Background

The 2020 table grape export season saw a large proportion of shipment rejections due to quality claims resulting in the industry losing a significant amount of potential income. There are many costly value-adding activities that take place between harvesting and foreign consumption, which are still billed whether the product is viable or not. There is, therefore, currently a gap in the industry's understanding of how fruits' condition changes due to the numerous variables associated with exporting a fresh product. Exporters currently make use of in-house quality controllers (QC) to try and ensure a standard across the multitude of producers. Unfortunately, the adjustment has not substantially reduced the proportion of consignments arriving in poor quality. Table grape quality is evaluated and graded on arrival, according to a three-point Likert scale, namely green, amber, and red. A green indicates sound quality, an amber a potential claim (QP), and red a definite quality claim (QC). Arrival inspections are conducted either by inhouse or independent fruit quality surveyors.

The proportion of poor quality arrivals are further exacerbated due to the current fruit export model, which sees allocation decisions being made long before the product is to be consumed.

Exporters are incentivised to maximise producer returns by securing the best possible prices. If the producers deliver a substandard product to the forementioned programs, the fruit will be subjected to financial claims, such as sorting and repacking costs, which are subtracted from the sales price. This results in a return below the prior negotiated price, reducing the final net return to the grower. There, is therefore, a need to find the balance between volume and fruit quality to ensure maximum returns.

Pre-season, exporters meet with retailers and negotiate the following season's prices and volumes suppliable per variety group per week. During the season, exporters receive weekly harvest forecasts from producers prior to the pack week. According to these forecasts, fruit is allocated to the various retail programs secured. Currently, the quality of the product is not being evaluated thoroughly enough, resulting in sub-optimal fruit being allocated and exported.

1.2 Problem Statement

Producers and packhouse managers would ideally like to pack 100% of the grapes produced, which is unfortunately to their own detriment. The threshold of acceptable quality is currently decided purely based on a producer's perception of their own product. Exporters have tried to overcome this problem by employing the in-house quality controllers, to better police the packhouses, but ultimately producers still have the final say as to what ends up in the box. The current model is flawed as exporters allocation decisions are based on the producers' perception of their products' quality. Instead, information should flow between the producers (point of production) and receivers / consumers (point of consumption) so that informed decisions can be made at the point in the chain where unnecessary additional costs can be avoided. What instead occurs is substandard fruit being packed and shipped halfway around the world, at great expense, only to be rejected due to poor quality.

The potential factors attributing to poor arrival quality are known, but how and to what extent these factors play a mutual role in causing claims has not yet been investigated in an industry setting.

This study, therefore, determines how and to what extent these factors cause quality deterioration and if there is in fact a link between harvest quality and arrival quality. In addition, machine learning modelling techniques are used to predict the arrival quality prior to export, to aid exporters allocation decisions to mitigate losses. The effect of the supply chain is also considered and evaluated.

1.3 Aims and Objectives

The aim of this study is to determine which variables negatively contribute to the arrival table grape quality of an export orientated supply chain between South African and Europe. Machine learning techniques are then applied to the variables identified, with the aim of developing a classification model which predicts the probable arrival quality scores of table grapes.

Poor arrival quality results in a loss of income for producers and exporters, as well as end-of-chain food waste. Effective upstream diagnosis of the probable quality of table grapes on arrival in Europe will improve allocation decisions, allowing exporters to maximise their returns while reducing waste downstream.

1.3.1 Objectives

The objectives of the study are:

- To determine the variables that are associated with arrival table grape quality.
- To identify whether the export supply chain between South African and Europe contributed to adverse table grape quality for the 2020 season.
- To apply and build appropriate machine learning models for the classification task.
- To identify and evaluate a model that performs well for this classification task.

- To understand how the best performing model applies the variables to predict the arrival quality of table grapes.

1.4 Research Questions

To achieve the study's aims and objectives, the following research questions are investigated:

Primary Question

What are the causes of poor table grape quality on arrival, and can these variables be modelled to predict the arrival quality prior to export?

Secondary Questions

1. What are the preharvest factors that influence table grape quality?
2. What are point of harvest factors that influence table grape quality?
3. What are the postharvest factors that influence table grape quality?
4. What are the characteristics of a product that arrives in poor quality?
5. Does the supply chain (from the point of loading onwards) contribute to poor arrival quality of table grapes?
6. Can machine learning be employed to predict the downstream quality of table grapes?
7. Do models perform equally well at predicting the three arrival score class-labels?

1.4.1 The relationship between the research objective and questions

The objectives and related research questions for this study are shown in *Table 1.2*.

Table 1.2: Research objectives and related research questions

Research Objectives	Research Questions	Addressed in Chapter
To determine the variables that are associated with arrival table grape quality.	1. What are the preharvest harvest factors that influence table grape quality?	Two and Seven
	2. What are point of harvest factors that influence table grape quality?	Two and Seven
	3. What are the postharvest factors that influence table grape quality?	Two and Seven
	4. What are the characteristics of a product that arrives in poor quality?	Seven
To determine the variables that are associated with arrival table grape quality.	5. Does the supply chain (from the point of loading onwards) contribute to poor arrival quality of table grapes?	Two and Seven

To identify whether the export supply chain between South African and Europe contributed to adverse table grape quality for the 2020 season.		
To apply and build appropriate machine learning models for the classification task. To identify and evaluate the best performing model with relevant performance metrics.	6. Can machine learning be employed to predict the downstream quality of table grapes?	Three, Six, and Seven
To understand how the best performing model applies the variables to predict the arrival quality of table grapes.	7. Do models perform equally well at predicting the three arrival score class-labels?	Seven

Source: Author's own

1.5 Conceptual Framework

A conceptual framework is a structure that the researcher best believes can explain the natural progression of the phenomenon being studied. The conceptual framework describes the relationship between the main concepts/theories and gives an explanation of how the research problem was explored (Dickson, Adu-Agyem & Emad Kamil, 2018). It is the simplest way in which a researcher can present their remedies to a problem and accentuates the reason why a research topic is worth studying (Grant & Osanloo, 2014).

Figure 1.1 displays the conceptual framework used in this study. The framework explains how data was collected, focusing on both quantitative and qualitative techniques to determine what the main causes of poor arrival quality of table grapes are. The variables identified were then used to train machine learning models to predict the arrival quality scores.

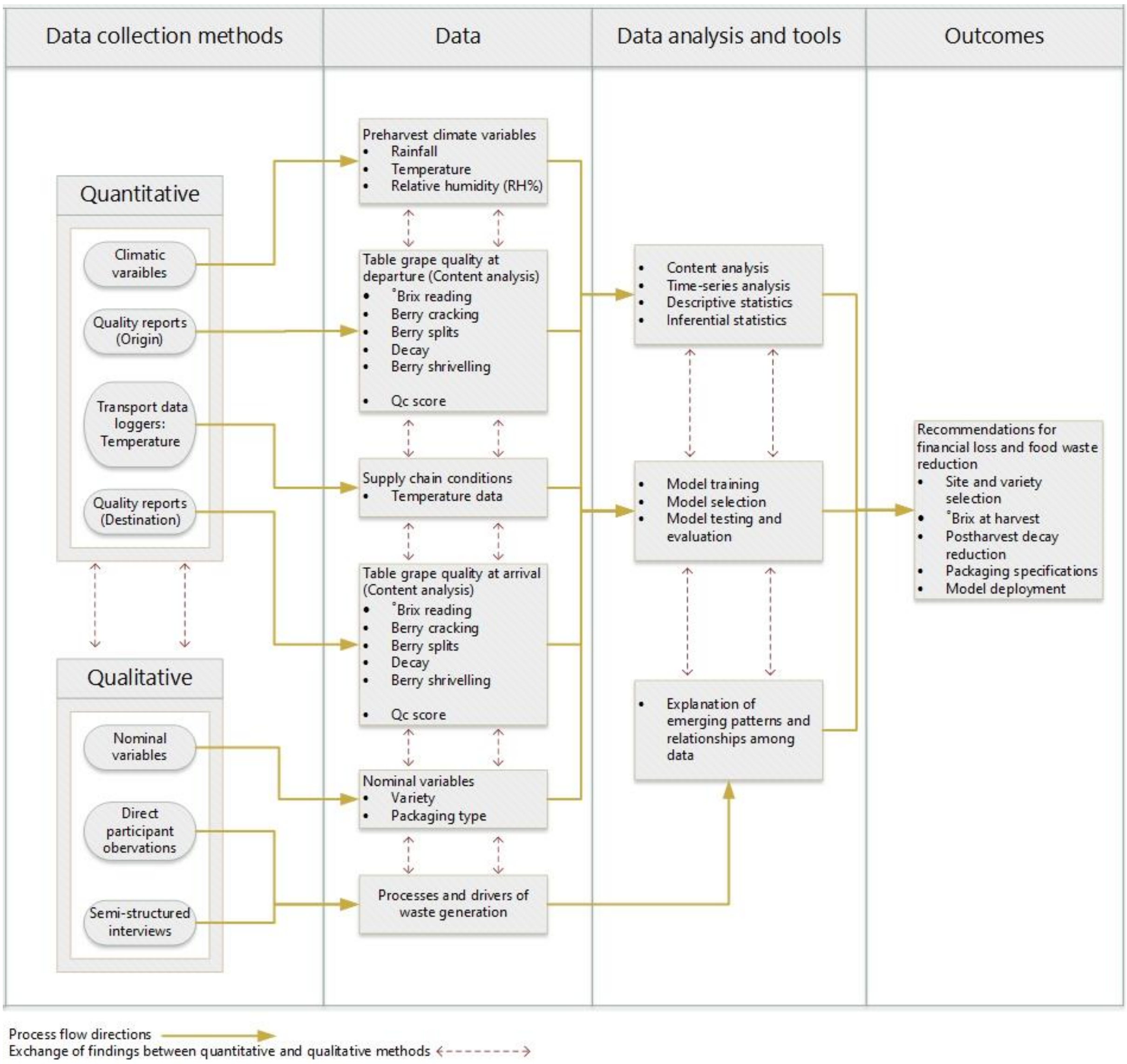


Figure 1.1: Conceptual framework of research study

Source: Adapted from Papargyropoulou, Wright, Lozano, Steinberger, Padfield & Ujang, 2016

1.6 Chapter outline

The outline of the thesis is as follows:

1.6.1 Chapter 2: Literature Review - General

Chapter 2 includes all the relevant literature pertaining to table grapes. The background of the South African table grape industry is provided, which includes previous studies that have looked at the South African table grape cold chain, as well as international studies conducted on fresh fruit and vegetable (FFV) food waste throughout the food supply chain. The factors influencing as well as the common disorders that reduce table grape quality are explored, and potential best practises that can be implemented to mitigate losses are identified.

1.6.2 Chapter 3: Literature Review – Machine Learning

Chapter 3 examines all the relevant literature related to machine learning and predictive modelling. The basic machine learning principles are introduced with specific reference to model complexity decisions and how those impact the interpretability of the results. The techniques to apply depending on the predicted variables level of measurement and the types of models available are also discussed. Feature selection and different model evaluation techniques are considered to assess the prediction performance.

1.6.3 Chapter 4: Research Design and Methodology

Chapter 4 discusses the flow of the research methodology. This study was aided by the research process onion, which summarises the important considerations that were undertaken during the research.

1.6.4 Chapter 5: Data Pre-processing

Chapter 5 described all the data extraction, wrangling, joining, and cleaning steps undertaken to extract the raw data from the five unique sources and to then join the data creating a complete dataset for analysis and modelling.

1.6.5 Chapter 6: Modelling

Chapter 6 explained the steps in the model training process which saw the building and training of four types of machine learning models. The model that performed best on the training dataset was retained for further evaluation.

1.6.6 Chapter 7: Data Analysis

Chapter 7 contains the data analysis for the complete dataset. The data analysis seeks to explain the relationship between the independent variables and the dependant variable, the arrival quality scores of table grapes. The data is illustrated graphically and tabularly and includes discussion points identifying problem areas. The analysis starts with descriptive statistics, then moves on to inferential statistics, and ends off with the evaluation of the machine learning modelling results.

1.6.7 Chapter 8: Interpretation of Results

Chapter 8 consists of interpretations of the data analysis conducted in *chapter 7*, which include possible explanations for the outcomes found. Solutions are also suggested to reduce the poor arrival quality scores that result in financial claims and food waste.

1.6.8 Chapter 9: Conclusions and Recommendations

Chapter 9, the final chapter, gives a summary of the findings and recommendations that stakeholders can implement throughout the supply chain to reduce the effect of table grapes arriving in poor quality. Suggestions are made regarding future research avenues in this field.

Chapter 2: Literature Review - General

The literature review gives a brief contextualisation of the South African table grape industry historically and today, how a postharvest table grape cold chain operates, what food waste is and how it pertains to fresh fruit and vegetables. The pathological and physiological disorders that influence table grape quality are explored and various mitigation practises are considered.

2.1 Historical background

Grapes, from the genus *vitis,* cultivation records date back to the Neolithic era (6 000 to 6 500 BC) in the Near East, around the Caspian Sea. Around 4 000 BC, when the Hittites migrated from modern day Turkey to the Greek island of Crete, they introduced grapes to Europe (Fresh Produce Exporters' Forum, 2016). The Europeans have since had a long history of grape cultivation and celebration, who ultimately spread the plant globally.

Grape vines arrived in South Africa in 1652, the same time the Dutch settlers founded a supply station at the Cape of Good Hope on behalf of the Dutch East India Company (PPECB, 2003). The vineyards were planted to produce wine to ward off sailors' scurvy as they continued their voyage round the tip of Africa. This was during a time before it was commonly known that the disease was caused by a lack of vitamin C. Fortunately for South African grape producers, the identification of the real source of sailors' poor health did not result in the destruction of the numerous vineyards planted.

Grapes were commonly used to produce wine or eaten fresh locally. This all changed when a French engineer by the name of Charles Tellier started experimenting with refrigeration in 1868. This ultimately allowed the South African fruit export industry to germinate. The first trials began with grapes wrapped in newspaper, which astonishingly survived the 19-day sea voyage (Symington, 2008). The success of that initial trial prompted the first commercial volumes of fresh fruit, consisting of 14 cases of peaches, which were exported in January 1892. The peaches arrived in Southampton in perfect condition and by June of that same year, the Cape Fruit Syndicate had exported a total of 1 900 cases of grapes and 6 000 cases of apples and pears (Fresh Produce Exporters' Forum, 2016).

Since those early days, the South African table grape industry has become a global leader, exporting the sixth largest volume globally with a market share of 8% (SATI, 2022: 41). South Africa produced a total of 66 149 984 4.5kg equivalent cartons in the 2019/2020 season, an increase of over 30% from the 2013/2014 season (SATI, 2020).

2.2 Industry review

According to the South African Table Grape Industry (SATI, 2019), South Africa is the Northern Hemisphere's oldest and most reliable table grape supplier. The South African table grape industry consists of five main production regions located across the country. The regions consist of those concentrated in the Western Cape, along the Orange River and the remainder are in the Northern Provinces. *Figure 2.1* identifies the five regions and indicates the production share emanating from each geographic area.

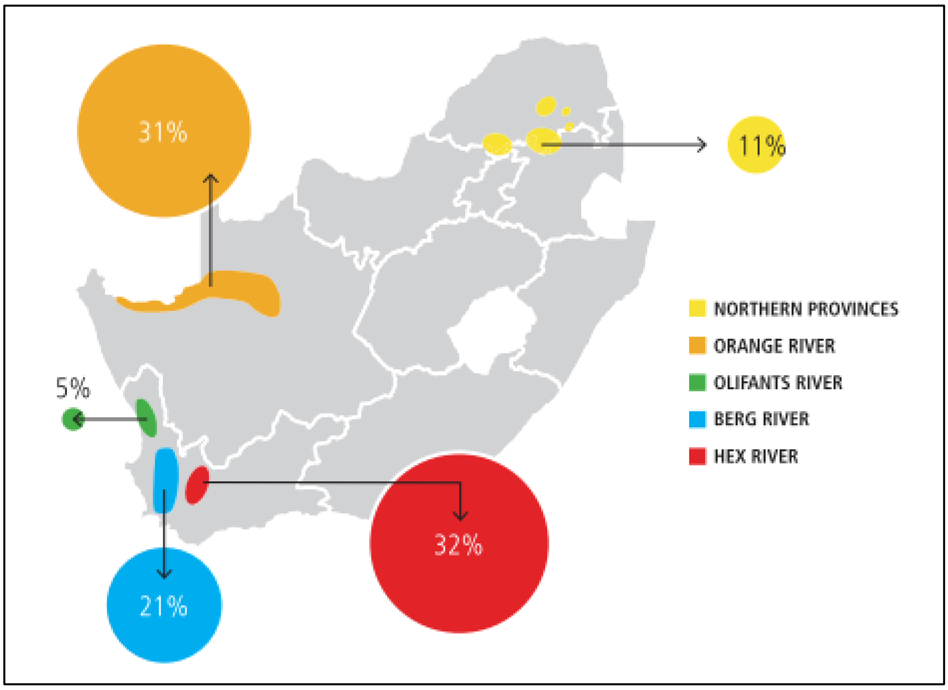


Figure 2.1: South African production regions

Source: SATI, 2018

As of 2019, the total South African land area planted with table grapes consisted of 21 798 hectares. The planted area is distributed with the Hex River accounting for 30% (6 619 ha), Berg River 24% (5 210 ha), Olifants River 5% (1 185 ha), the Orange River 28% (6 195 ha) and the Northern Provinces 12% (2 589 ha) respectively. To contextualise the scale of the table grape industry, one must compare the area planted to the total area planted of all deciduous fruit cultivars. Deciduous fruits refer to trees, shrubs and vines that lose their leaves during winter. The total area planted of deciduous fruit comprises of 80 738 ha, where table grapes account for 27% thereof.

South Africa has a varied climate, which impacts the growing season of the various regions. The growing season is impacted by local weather conditions, elevation, and the proximity to the equator. Table grapes grow best in temperate regions, with hot dry summers and cold wet winters, known as a Mediterranean climate. *Table 2.1* indicates the harvest periods of each production region throughout the year.

Table 2.1: Production periods and volumes of various regions

Province \ week	Oct	Nov.				Dec.				Jan.				Feb.				Mar.				Apr.				May
	44	45	46	47	48	49	50	51	52	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Northern Provinces	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■										
Orange River	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■									
Olifants River					■	■	■	■	■	■	■	■	■	■	■	■	■	■								
Berg River					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				
Hex River								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Source: adapted from SATI, 2022

Due to the different growing seasons, the regions located in the northern part of the country harvest earlier due to their proximity to the equator. *Figure 2.2,* which displays the weekly production volume of each region, further reinforces this phenomenon.

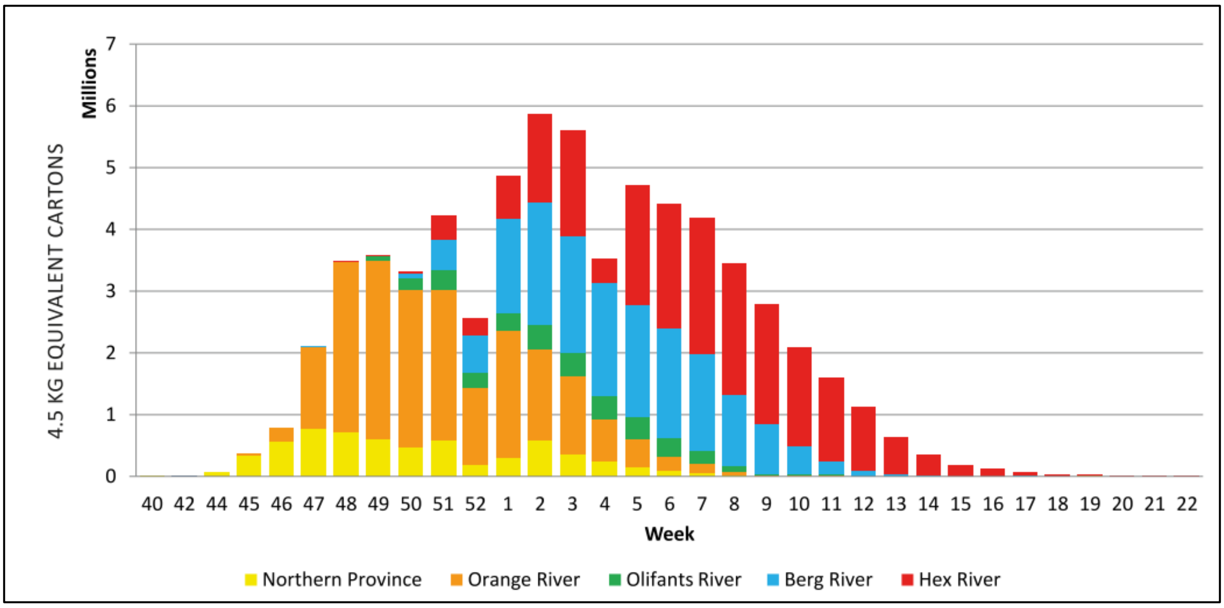


Figure 2.2: Weekly production volume per region

Source: SATI, 2020: 12

From a global trade perspective, the northern hemisphere nations are in the depths of their winter during South Africa's peak production weeks. This sets up a market climate where northern hemisphere nations are unable to produce their own table grapes locally, giving southern hemisphere producers the opportunity to fill an undersupplied market and, in turn, realise higher prices. These factors have resulted in South Africa having an export-orientated table grape industry exporting a total of 77 732 736 4.5 kg equivalent cartons (368 366 tonnes) (SATI, 2022). This production volume ranks South Africa as the sixth largest table grape exporter globally, as seen in *Figure 2.3,* with a market share of 8% in 2021.

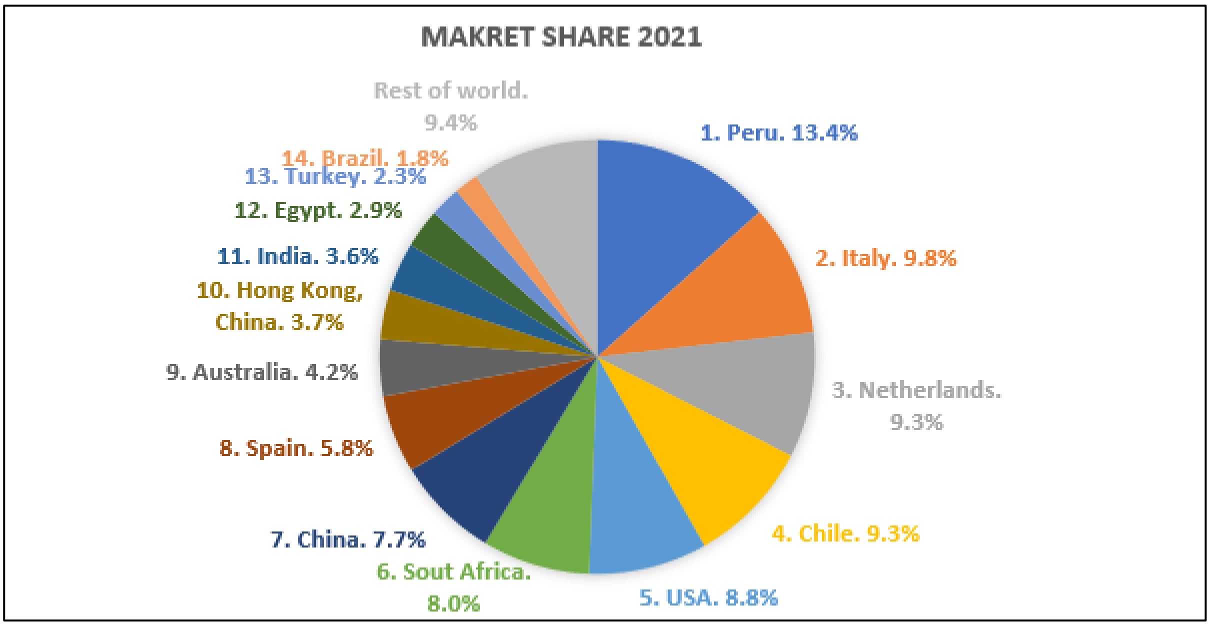


Figure 2.3: Market share of largest table grape exporting countries

Source: adapted from SATI, 2022

The high export market share is achieved in spite of the fact that South Africa only produces a total of 315 000 tonnes of the world's 27 million tonne table grape production. This ranks South Africa thirteenth in terms of total table grape production (*Statistical Report on World Vitiviniculture*, 2019), equating to a total production percentage of 1.16%.

The main destinations for South Africa's table grapes are displayed in *Figure 2.4.* The EU remains South Africa's largest fruit destination, receiving 136 873t of table grapes and the UK in second place receiving 64 438t. The Far East has recently become the third largest recipient of South African table grapes, receiving 70% more fruit in the 2018/19 season compared to the 2015/16 season (growth from 2 355 329t to 3 366 223t).

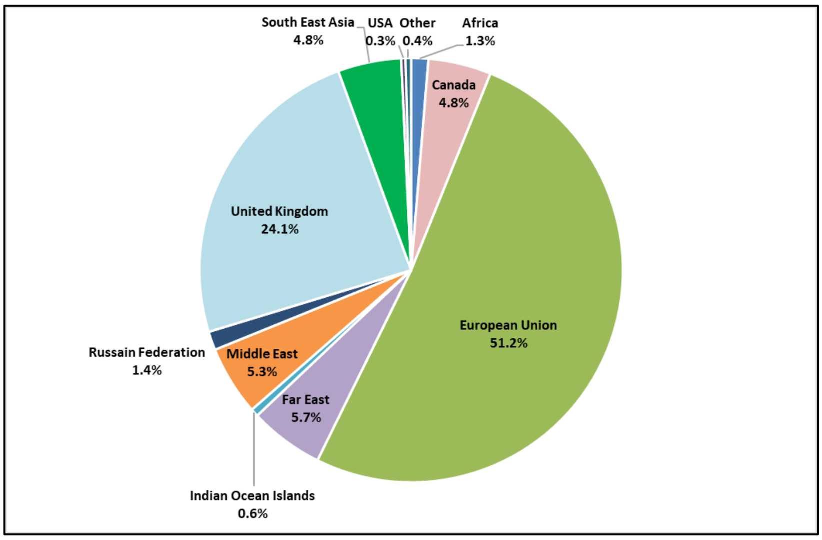


Figure 2.4: Main export destinations for South Africa's table grapes

Source: adapted from Phaleng & Tshitiza, 2019

2.3 Supply chains

The Council of Supply Chain Management Professionals (CSCMP) (2017), define a supply chain as "a process that starts with unprocessed raw materials and ends with the consumers using the finished product."

Supply chains are closely connected to logistics in that a supply chain is the many ways in which a product is transformed along the steps in a logistical chain. Therefore, the definition of a supply chain is most similar to that of the value chain, except the basis of a supply chain is logistics (Mashabela, 2007: 14).

Physical distribution is an important process of the table grape industry and adds value in the form of place and time utility as the fruit is moved from the point of production to the market where the demand exists. There is also an aspect of form utility in terms of the various packaging alternatives. The research context examines how table grape quality varies along an export orientated supply chain between South Africa and Europe.

2.3.1 Cold chains

From a supply chain perspective, a cold chain is defined as the seamless movement of fresh, chilled or frozen products from production region to the market, through various storage and transportation mediums without any change in the optimum storage temperature and relative humidity (PPECB, 2013). Goedhals-Gerber, Stander & Van Dyk, (2017: 363) define a cold chain as a temperature-controlled supply chain that allows for national and trans-national trade in perishable products such as fresh fruit and vegetables.

A successful cold chain offers a competitive advantage in quality maintenance, a prolonged shelf life for the end consumer, increasing profits for producers, and reducing food waste throughout the cold chain. It is, therefore, of vital importance that the cold chain is maintained and performs well.

2.4 Postharvest supply chains

Traditional postharvest food systems comprise of interrelated activities spanning from the time of harvest through crop processing, marketing, up until the point of sale to the final consumer (pre-consumer) (Gogh, Aramyan, Sluis, *et al.*, 2013). A postharvest system includes all activities in the chain intended to add value to the final product.

For the scope of the research, preharvest activities (plant propagation, irrigation, fertilisation, cultivation) are excluded.

2.4.1 Generic export orientated South African table grape supply chain

A generic South African table grape supply chain consists of multiple interrelated stakeholders that ensure the product effectively moves from point of production to consumption. *Figure 2.5*, displays the typical stakeholders' and structure of a table grape export chain originating from South Africa. Refer to *Appendix A* for a more detailed process flow.

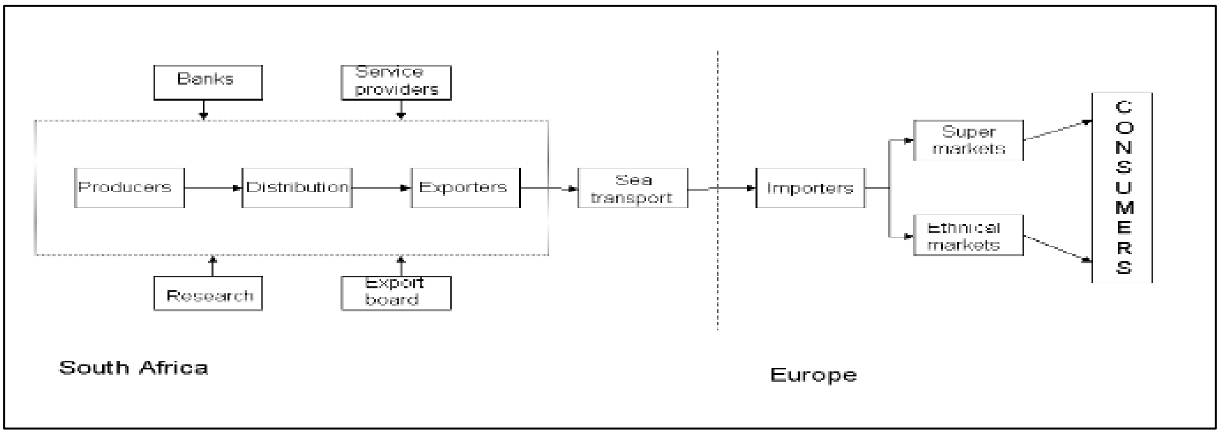


Figure 2.5: Table grape export chain

Source: Trienekens & Willems, 2007

<u>Producer/ packaging</u>

A producer's core business is to cultivate a high quality product within "Good Agricultural Practice" protocols (DAFF, 2017). Farmers need to be aware of ever-changing market demands and to supply those varieties cost effectively with consistency and reliability.

Farmers produce their product and package it either at small farmer run packhouses or at large centralised packhouses, packaging several farmers' fruit under a single label, often consisting of multiple product lines. There are various packaging alternatives, but the most common variants consist of: (i) 4.5kg cartons; (ii) 5kg trays containing clamshell punnets (10 x 500g); and (iii) 5kg trays containing open-top punnets (10 x 500g). All three of these variants are unitised on pallets. Four and a half kilogram (4.5kg) cartons are the most common packaging option and form pallet dimensions of 180 stacked cartons per pallet.

<u>Cold Storage</u>

The packaged fruit, once palletised, is mostly taken to a local cold storage facility, either run as a cooperative or privately owned. Cold stores are responsible for receiving, handling, and cooling the table grapes to the required temperature to ensure sufficient freshness and shelf life at the end of the supply chain. The correct loading protocols into truck or containers must be followed according to approved regulations stipulated by the Perishable Product Control Board (PPECB).

The cold stores act as a logistical node where the fruits' temperature is brought down to -0.5°C through the process of forced air cooling (FAC) to reduce postharvest losses (Henning & Chetty, 2019). The FAC process can take up to 24 hours to get the pulp temperature down to the required level.

Once the fruit has gone through the FAC process, it is ready to be loaded into shipping containers or refrigerated trucks. Forty-foot reefer high-cube containers are used and fit either 20 or 21 pallets depending on the size of the pallet base. The Hex River production region's cold storage facility, Hexkoel, loads between 50 and 100 shipping containers each evening, depending on which cultivar is being harvested at that time.

Transporters (freight forwarders)

The fruit is then transported by container trucks to the port where the containers are loaded onto ships destined for various international markets. The inland transportation is contracted out to various local logistics service providers. Container trailers are equipped with modern cooling technology such as 'gensets', which supply electricity directly to the reefer container to keep the fruits' temperature at a constant. It is not a requirement to make use of a genset during transportation, but the general practise is to utilise a genset trailer when the travel distance is longer than 2 hours' drive from the port (Herholdt, 2021).

The containers arrive at the port terminal and are subsequently loaded onto the ships. The Port of Cape Town suffers from heavy winds during the deciduous fruit exporting period between December and February. This often results in loading delays at the port. Due to the produce's short shelf-life, as discussed above, these loading delays can have major consequences for exporters.

Delays longer than a one day can result in significant costs, including:

- Ship demurrage costs
- Labour costs – stevedoring
- Labour costs – shoreside
- Transfer of fruit between inland cold stores
- Lost sales opportunities
- Risk of container/ pallet being rejected by quality controllers (QC) at foreign retailer due to delays

Exporters

The primary role of exporters is to market and sell producers fruit at the best market price attainable through negotiations (DAFF, 2017). Prior to deregulation in 1997, South African fruit was exported through a single marketing channel consisting of two governmentally regulated bodies, Outspan and Unifruco Ltd (PPECB, 2003). This allowed for supply chain simplicity and optimisation, but at the detriment to farmers who would receive one market price regardless of the quality of fruit produced (Trienekens & Willems, 2007). The market has since been deregulated, allowing for greater competition resulting in a plethora of new marketing agents. According to the Fresh Produce Export Forum (2018:7), there are currently 110 exporters that account for 90% of the fruit exported from South Africa. Seventy of those exporters either market only table grapes or other deciduous fruits including table grapes.

Exporters' activities involve facilitating fruit transportation from production regions to foreign importers. These activities include consolidating shipments, negotiating prices with freight forwarders and importers as well as opening new markets. Their operations are extremely reactionary due to market price fluctuations, which are influenced by many factors including:

- Global supply of fruit on the market from foreign producers
- Exchange rate fluctuations
- Consumer preferences and trends
- Farm accreditations

There are two distinct fruit sales channels available to growers. The first approach is to sell the produce directly to an importer, with or without the help of an agent (larger producers). Second, to supply fruit to a conglomerate that pools fruit to take advantage of economies of scale (DAFF, 2017).

Some growers have pooled their fruit and resources and have founded private cooperative export companies, an autonomous association of growers who jointly own and democratically control the enterprise. This allows growers to control the entire marketing chain from production to delivery, cutting out any potential third parties allowing for greater value-added. This method also allows for branding, which is in turn supported by partnerships with breeding programs. This gives growers access to new grape varieties tailored to various markets, allowing for greater financial returns.

Markets

The foreign markets that received the largest quantities of South African table grapes for the 2021/22 season consisted of the EU (39.9 mil.), UK (16.6 mil.), Canada (4.4 mil.), Middle East (3.9 mil.) and South East Asia (3.5 mil.) 4.5kg equivalent cartons (SATI, 2022) . There are eleven major market regions, where the five listed receive the largest portion of produce exported.

Market agents (traders), mentioned above, act as middlemen who establish contact between producers/ exporters and buyers from importing countries. They work off commissions, usually taking between 2% to 3% commission on the sales price depending upon the Incoterm agreed. Importers by contrast buy and sell their own capacity, assuming full risk unless working on consignment. Their margin will usually consist of 5% to 10% commission due to their increased risk as well as value adding responsibilities, such as customs clearing and distribution compliance paperwork. An example of an account sale for fruit sold on consignment can be found in *Appendix B*, where the importer deducted a 5% commission from the sales price for their services rendered.

Few exporters have long standing contracts with retailers who deliver directly. This places more risk on retailers who can then be forced to receive substandard fruit, but due to the increasing importance of production and ethical practise standards (GLOBALGAP, GRASP, TNC, FSMA, SPRING SIZA, BRC/HACCP, etc.) long term planning contractual relationships are expected to increase (DAFF, 2017).

2.5 Food loss/ food waste / postharvest loss

The Food and Agricultural Organisation of the United Nations (FAO) states that a distinction must be made between two concepts, these are defined below:

- **Food loss** is the decrease in mass (dry mass) or nutritional value (quality) of food that was originally intended for human consumption. This can include spills, spoils or loss that occur before the produce reaches the consumer resulting from current institutional food production supply system frameworks, unintended agricultural processes or technical limitations in the supply chain infrastructure.
- **Food waste** refers to food that is appropriate for human consumption but is discarded whether or not it is kept beyond its expiry date or left to spoil. This is the removal of food fit for human consumption due to a conscious decision or negligence on the actors' part.

The FAO also define a third term ***Food wastage,*** which encompasses both the food loss and food waste concepts, which can be interpreted as food that is lost due to deterioration or waste.

The Postharvest (PH) loss can be described as "losses between harvest and the onward supply of produce to market, which broadly equates to waste in the food supply chain" (Gogh *et al.*, 2013). PH loss is, therefore, food that is lost due to suboptimal performance by supply chain actors throughout, and therefore, encompasses aspects of both food loss and food waste. Factors that cause postharvest waste along a food supply chain are summarised in *Table 2.2*.

Table 2.2: Factors causing postharvest waste along food supply chains

Postharvest supply chain stages	Factors causing loss
1) Harvesting	• Incorrect harvesting techniques • Careless handling • Distance between orchard and packhouse • Packhouse ambient temperature
2) Transport	• Carless handling • Temperature during transport & transhipment
3) Shipment rejections	• Quality rejection due to food safety standards • Lack of skilled work force • Cost of certification • Lack of enabling environment
4) Processing	• Various processing steps • Packaging damage • Technical malfunctions • Demand sizes
5) Retail level	• Improper shelf-life labelling • Product flaws • Discounts • Overstocked shelves
6) Consumer	• Discounts • Lack of planning (unaware of what food is at home) • Improper storage techniques • Oversized dinner plates • Over preparation resulting in leftovers being discarded • Insufficient packaging (protection, apportionment, shelf life, size [too big packaging], resealing properties)

Source: adapted from Stroecken, 2017

2.5.1 The extent of global food wastage

Food wastage is significantly different along various food value chains and depends on the type of product as well as the region(s) where production and consumption occur. Global dietary transitions, moving away from starchy staples toward more delicate, short shelf-life produce, is associated with greater probabilities of food loss due to their highly perishable nature (Louw, 2017:9).

According to Gustavsson *et al.* (2011:4), supply chain food waste differs according to a country's income distribution. It has been found that medium/high-income regions tend to have a greater proportion of waste occurring near the end of the supply chain (roughly 40% at retail/ consumer stage), whereas lower-income nations tend to waste earlier on in the chain (roughly 40% during postharvest/ processing level). These results indicate that the net waste amongst low- and higher-income regions is roughly equivalent. The cost of food waste at the end stage of the supply chain is far higher since value-adding activities such as transportation, storage and processing have to be added on top of the initial production and resource outlays.

With a specific focus on fresh fruits and vegetables (FFVs), Gustavsson *et al.* (2011:7) determined that the major cause of waste is grading requirements set by retailers. *Figure 2.6* further outlines how waste is spread across the supply chain in various regions.

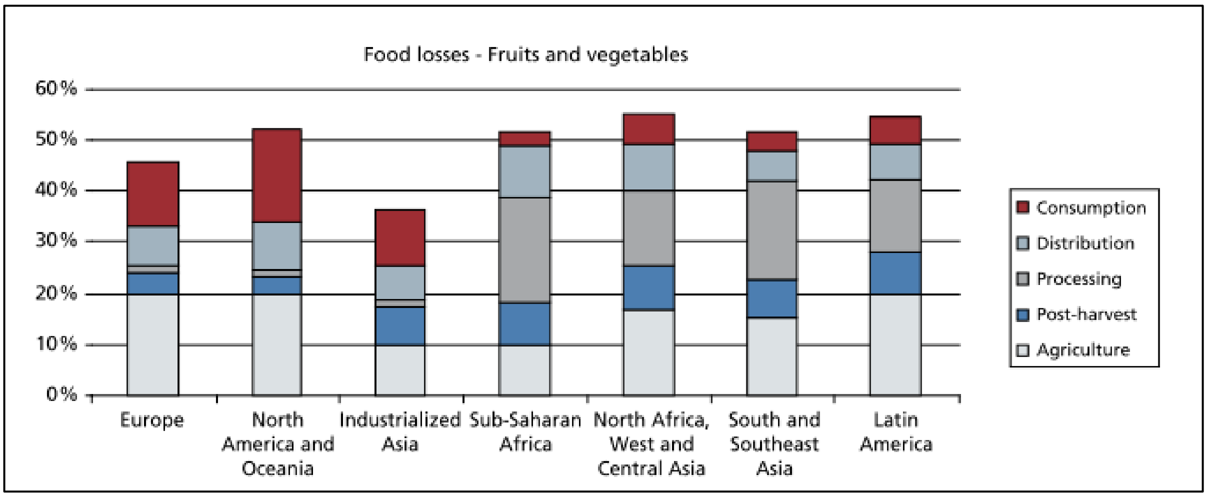


Figure 2.6: Parts of the initial production lost or wasted at different stages of the food supply chain (FSC) for FFV in different regions

Source: Gustavsson et al., 2011

The supply chain applicable to this thesis spans two continents that fall into two contrasting income distributions. The production region, South Africa, is considered a low-income region whereas Europe, the region of final consumption, is a high-income region. According to the literature, the implications of this global supply chain should constitute a sizable net percentage of postharvest food waste from the total production initially intended for human consumption. To further bolster this point, it has been estimated that 70% of The EU's food waste arises in the end chain role players, namely food services, retailers and households (Stenmarck, Jensen, Quested, *et al.*, 2016).

2.6 Packaging logistics

Table grapes are sold in two packaging types, as mentioned in *section 2.4.1*, bulk or punnets. The punnet packaging design can further be subdivided into those that have lids that can open and close (clamshell punnets) or those that are heat-sealed with flow wrap film, which can only be opened once, preferably by the consumer. With bulk packaged grapes or clamshell punnets, berries that are mouldy/rotten/decayed or exhibit other unfavourable physiological conditions, can be removed by retail staff. In contrast, the sealed punnet's, which cannot be opened, entire contents will be discarded due to one or two suboptimal berries. This results in wastage of edible food only because a diminutive portion of the packaged content is suboptimal.

2.7 Factors influencing table grape quality in the cold chain

There are a variety of factors which have a marked impact on fruit quality. These factors consist of four major physiological processes, which are transpiration (better known as moisture loss), condensation, respiration, and mechanical damage. These processes are directly influenced by a myriad of environmental factors, which consist of – temperature; relative humidity; mass transfer due to the rate of airflow and the fruit's skin permeability; packaging; and poor handling.

Table grapes are non-climacteric fruit; they do not continue to ripen after harvest as there is no dramatic increase in respiration or ethylene production coincident with ripening. The response of fruit ripening and the low respiration rate determine that softening is not a relevant postharvest problem as in the majority of other fruits (Zoffoli, 2008: 415).

2.7.1 Psychrometry: balancing temperature and humidity

Psychrometry is the science of determining the thermodynamic properties of gas-vapour mixtures (Singh & Heldman, 2014). By measuring some of the environmental factors, such as the relative humidity (RH%) and the ambient temperature (dry-bulb temperature), it is possible to predict the fruit pulp (wet bulb) temperature, as well as the rate of transpiration or condensation. The use of psychrometry is, therefore, vital when managing a perishable product cold chain to ensure good end of chain product quality.

Figure 2.7 is an example of a psychrometric graph, which represents the interactions of various thermodynamic properties such as dry and wet bulb temperature, humidity, enthalpy, air density, and the water vapour pressure.

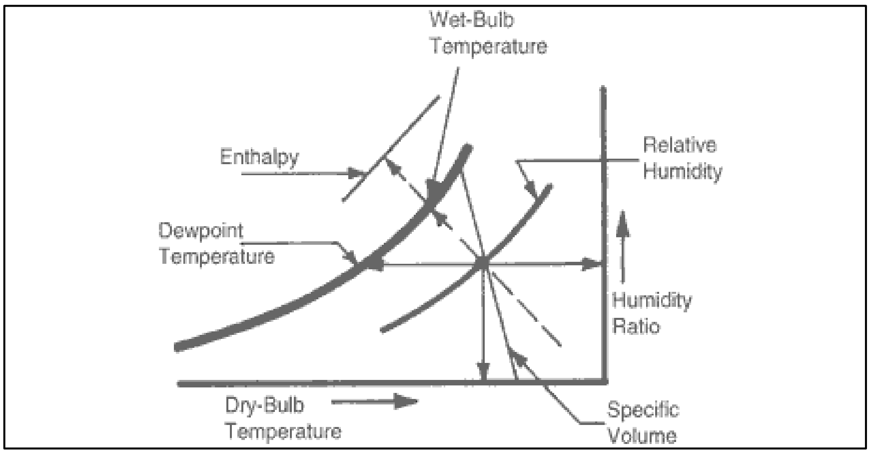


Figure 2.7: Elements of a psychrometric chart

Source: Becker, Misra & Fricke, 1996

The driving force behind moisture loss is the vapour pressure deficit (VPD), which can be determined from psychrometric chart readings. As seen in *Equation 1*, VPD can be determined by subtracting the vapour pressure of the air from the saturated vapour pressure – the vapour pressure at the surface of the fruit.

$$VPD = VP_{sat} - VP_{air} \tag{1}$$

Where:

Vpd	=	the difference between the vapour pressure (kPa) at the fruit surface and the vapour pressure (kPa) of the surrounding air
VP_{sat}	=	the vapour pressure at the surface of the fruit
VP_{air}	=	the vapour pressure of the ambient air

Equation 1: Vapour pressure deficit

Source: Ngcobo, 2013

Psychometry is relevant for the postharvest handling of agricultural products because fruits and vegetables continually transpire, hence the storage environment needs to be manipulated to ensure quality.

2.7.1.1 Relative humidity (RH%)

Relative Humidity is the measure of the amount of water air can hold at a certain temperature. RH is a ratio of the density of water vapour in the air to the density of saturated water vapour at the dry bulb temperature of the air (Singh & Heldman, 2014). RH is not an absolute amount but is rather a relative indication of the amount of moisture in the air at a specific temperature. RH can, therefore, be calculated according to *Equation 2*.

$$RH\ (\%) = \left(\frac{water\ vapour\ content}{water\ vapour\ capacity}\right) \times 100 \tag{2}$$

Equation 2: Relative Humidity

Source: Thompson, Mitchell, Rumsey, Kasmire & Crisosto, 2008

When air temperature increases, water molecules in the air are warmed, giving them more energy. With an increase in energy (higher temperatures), the water molecules are less likely to condense implying that warm air can contain more water molecules than cooler air.

2.7.1.1.1 Dew point

The dew point is the saturation temperature of moisture present in a sample of air, which can be defined as the temperature at which water vapour changes into liquid (condensation). This occurs when the RH of the air is at 100% (Fresh Produce Exporters' Forum, 2016).

Dew-point temperature is when the air-vapour mixture is cooled at a constant air pressure and constant humidity ratio. A temperature will be reached where the air-vapour mixture becomes saturated. Further lowering the temperature will result in condensation (Singh & Heldman, 2014).

2.7.2 Storage conditions

The process of forced air-cooling (FAC) or 'pressure cooling', rapidly pulls air through the packaged cartons of fruit, greatly increasing the cooling rate and subsequently reducing the cooling time (taking a 1/4 to 1/10 of the time needed for room cooling). The process effectively increases the surface area being cooled from that of the carton to that of the product contained within – the grapes, changing the cooling process from a conductive one to a more convective process. Room cooling uses a conductive cooling approach, a slow process involving heat transfer from one fruit to the next (within the pallet), as the 'cold front' moves through the pallet, visible in *Figure 2.8*.

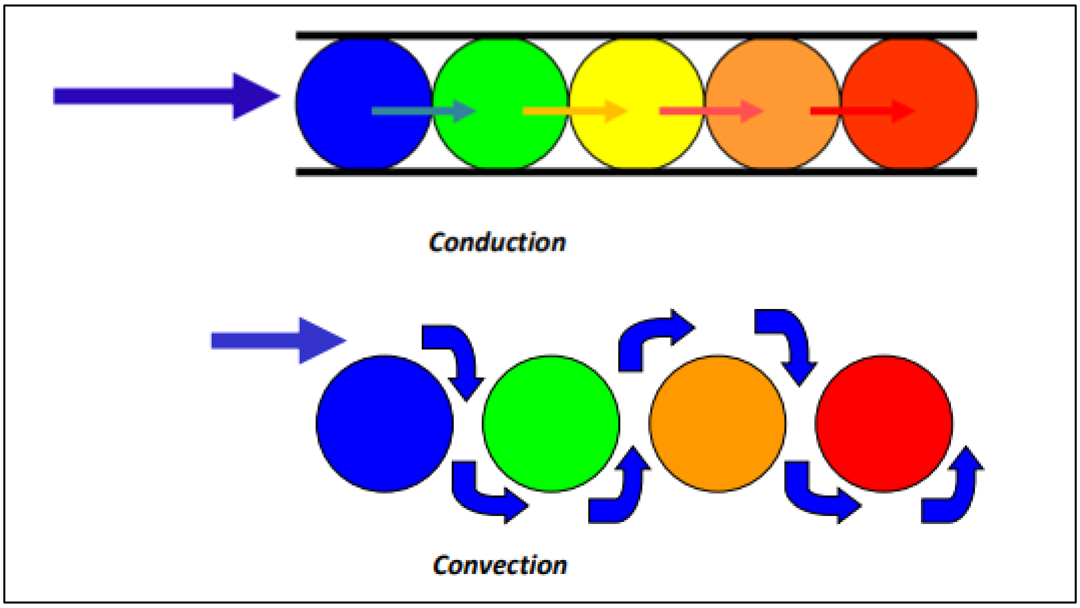


Figure 2.8: Heat transfer principles

Source: Strydom, 2019

FAC uses a method of 'sucking' cold air through the pallet load and cartons within. This method lowers the fruit pulp temperature in two ways (i) through forced convection and (ii) evaporative cooling.

Convection is the transfer of heat by the movement of a fluid (liquid of gas) between areas of different temperatures (Stevens & Fuller, 2015). The process of convection is when cold air meets a hot object, the fruit, resulting in an increase in air temperature and a decrease in fruit temperature. Due to conduction – the passing of thermal energy between materials that are in direct contact – the heat from the fruit is transferred to the air just surrounding it. In cold air, the molecules are more closely packed together than in warm air, so cold air is denser. As new cold air meets the less dense warm air, it sinks below the warm air since it has a greater air pressure – as seen in *Figure 2.9*. Forced convection speeds up the process of heat transfer caused by an external force, in this case a fan.

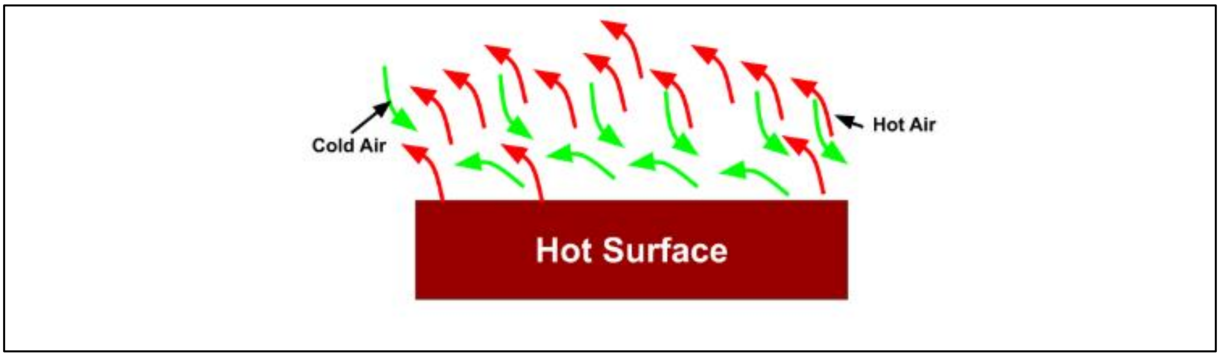


Figure 2.9: Convective air currents

Source: "Convection Heat Transfer – Natural and Forced Convection", 2019

Evaporative cooling occurs when air, that is not too humid, passes over a wet surface; the faster the rate of evaporation, the greater the cooling (Liberty, Okonkwo & Echiegu, 2013). With little or no airflow, the rate of evaporation is low as existing water vapour is not being removed and instead remains stationary. With the introduction of uninterrupted air flow, continuous removal of the existing water vapour occurs, allowing new water vapour, escaping from the fruit, to take its place. The air flow, therefore, enhances the rate of evaporation, reducing the cooling time (Schouten, Lemckert, Parisi, *et al.*, 2011). The downside to the evaporative cooling effect is that it results in moisture loss, which can have a multitude of negative effects on the commodity. The evaporative cooling effect of the FAC process can be minimised while still maximising the convective cooling effect by circulating cold air with a high relative humidity (Holcroft, 2015).

The application of FAC not only increases the rate of cooling but also negates the formation of condensation since air that is pulled through the pallet is colder than the produce, therefore, nullifying the formation of condensation.

2.7.3 Postharvest water relations

Water loss in agricultural products results in weight loss, wilting and shrivelling, while free water or condensation facilitates the growth of pathogens (Holcroft, 2015). At the point of harvest, the produce will be at its maximum water content, with a fresh appearance and crisp texture. Table grapes, for example, consist of 82% water at the point of harvesting (Holcroft, 2015).The harvesting process removes the fruit from its water supply. This is the point when the produce begins to lose weight due to water loss. There are serious economic consequences due to water loss such as the reduction of saleable weight, degradation of the product's appearance caused by wilting and shrivel, resulting in a reduced shelf-life as well as lower profitability.

2.7.3.1 Water loss/ Mass transfer

The process of mass transfer for horticultural products occurs due to the difference in vapour pressure (VPD) between the fruit surface and the surrounding air. This process is known as transpiration.

If transpiration is not decelerated, it induces wilting, loss of firmness, crispness and juiciness, with simultaneous deterioration in appearance, texture and flavour (Gopala Rao, 2015). The factors that affect the rate of transpiration are temperature, relative humidity, atmospheric pressure and the extent of any mechanical damage. Moisture loss is driven by the difference in water vapour pressure between the product surface and the surrounding air (Becker & Fricke, 2014), so to minimise transpiration, fruit should be stored at low temperatures with high relative humidity.

According to Nelson (1978), there are at least three symptoms experienced by table grapes which are directly linked to water loss: (i) shrivelled stems (also known as dry stem), which become brittle; (ii) stem browning, which occurs as dehydration increases (often results in berry shatter or loose berries); (iii) berry softening followed by the formation of wrinkle like markings.

Mass loss occurs due to the diffusion of moisture through the product's surface, which evaporates into the surrounding air (Hatch, 1989). The maximum permissible amount of water loss (the point before the effects of water loss is visible) is about 4–5% loss of total mass when considering table grapes. Mass loss beyond this point tends to result in manifestation of the symptoms listed above (Nelson, 1978).

2.7.3.1.1 Transpiration (mass transfer) coefficient

The mass transfer/transpiration coefficient depends on two factors, namely the air velocity at the surface of the fruit and the fruit skins' resistance to moisture migration (Hatch, 1989) (refer to *Equation 3*).

$$k_i = \frac{1}{\frac{1}{k_a} + \frac{1}{k_s}} \tag{3}$$

Where:

k_i = overall mass transfer coefficient (k value)

k_a = air film mass transfer coefficient

k_s = skin mass transfer coefficient

Equation 3: Mass transfer coefficient

Source: Becker *et al.*, 1996

According to Becker *et al.* (1996), the air film mass transfer coefficient ($\mathrm{k_a}$) describes the convective mass transfer, which occurs at the surface of the fruit depending on the airflow rate. This variable is especially important if different refrigeration methods are applied, such as force air cooling (FAC) or room cooling, where air is circulated through the palletised product to remove heat (Hatch, 1989). This coefficient can also be impacted by the number and size of perforated ventilation holes in the table grape packaging.

The skin mass transfer coefficient ($\mathrm{k_s}$) is based on the proportion of the fruits' skin that is covered by pores, as well as the surface structure and physiological characteristics thereof. The skin of a table grape is covered by a natural wax layer known as the 'bloom', which resists moisture loss. Poor handling of the table grape cluster can cause mechanical damage resulting in stem scars, wounds, surface cracks, bruising and the removal of the bloom layer. The damage to the grape clusters can create physical openings, such as hairline cracks, allowing direct exposure of the fruit to the surrounding air (Hatch, 1989). Increased mechanical damage can, therefore, also have an effect on the skin mass transfer coefficient ($\mathrm{k_s}$).

2.7.3.2 Condensation

Condensation occurs when warm moist air meets a cold surface. Warm air contains more moisture than cold air at the same RH. When warm air meets a cold surface, the air temperature at the surface drops below the dew point causing the air, which is saturated with water vapour, to release the excess moisture in the form of condensation. Condensation, with regard to table grapes, can occur once the fruit pulp temperature has already been brought to the recommended storage temperature of -0.5°C. If the cold chain experiences temperature breaks, whereby the storage temperature increases, residual condensation can occur within the packaging and on the fruit's surface. Due to the recommended table grape storage conditions of -0.5°C at a RH of 95% (Henning & Chetty, 2019), a small temperature fluctuation from these conditions can cause the dew point to be reached, resulting in condensation (Linke & Geyer, 2013).

Freshly harvested table grapes will not absorb moisture from the air even if the RH is close to the dew point, however, table grapes will absorb water if they are immersed in it. Condensation within the packaging's plastic liners can lead to postharvest water absorption, resulting in berry cracking, also known as "bag split" (Fresh Produce Exporters' Forum, 2016). Burger (2000:158) noted that when the storage temperature is raised from -0.5˚C to 10˚C, any time during the cold storage period, condensation is initiated within the bag. Available moisture is absorbed osmotically by the grapes, increasing the turgor pressure within the berries aggravating the risk of berry cracking or split.

During the condensation of water vapour, heat is also released causing increased temperatures at the point of condensation (Linke & Geyer, 2013). The process of condensation consists of two heating properties, namely heat conduction and latent heating. Heat conduction can be described as the movement of heat energy between objects of different temperatures when they come in direct contact with one another. Latent heat is the energy released or absorbed when a phase change occurs without changing the temperature. An example of a phase change involving latent heat can occur when a substance condenses or is vaporised at a specific temperature and pressure. To get water to change to a water vapour, the liquid needs energy in the form of heat to be added. This is done by boiling water, an endothermic reaction. The equivalent amount of heat energy is released when water vapour is condensed.

2.7.4 Respiration

Postharvest, fruit continues to 'live' by performing a chemical process where sugars and oxygen is converted into carbon dioxide, water and heat – this is known as respiration (Becker & Fricke, 2014). The chemical reaction is denoted in *Equation 4*.

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + \ 2667\ kJ \tag{4}$$

Equation 4: Respiration chemical reaction

Source: Becker *et al.*, 1996

Plants require energy to grow in their environment. Through the process of cellular respiration, plants can break down glucose into Adenosine 5'-triphosphate (ATP), which provides energy to carry out various functions. Picked fruit continues this aerobic procedure even though there is a finite supply of glucose, which when exhausted, will cause the metabolic process to stop and lead to decay. Respiration and the metabolic rate are directly related to the storage temperature of the produce, where at lower pulp temperatures the rate of respiration slows, increasing the product's shelf life (lal Basediya, Samuel & Beera, 2013).

As stated, the aerobic process of respiration results in the production of carbon dioxide, water and heat. One can, therefore, determine the respiration rate by either measuring the rate of oxygen uptake or carbon dioxide production. The other by-product generated, heat, increases the temperature of the commodity, which, in turn, increases the VPD and transpiration meaning further moisture loss (Becker *et al.*, 1996). The heat generation W [J/(kg·h)] and carbon dioxide production $\dot{\mathrm{m}}_{\mathrm{CO_2}}$ [mg/(kg·h)] at different pulp temperatures for table grapes is indicated in *Figure 2.10*.

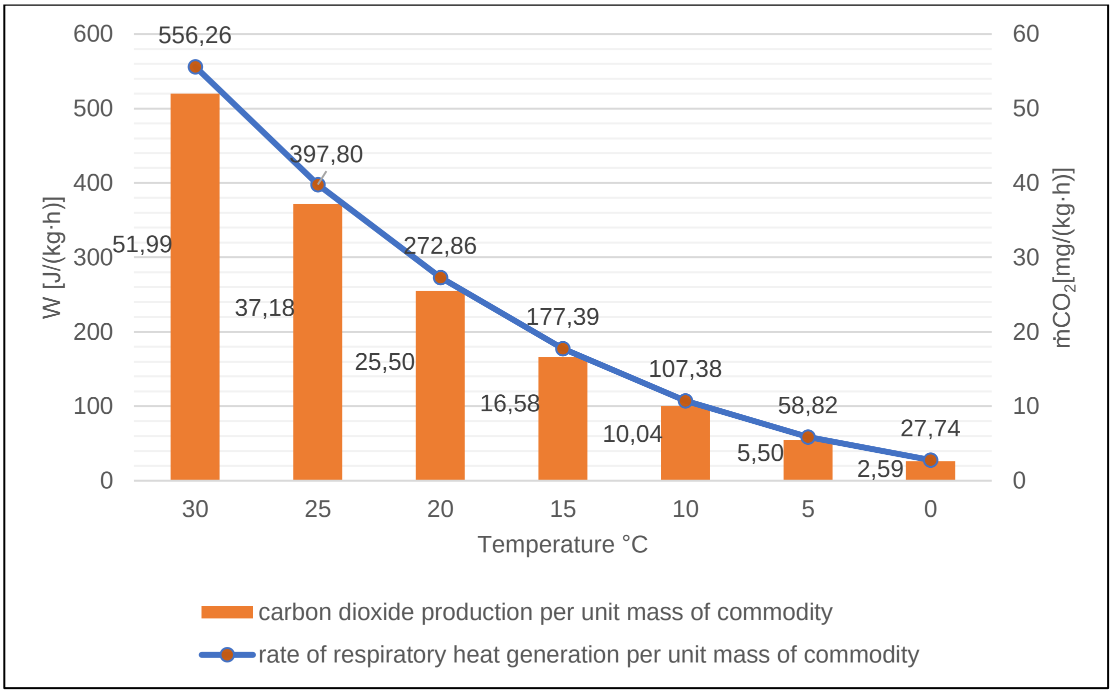


Figure 2.10: Rate of respiratory heat generation and carbon dioxide production per unit mass of table grapes at various pulp temperatures

Source: adapted from Becker et al., 1996

Figure 2.10 indicates that as the pulp temperature is decreased (through cooling), the rate of heat generation and respiration also decreases. Note that for every 10°C decrease in temperature, the rate of respiration, measured by carbon dioxide production, approximately halves. According to Crisosto, Mitcham & Kadar (1998), the rachis, or grape stem, rate of respiration is approximately 15 times higher than that of the berry. The figure indicates the importance of field heat removal and good cold chain practises to reduce the fruits' pulp temperature to prolong the shelf life through delayed senescence and minimised moisture loss.

2.7.5 Effect of mechanical damage on quality

Mechanical damage can be defined as deformation, superficial rupture and destruction of fruit tissue due to external forces (Montero, Schwarz, Santos, *et al.*, 2009). Mechanical damage is often inflicted to fruit during field operations, on grading and packing lines, during transportation and throughout end of line handling where produce is presented by retailers on display (Li & Thomas, 2014).

This improper handling can result in structural, tissue and cell damage to fruit caused by five major damage types. These consist of impact, compression, abrasion, puncture, testing, or several actions combined (Li & Thomas, 2014). The effect of the damage results in structural failure and often increases susceptibility to decay and microorganism growth.

A study conducted by Fischer, Craig, Watada, Douglas and Ashby (1992), simulated in-transit vibration damage to packaged table grapes. The results of the study were determined by the physiological loss in weight (PLW) of the grape clusters. This was accomplished by comparing the initial fresh weight and subsequent loss in weight, resulting in a berry shatter percentage, calculated by dividing the weight of the free berries from the weight of the packed grape cluster (Kaur, Arora, Gill, *et al.*, 2019). Berry shatter, also known as berry drop or berry abscission, is the term used to describe the loss of berries from the cap stem (Burger, 2000). The results of the study concluded that after vibrational treatment had been applied, the berry shatter percentage was 16%, enough to warrant a quality assessment downgrade resulting in subsequent financial loss (Fischer *et al.*, 1992). The study also determined that the effect of vibration on the fruit resulted in berry browning as well as an increase in the fruits' respiration and transpiration rates, resulting in a decreased shelf life.

2.8 Quality standards impacts of food loss and waste

Product quality standards are a tool used by the governing organisations of both importing and exporting countries, to control the quality of fresh produce, formalised in national legislation. The specifications outline the minimum required quality levels, ripeness, taste, smell, safety, shape, skin, uniformity across products within one package and weight (de Hooge, van Dulm & van Trijp, 2018).

South Africa's table grape quality standards are governed by the Department of Agriculture, Forestry and Fisheries, now known as the Department of Agriculture, Land Reform and Rural Development (DALRRD) stipulated under section 4(3)(a)(ii) of the Agricultural Product Standards Act, 1990 (Act No. 119 of 1990), amendment No. 754 of 21 September 2012 (1998-15) (DAFF, 2015). The European Union conversely has its own quality-grade classifications for table grapes (European Union, 2017), which are relatively aligned to DALRRD's requirements. These two legislative bodies enforce marketing standards that govern the export and import of produce from and to the above-mentioned regions.

2.9 The Perishable Products Export Control Board (PPECB)

The Perishable Products Export Control Board (PPECB) inspects all fresh produce to ensure that it conforms to export quality requirements. The organisation was established in 1926 to oversee the shipping of perishable export products, the collection of growers' estimates, and the allocation of pre-cooling and refrigerated shipping space (PPECB, 2003: 16).

Nowadays, the PPECB is constituted and mandated in terms of the Perishable Products Export Control Act (PPEC Act), No 9, of 1983 to perform cold chain services. The PPECB also delivers inspection and food safety services assigned by DALRRD under the APS Act, No.119 of 1990

(*Fresh Food Trade SA*, 2021: 122). DALRRD audits the PPECB at the inspection points, which consist of the packhouses, cold stores, container depots, and ports.

Fresh fruit exported to the European Union (EU) must be inspected on arrival to assess if it conforms with the market's quality and labelling standards before it can be cleared through customs. The EU has certified certain countries, including South Africa, with approved pre-export inspection services. The PPECB is officially recognised by the European Commission (EC) as a certification body, and therefore, has the authority to issue their own conformity certificates, which act as proof that the product meets EU market standards.

In terms of EC Regulation 1148, the PPECB is authorised to inspect and pass fruit destined for the EU markets on behalf of the EU inspectorate. South African fruit entering the EU markets, therefore, does not have to undergo quality inspection on arrival at EU ports of destination if the fruit has been approved by the PPECB, which is indicated with a "passed for export" stamp, regarded as a symbol of quality assurance (*Fresh Food Trade SA*, 2021: 123). Table grapes are one of the fresh fruits that are permitted for pre-export inspections.

Even though there is an independent quality control authority in the form the PPECB, table grapes exported to Europe still arrive in substandard condition resulting in financial losses and food waste.

2.9.1 Cosmetic specifications effect on food waste

Due to the quality standards enforced upon table grapes, there is potential for produce to be rejected due to cosmetic aspects or the products' appearance. The main topic of contention is that due to cosmetic specifications, produce that meets all other quality and safety requirements is wasted, because the shape, colour, weight and size does not meet the 'usual' aesthetic standards (de Hooge *et al.*, 2018).

Cosmetic specifications are relevant in the sense that the fruits' appearance sets the consumers' first impressions of the product when on the shelves. These specifications, therefore, establish consumer expectations of what good quality fruit is and renders consumers unwilling to purchase suboptimal ones due to the product not being perceived as 'good quality' resulting in further retail waste.

The relevance of produce with cosmetic defects on this study is twofold. Firstly, this study looks at the supply chain's effect on waste. Waste that occurs due to packhouses sorting and removing of undesirable berries early in the supply chain will result in food waste, not due to the supply chain, but rather due to cosmetic specifications. Secondly, waste that occurs near the end of the supply chain where cosmetic defects are present is caused by supply chain factors such as poor handling, temperature and humidity deviations.

An example of this is stem browning (rachis browning). Stem browning can be described as the dehydration of the rachis of the grape cluster due to absence of an epidermal cuticular wax, which would act as a barrier against water loss (Kaur *et al.*, 2019). The dehydration results in desiccation (dryness), senescence and tissue softening during postharvest storage. An example can be seen in *Figure 2.11*. The cause of desiccation can be attributed to poor cold storage where the grape clusters are exposed to an uncontrolled ambient environment (Li, Kaplunov, Zutahy, *et al.*, 2015).

Figure 2.11: Stem (rachis) browning

Source: Kaur et al., 2019

Stem browning is applicable to the cosmetic specifications' effect on food waste in that consumers do not eat the stem, but a brown stem has a direct impact on consumer perceptions of freshness and influences the marketability of table grapes.

2.9.2 Postharvest quality factors

The word "quality" is derived from the Latin *qualitas*, meaning – attribute, property, or basic nature of an object (Lopez Camelo, 2004:88). Quality is defined as "the standard of something as measured against other things of a similar kind; the degree of excellence of something". This means that a product of better quality is superior to another in several objective or subjective characteristics.

For table grapes to sell for a good price at global fruit markets, the product needs to possess a certain set of characteristics required by retailers and consumers alike. These include: good appearance (e.g. shape, berry colour, rachis/stem greenness), good texture consistency (e.g. berry firmness), absence of defects (e.g. decay, bruising, skin/flesh browning, cracked/split berries), intact whole bunches (e.g. no loose, dry, or watery berries), and good flavour (Defilippi, Rivera, Preez-Donoso, *et al.*, 2019).

2.9.2.1 Perception of quality

To evaluate fruit quality, one first needs to be familiar with how quality is determined by consumers. Quality is determined by a consumer via simultaneously evaluating a product's characteristics' according to subjective and objective criteria. The process uses four of the human senses – sight, touch, smell, and taste – to compare the product's characteristics to previous experiences with the same product type. For example, sight can be used to determine ripeness according to the product's colour. Touch can further be used to evaluate ripeness by squeezing the fruit to determine its firmness and texture. Finally, taste is used to determine the flavour. Smelling the fruit's aroma is a less commonly used evaluation technique, but is still nonetheless relevant, as fruit with a bad odour immediately indicates over-ripeness and decay. Lopez Camelo (2004:90) developed a consumer perception of quality framework, *Figure 2.12*, which breaks down the consumer's perception of fruit quality into various components – consisting of: appearance/condition, freshness/ripeness, flavour/taste/aroma, nutritive value, and safety.

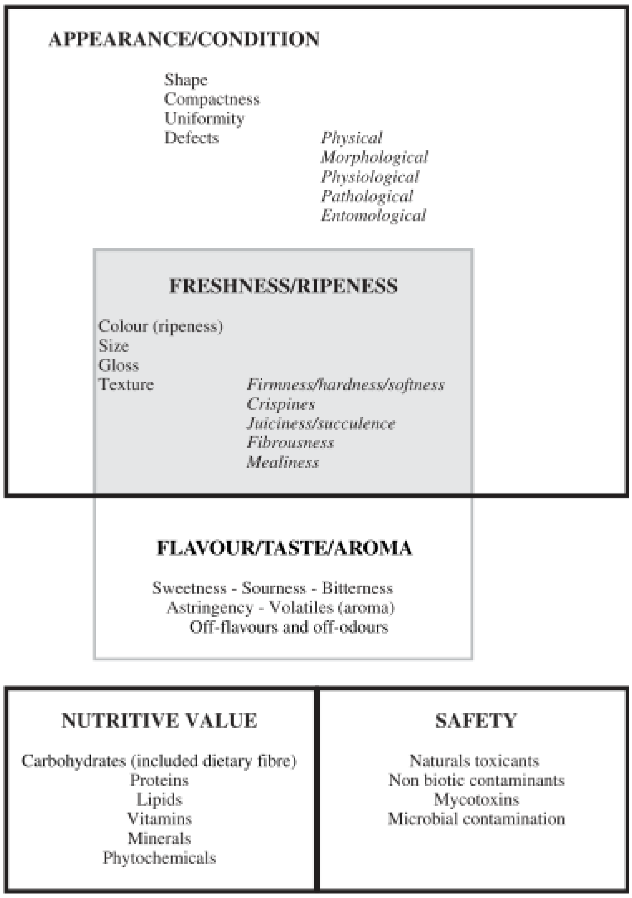


Figure 2.12: Consumer perception of quality

Source: Lopez Camelo, 2004:90

As indicated in *Figure 2.12*, producing a quality product consists of many characteristics, which begin when the vine is first planted. The soil selected, water availability, climatic conditions during the growing phase, pest and disease control, and fertilisation all impact quality at the point of harvest.

A distinction between the term fruit ‘quality’ and ‘condition’ must, therefore, be made although the two are often used interchangeably. Overall fruit quality is an aggregation of the fruits’ various characteristics at the point of harvest, demonstrating various ‘degrees of excellence’. Good quality fruit will be devoid of defects whereas poor quality will have numerous permanent defects. In contrast, the condition of fruit is not permanent as fruit is a living organism constantly undergoing various changes due to maturation and senescence influenced by environmental and storage conditions. The distinction implies that good quality fruit at harvest’s condition can deteriorate resulting in an undesirable final product.

For the sake of this study, fruit quality refers to fruit characteristics that are impacted by postharvest practises – the condition. Characteristics such as size, shape, and colour are excluded as these characteristics are determined by pre-harvest practices and do not change overtime.

2.10 Postharvest pathological and physiological disorders in table grapes

Physiological storage disorders are non-pathological injuries to the produce caused by abnormal changes in cell metabolism. These physiological disorders can impair internal and/or outer appearance of the product, as well as effect flavour, nutritional quality and a reduction in postharvest life and marketability of fruit and vegetables (Tonetto de Freitas & Pareek, 2019). One must also be aware that although many postharvest disorders are caused by postharvest factors, pre-harvest factors also do play a role.

2.10.1 Mycotoxins

Mycotoxins are toxic compounds, which are naturally produced by various types of moulds (fungi). There are numerous kinds of mycotoxins, such as: patulin, fumonisin, aflatoxins, zearalenone, ochratoxin A and nivalenol/deoxynivalenol, which are harmful to humans if consumed (World Health Orgainsation - Fact Sheet, 2018). Consumption of these toxins can result in severe illness in the short-term and have been linked to long-term negative health effects, including cancer and immune deficiency. Exposure to mycotoxins, therefore, needs to be kept to a minimum to reduce the potential health risks associated with the consumption of such toxins.

2.10.2 Postharvest microorganisms and pathological disorders

The major postharvest issues for table grapes are desiccation, bruising and decay (Cappellini, Ceponis & Lightner, 1986). The causes of these postharvest issues can be attributed to pathogens or poor handling techniques throughout the supply chain.

The major pathological decay culprits are *Botrytis cinerea, Cladosporium herbarum, Alternaria* (*Alternaria alternate*), *Rhizopus stolonifer, Aspergillus niger, Penicillium* spp. and Soft tissue break down (STB) (Barkai-Golan, 2001).

Botrytis cinerea (*B. cinerea*) or better known as *Botrytis*, causes a grey mould rot that can have a devastating effect on packaged fruit. It is an airborne pathogen that is ever present in a vineyard, as seen in *Figure 2.13*, but lays latent and only manifests during postharvest storage when conditions are more favourable for growth (constantly wet and humid) (Fresh Produce Exporters' Forum, 2016; Peacock & Smilanick, 1996).

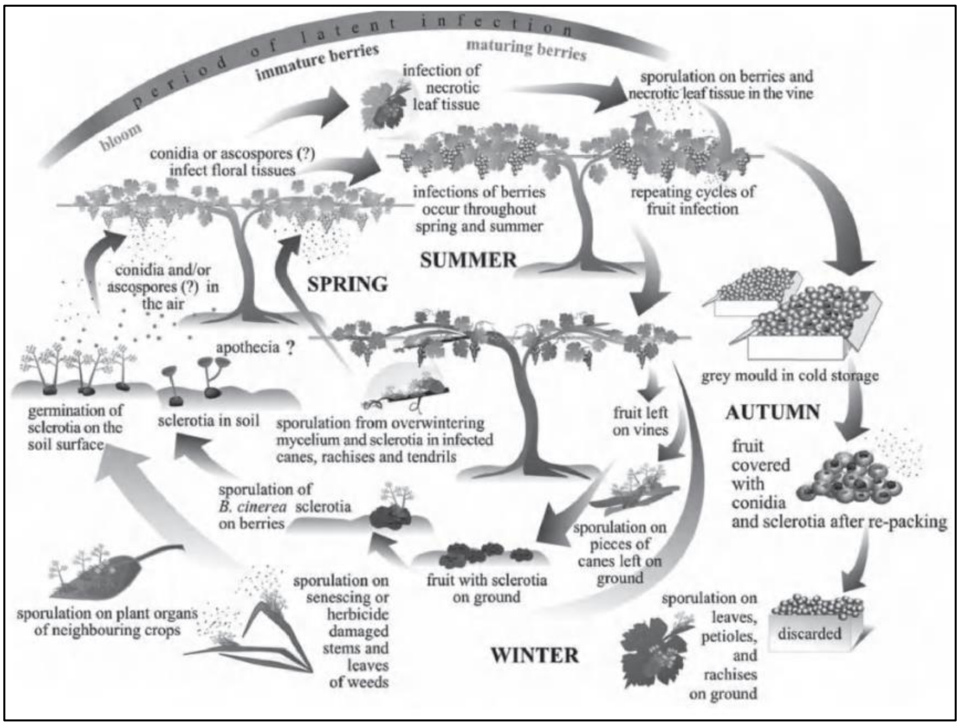


Figure 2.13: Life cycle of Botrytis cinerea and grey mould in wine and table grape vineyards

Source: Elad, Williamson, Tudzynski, et al., 2007:245

The infection takes place when spores are deposited on the surface of the plant material, and then germinate using a root-like mycelial outgrowth (De Visser, Nannes, Van Bokhoven, *et al.*, 2015). The germination of *Botrytis* requires moisture, which usually occurs during a wet growing season resulting a severe *Botrytis* epidemic (Hill, Beresford & Evans, 2019:84).

Botrytis infections occur pre-harvest, across many of the phenological growth stages of the grapevine. Phenology is the study of growth and development rates related to climactic factors that occur periodically. Grape producers use the phenology to keep track of the development stages of the grapevine to timeously apply needed interventions to maximise vine output and fruit quality.

According to the Eichhorn-Lorenz (E-L) system adapted from the BBCHP scale, seen in *Table 2.3*, infection can occur from flowering (ELP 19), through to ripening of berries (ELP 34). The susceptibility of infection increases from *veraison* (berry's start to colour), ELP 35, to harvesting, stage 37-39 (González-Domínguez, Caffi, Ciliberti, *et al.*, 2015). According to Elad *et al.*, 2007:246-252, there are six infection 'path-ways', ranging from a period of early development, through to *veraison* and ripening. Only the latter stages of development are focused on in this thesis, as late-stage infection manifests post-harvest. At these latter stages, infection susceptibility increase as structural and biochemical changes occur during berry maturation (González-Domínguez *et al.*, 2015; Molitor, Baus, Hoffmann, *et al.*, 2016).

The late stage infection of the plant tissue primarily occurs through wounds in the fruit skin (Molitor, Baus, Hoffmann, *et al.*, 2016:232).

Table 2.3: Description of the phenological stages of the grapevine according to Eichhorn and Lorenz, modified by Coombe (1995)

Phenological scale (ELP)	Phenological stage
1	Winter bud
4	Green tip, first leaf tissue visible
7	First leaf separated from shoot tip
12	5 leaves separated; shoots about 10 cm long; inflorescence clear
17	12 leaves separated; inflorescence well developed; single flowers separated
23	17-20 leaves separated; 50% caps off, full bloom
26	Cap-fall complete
27	Setting; young berries enlarging (> 2 mm diam.), bunches at right angles to stem
31	Berries pea-size (7 mm diam.)
32	Beginning of bunch closure, berries touching (if bunches are tight)
33	Berries still hard and green
35	Berries begin to colour and enlarge: veraison
38	Berries harvest-ripe (22°Brix)
39	Berries over-ripe
41	After harvest: cane maturation complete
43	Beginning of leaf fall
47	End of leaf fall

Source: Lorenz et al., 1995; Verdugo-vásquez, Fuente, Ortega-farías, et al., 2017

The decay is made evident with brown discoloration, which later becomes covered with an abundance of grey-brown spores (also known as grey mould), as seen in *Figure 2.14*. If *Botrytis* is left untreated, the spores will multiply on the fruit within the packaging, resulting in shelf-life reduction and potential loss.

Figure 2.14: Botrytis cinerea (grey mould)
Source: Fresh Produce Exporters' Forum, 2016

According to Du Plessis, (1935:4), the first description of *Botrytis* was by Person in 1801, with the fungus occurring on decaying leaves and organic matter.

Cladosporium (*Cladosporium herbarum*) is a form of fungus that can grow at low temperatures, and tend to be slow-growing and form black spots on foods (Barkai-Golan, 2001). *Cladosporium* rot is a form of rot that is initiated in the vineyard and occurs as a minor wound-associated disease in grapes after a long cold storage period (>60 days). The rot is characterised by the development of dark green necrotic circular lesions with a velvety appearance that affect the top part of the berry (Zoffoli & Latorre, 2011). The growth of the *Cladosporium* can be detained at constant temperature management around 0°C, therefore, emphasising the need for constant temperature management and the avoidance of fluctuations.

Alternaria (*Alternaria alternate*) blight is a fungus that attacks berries with damaged tissue, especially the region around the stem cap (pedicel) (Kakalíková, Jankura & Šrobárov́, 2009). According to Swart and Lennox (1995:3), *Alternaria* causes bunch rot, which is characterised by firm, superficial, dark-brown to black lesions on the berries near the pedicels, and fluffy grey tufts of fungus growing on the rachises (stems) and pedicels as seen in *Figure 2.15*. *Alternaria* is an opportunistic pathogen that gains entry to the plant via wounds but does not spread from berry to berry.

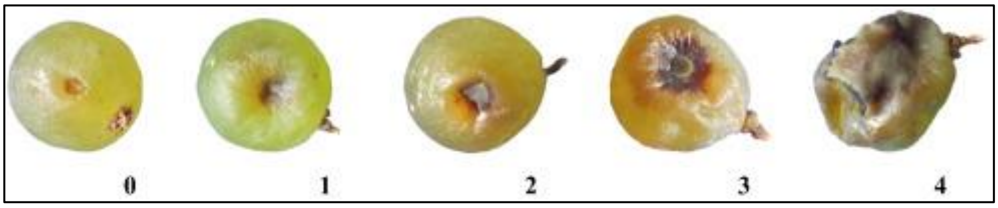


Figure 2.15: Withering process of Alternaria infection in grape berries
Source: Lorenzini & Zapparoli, 2014

Rhizopus (*Rhizopus* spp.) rot is a fungus that produces no visible symptoms on the berries. When slight pressure is applied to the infected berry, it explodes, turning into a watery mess (Fresh Produce Exporters' Forum, 2016). The infection continues to spread throughout storage by means of juice from infected berries coming in contact with adjacent healthy berries (Barkai-Golan, 2001). In some instances, *Rhizopus* is visible when the fungus creates a cobweb-like white mycelia with numerous sporangia (white when young, later turning black) (see *Figure 2.16*). Optimal growth development temperatures are between 20-25°C, and thus, can be suppressed through cold storage. However, growth will resume when the fruit is transferred to ambient shelf conditions.

Figure 2.16: Rhizopus rot

Source: Ministry of Electronics and Information Technology: Government of India, 2019

Aspergillus (Aspergillus niger) or black mould rot, is a postharvest disease that effects berries that have suffered injuries caused by poor postharvest handling techniques (Shankar, Sharma, Raj Boina, *et al.*, 2014). The pulp of infected berries is reduced to a watery consistency and results in berries looking black in colour (see *Figure 2.17*) (Fresh Produce Exporters' Forum, 2016). The source of the fungus is from spores often found in the air or soil in the vineyards and can in the packaging facilities if proper sanitation practices are not routinely implemented. Since the fungus does not grow at a temperature above 5°C, good cold storage techniques should be an effective fungus suppressant as long as there are no serious cold chain breaks (Barkai-Golan, 2001).

Figure 2.17: Black mould Aspergillus

Source: Shankar et al., 2014

Penicillium (Penicillium spp.) or better known as blue mould rot, is a blue-green fungus that only attacks damaged tissue or wounded berries. It infects the entire berry and turns it into a watery mass (see *Figure 2.18*). Infection can occur in injured berries by conidia, dispersed by air currents, insects and the wind. *Penicillium* is unlike many of the other forms of fungus because it continues to cause decay at a temperature of 0°C, albeit at a slower rate (Barkai-Golan, 2001). The current blue mould control method is to use SO_2-generating pads, which have been ineffective (Witbooi, Fourie & Taylor, 2010). The downside to this control method is that red varieties, such as Red Globe, are sensitive to high SO_2 concentrations, which can paradoxically promote berry injury and may lead to soft internal breakdown. This internal breakdown can in turn favour *Penicillium* infection. The present incidence of *Penicillium* is low (<2%), however, importing countries market tolerance is (<0.5%), therefore, resulting in potential financial implications for growers such as price reduction or rejection (Franck, Latorre, Torres, *et al.*, 2005).

Figure 2.18: Blue mould (Penicillium)
Source: Shankar et al., 2014

Soft tissue breakdown (STB) also referred to as 'Melting decay' is characterised by severe maceration of grape berries flesh tissue. STB symptoms in storage manifest in cracking and dissolution of the epidermis (Morgan & Michailides, 2004), followed by sunken decayed areas, usually with absence of superficial growth (see *Figure 2.19 A & B*).

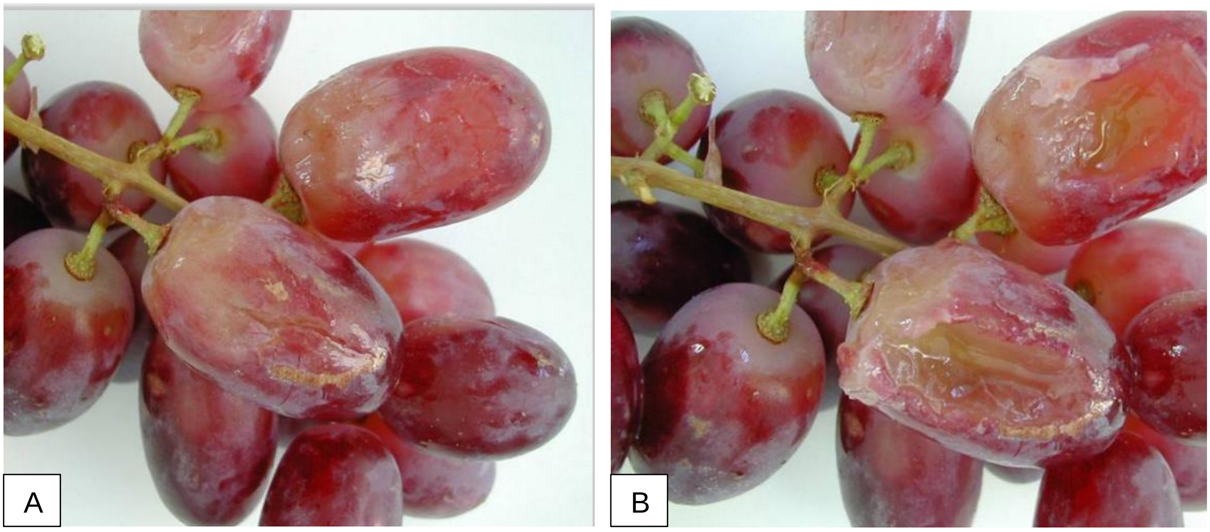


Figure 2.19: Soft tissue breakdown on crimson seedless – a) epidermis cracks; b) sunken/macerated tissue
Source: Witbooi et al., 2010

The cause of STB, according to Witbooi *et al.* (2010:45), is instigated by a conglomoration of various pathogens, comprising of *Rhizopus stolonifer, Aspergillus niger, Penicillium* spp., as well as other yeasts and bacteria, although playing a smaller role. From a postharvest decay control perspective SO_2 does not regulate the formation of STB, but instead contributes to the problem. STB develops on berry wound sites caused by either mechanical damage in the harvesting process or from split berries, caused by insufficient packaging perforation ventilation. There is currently no satisfactory postharvest commercial control method against the potential formation of STB. The only current guidelines are to minimise cold chain breaks to avoid the formation of condensation within the cartons.

2.10.3 Physiological disorders

One of the most pervasive physiological disorders ailing soft fruits is cracking of the skin and splitting of the underlying flesh. The disorder can present itself both pre- and post-harvest resulting in significant commercial losses by reducing the quality and yield.

Swift, May and Lawton (1974) defined the difference between berry cracking and splitting in grapes. Cracks are a fine superficial fracture of the fruit skin, whereas splitting is an extreme form of cracking where cracks penetrate deep into the flesh.

The rate at which turgor pressure builds up within the grape berry skin determines if it will crack or split. Turgor develops when the protoplasts that absorb water osmotically cease expanding due to the relatively inflexible cell wall matrix (Burger, 2000:31). The epidermal tissue (skin) limits berry expansion and if the turgor pressure is high, the epidermis will rupture. When skin cells rupture, the internal flesh of the grape is exposed where the fungal pathogen, such as *Botrytis* can then infect the fruit and cause decay manifesting postharvest.

It has been observed that table grapes susceptibility to berry cracking varies depending on the variety. By measuring the amount of cell turgor required to crack the skin of mature berries, it was observed that 50% of berries cracked under 15 atmospheres turgor for vulnerable varieties and 40 atmospheres turgor for resistant varieties (Beede, n.d.). It was also noted mature fruit with a sufficient sugar content could absorb enough water to crack the skin under the right environmental conditions. These conditions were high soil moisture and low water use by the vine.

The mechanism that leads to the disorder can occur through two avenues:

1) Direct absorption through the grape's skin. When moisture occurs on the berry surface, water diffuses through the skin into the interior of the fruit. The rapid influx of moisture into the berry results in cracking. The moisture source could be raindrops falling directly on the ripe grape bunches or by condensation forming on the berries within the vine canopy or within the packaging liner postharvest.

2) The pre-harvest uptake of moisture by the vine, either from the soil through the root system or to a lesser extent through the leaves. The environmental factor causing preharvest splitting are a sudden significant increase in soil moisture due to rain or late irrigation, prior to complete maturity, exacerbated by a period of drought in the growth phase (Burger, 2000:23).

Beede (n.d.:2) noted that different table grape varieties have different levels of susceptibility to berry cracking.

2.11 Pre-harvest climatic conditions' effect on pathological and physiological disorders postharvest

Botrytis cinerea, mentioned above, is one of the most destructive fungal pathogens affecting table grapes, annually causing severe economic losses (Peacock & Smilanick, 1996).

In terms of the preharvest grape lifecycle, young immature berries have been found to be highly resistant to *Botrytis*. However, these defences tend to wane as maturation progresses, as natural occurring fungistatic substances within the grape cluster decline and micro-cracks form as the cuticle deposits covering the berry skin decrease, beyond *veraison. Botrytis* infection is, therefore, more common in ripe berries (Molitor *et al.*, 2016). A common perception amongst table grape producers is that rain in the weeks prior to harvest result in more severe post-harvest decay compared to if there had been dry conditions over that same period (Hill, Henshall & Beresford, 2017). This insight was first noted by Du Plessis (1935:70), in his doctoral research on the wastage of export grapes, where grapes picked shortly after rainfall had an abnormally high infection percentage. This perception agrees with the literature, which states that the fungus penetrates wounds, or cracks/splits, in the berry's skin. The spores develop in the cracks and then spread across the entire berry. *Botrytis* is known as a necrotrophic fungus: it first kills its host plant and then colonises the dead tissue (Romanazzi & Feliziani, 2014). This pathogen can quickly spread through an entire vineyard as airborne spores travel from bunch to bunch.

The climatic conditions within which most pathogens can grow occur within the temperature range of 0˚C to 30˚C; optimally at 20˚C as well as, conditions with high relative humidity and free moisture. For pathogen germination to occur, spores need free moisture, nutrients and a certain duration within a given temperature range. For example, at temperatures between 18˚C and 24˚C, only two hours of free water are needed for germination to occur ("Bunch Rot Part 1: Botrytis cinerea", 2021). At temperatures greater and lower, however, the time taken for germination is extended.

Ciliberti *et al.*, (2015:1094) found that infection is highly influenced by relative humidity (RH): where at low levels of RH (65% RH), no disease occurred. It was also found that with wounded berries, disease incidence is much higher. This is probably due to the open wounds on the berry surface acting as a food source for the fungus to grow, allowing for rapid germination. Berry-to-berry infection is favoured by high relative humidity (Ciliberti *et al.*, 2015).

2.12 Best practices to reduce waste

Control of postharvest decay is best achieved through an integrated approach which incorporates some or all of the management interventions listed below.

2.12.1 Mitigation practices for suboptimal pre-harvest climatic conditions

Variable atmospheric conditions such as increased temperature and wind enhance evaporation post-rain, diminishing the extent of damage caused. Without any wind, existing water vapour, caused by the rain, is not efficiently removed, and instead remains stationary in the vicinity of the water body. Continuous airflow removes existing water vapour allowing newly escaping water vapour from the residual rain droplets to take its place, speeding up the rate of evaporation (Schouten *et al.*, 2011).

A precautionary measure that some farmers have used when heavy rainfall is forecast, is the application of plastic canopy covers (Fidelibus, Vasquez & Kaan Kurtural, 2016). These are fixed covers which are fastened along each row's trellising. The covers require a lot of labour for installation and need to be removed again once the rainy period has passed.

There is a view held amongst producers that the use of plastic covers is ineffective against moderate and stronger rainfall. In the event of a precipitation rate above 2.5 mm per hour, the vine's vascular transportation system forces the newly absorbed water from the roots to the ripe berries causing 'cracking' or splits. This view is supported by the research of Børve, Skaar, Sekse, Meland & Vangdal (2003), which found that fruit under protective covers, still suffer from cracks and splitting due to the creation of a vapour pressure deficit underneath the covering causing a reduction in transpiration. This being said, the study found that the beneficial effect of the covers outweighed the reduction in other quality factors.

There are new technologies entering the market such as the SOLARIG® GR Pro crop covers. The SOLARIG® GR Pro is a permanent crop protection cover that is retractable, visible in *Figure 2.20*. This allows for fast erection if rain/hail is forecast but is equally as quickly collapsed, post-rain, allowing the wind into the canopy, improving evaporation.

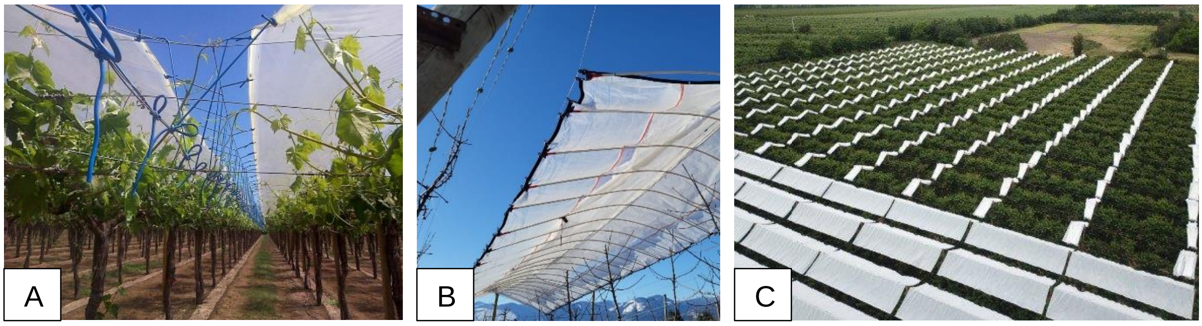


Figure 2.20: Retractable rain protective covers for fruit trees

Source: Bunyard, 2020

The permeability of the vineyard soil profile would play a role in how effectively the product would perform since the fallen rain would runoff the covers and collect in the middle of the rows. For example, coarse sandy soil with high granularity would allow for rapid soil drainage, giving the vine less of a chance to absorb the rainwater through the root system, lowering the likelihood of 'cracking' and 'splitting' of ripe grape berries. This is currently an untested product in the South African market, so the effectiveness is yet unknown.

It is commonplace for farmers to irrigate vineyards with mature fruit prior to rainfall to incrementally saturate the vine so that the precipitation drains into the soil and is not rapidly absorbed. This strategy is backed up by Bowcher, Clarke, Rogiers and Hardie (2012:6), who stated that soil moisture plays a role in reducing splitting as well-watered plants fruit do not increase in volume as much or as quickly as berries on water-stressed plants. Alternatively, a slow volume change may provide sufficient time to allow the skin to accommodate to an increase in berry volume without splitting. The effectiveness of this mitigation strategy still needs to be studied further.

2.12.2 Chemical application

Chemical control is the most effective strategy to reduce the incidence of grey mould in grapevines. According to Brink, Holz and Fourie (2006: 51), preventative fungicide should be applied three times during the growing period: "firstly, between budding and pre-bloom, to protect susceptible inflorescences (cluster of flowers); secondly, during bloom to pea-size stage, to reduce inoculum in the bunches and to prevent colonisation of floral debris; and thirdly, at bunch closure, to reduce inoculum of *B. cinerea* at various positions of the inner bunch, especially for cultivars are tightly clustered".

To effectively control pathogens, sufficient deposition and quantity of the fungicide applied to the susceptible plant tissue is essential, while not exceeding the acceptable residue level requirements for the various global markets (Brink, Calitz & Fourie, 2016). It has also, however, been noted that the timing of spray programmes to growing stages are less effective than spraying according to the environmental conditions at play. Fungicide applications before rain are more effective in reducing *Botrytis* infections than those applied after rain ("Bunch Rot Part 1: Botrytis cinerea", 2021).

2.12.3 Surface sanitation

Sanitation is of vital importance to control microbial contamination of fruits (De Simone, Pace, Grieco, Chimienti, Tyibilika, Santoro, Capozzi, Colelli, Spano and Russo, 2020). From a table grape postharvest perspective, the point of harvest to packaging poses high infection risk. Good vineyard and packhouse sanitation measures are of high importance to reduce the risk of microbial infection.

Vineyard sanitation practises include cleaning and decontaminating harvest trays and pickers trimming scissors prior to picking. Once a bunch has been picked, decayed berries should be clipped and discarded in a waste bin to reduce cross contamination between bunches in the harvest trays (lugs) (Fourie, 2008: 422). This ensures that fruit entering the packhouse is sanitary.

Packhouse sanitation requires all personnel to sanitise their hands and clipping scissors frequently. The cleaning and sanitation of work surfaces according to a fixed schedule with an approved chemical that won't leave residues is essential (Jansen, 2021).

Vineyard sanitation measures extend to the post season period, where any grape bunches not harvested or had fallen to the ground can be a source of inoculum the following spring. it is, therefore, of vital importance to either remove those bunches from the vineyard or to place them in the middle of the rows to be incorporated into the soil as mulch.

2.12.4 Fumigation chambers

Fumigation is the process of releasing and dispersing a toxic chemical in a gaseous state to reach and kill the target organism. Sulphur dioxide (SO_2) is a widely used postharvest fumigation agent for decay prevention of table grapes during cold storage. The current South African model makes use of slow-release SO_2 generating pads, which are placed within the packaging (Tessara, 2022), however, it has been argued that only applying this preventative measure is a little too late.

Other table grape production regions globally, make use of SO_2 chambers to fumigate the picked grapes prior to packaging. This allows for a sanitary product to enter the packhouse. A study funded by the Agricultural and Processed Food Products Export Development Authority (APEDA) of India, evaluated the effectiveness of industrial scale SO_2 chambers to prevent invasive pests, such as fruit fly, to meet certain export market phytosanitary requirements. The study found 100% mortality of adult fruit flies, but maggots or larvae were not destroyed (Vishwakarma, Bashir, Kumar, Yadav, Sharma & Lohakare, 2022).

The fumigation process (*Appendix C*) is implemented postharvest, prior to the fruit entering the packhouse. *Figure 2.21* illustrates the structure of a theoretical fumigation chamber, which can handle an industrial quantity of table grapes in harvest field trays.

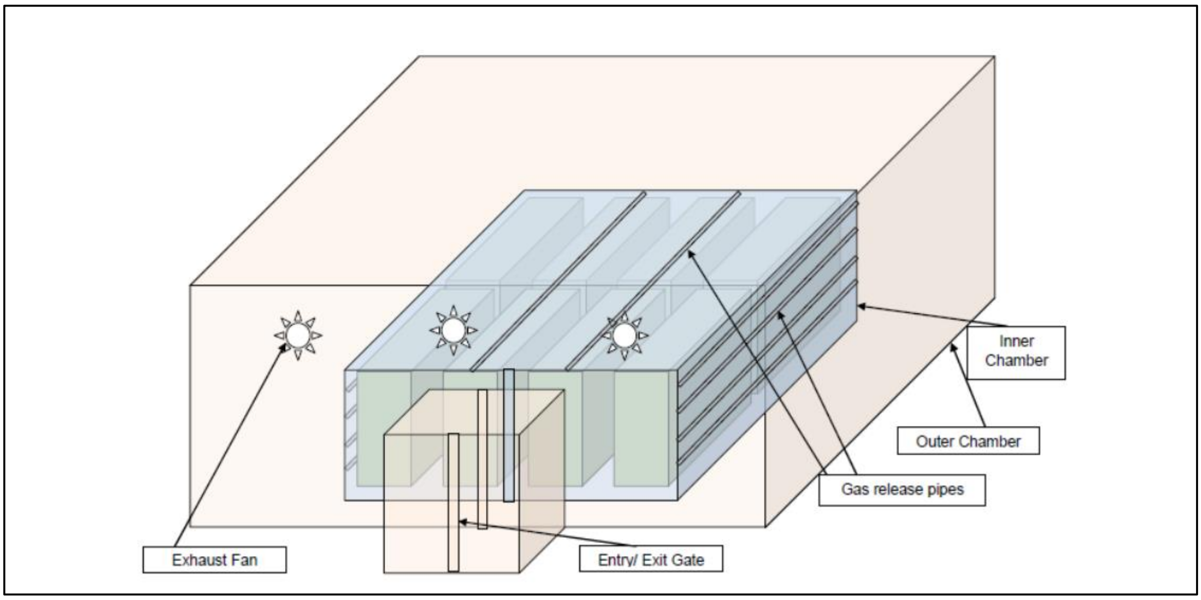


Figure 2.21: Isometric view of conceptual fumigation chamber

Source: Vishwakarma, Bashir, Kumar, et al., 2022

The entire fumigation treatment takes roughly thirty minutes and is, therefore, non-disruptive to the current workflow. For best results, clipping and trimming of rotten berries should be done prior to fumigation to keep the packhouse as sanitary as possible.

2.12.5 *Botrytis* pressure

As stated in *section 2.10.2*, infection can occur from flowering, through to ripening of berries. It is, therefore, of vital importance that producers apply the optimal fungicide spray regimes to minimise infection and postharvest loss. A best practice to forecast the decay potential of a specific vineyard block was established by Harvey (1955). The forecasting method requires 500 berries to be sampled from the specific block, surface sterilised, incubated for ten days at room temperature and high humidity, and then evaluated for *Botrytis* decay. The forecasting method at harvest was found to be highly correlated with the level of decay that developed during cold storage.

The incubation period takes roughly 10 days, whence early signs of *Botrytis* start showing. *Botrytis* is identified by the separation of the berry skin and tissue (slip skin) evolving into a velvety grey mould covering the entire berry. For that study, the forecast of decay threshold was no greater than 0.5%, roughly 2 berries out of the 500 sampled, otherwise the consignment was flagged for caution and extended periods of cold storage were not recommended.

2.13 Conclusion

The literature review illustrates that the fruit industry is a vital contributor to South Africa's economy. This being said, the product is highly perishable, with many potential factors leading to poor arrival quality at the end of the export orientated cold chain. For the industry to remain internationally competitive, factors that negatively impact the arrival condition need to be reduced. This could potentially be achieved by implementing the best practises suggested at the end of the chapter, increasing the proportion of sound quality arrivals for the European market, boosting grower returns while reducing food waste.

Chapter 3: Literature Review - Machine Learning

According to Burkov (2019:3), machine learning (ML) can be defined as the process of solving practical problems by 1) gathering a dataset, and 2) algorithmically building statistical models based on that dataset.

The General steps and decisions involved in a machine learning application are collecting data, preparing the data, choosing the models, training the models, tuning the model, evaluating the models, and making predictions. These steps are discussed in this chapter.

3.1 Machine learning basics

Some basic ML related concepts first need to be discussed to establish the context in which ML is applied to this study. These consist of the broad ML approaches available as well as how different levels of model complexity affect the prediction performance.

3.1.1 Machine learning approaches

ML approaches are divided into three broad categories consisting of supervised, unsupervised, and semi-supervised models based upon the learning process of the algorithm.

- **Supervised learning** is applied to datasets where for each observation there is an associated response, where the learning process aims to produce a model that accurately predicts the response of future unseen observations (Russell & Norvig, 2021: 1205).
- **Unsupervised learning** is used when the outputs (response) are not known or are not available at the time of model training, instead the approach tries to find informative patterns embedded in the input dataset (James, Witten, Hastie, Tibshirani, 2013).
- **Semi-supervised learning** consists of a dataset with both labelled and unlabelled observations, rather than consisting of an equal number of inputs and outputs. The unlabelled observations are kept, improving the learning algorithm (Burkov, 2019: 4) compared to only opting for an unsupervised approach.

This study focuses on supervised learning techniques as the data collected consisted of a collection of input and output pairs.

3.1.2 Bias-variance trade-off

The bias-variance trade-off is central to machine learning. A researcher typically wants to build as accurate a model as possible on the training data but one that can also make general predictions on new unseen data. This results in what is known as the bias-variance trade-off.

Bias refers to how well the machine learning algorithm captures the true relationship between the features and the response. If the model makes many misclassifications i.e., shows poor prediction performance, the model is said to have 'high bias', or that the model underfits.

Reasons for underfitting might be that the model is too simple for the prediction task's dataset (for example, linear models often underfit) or the features used for prediction are not informative enough of the dependant variable. The solution to the problem of underfitting is the application of more complex models or to select features with better predictive power. *Figure 3.1* shows that with increasing model complexity, the degree of fitting becomes more pronounced, demonstrated by different models being applied to the same dataset.

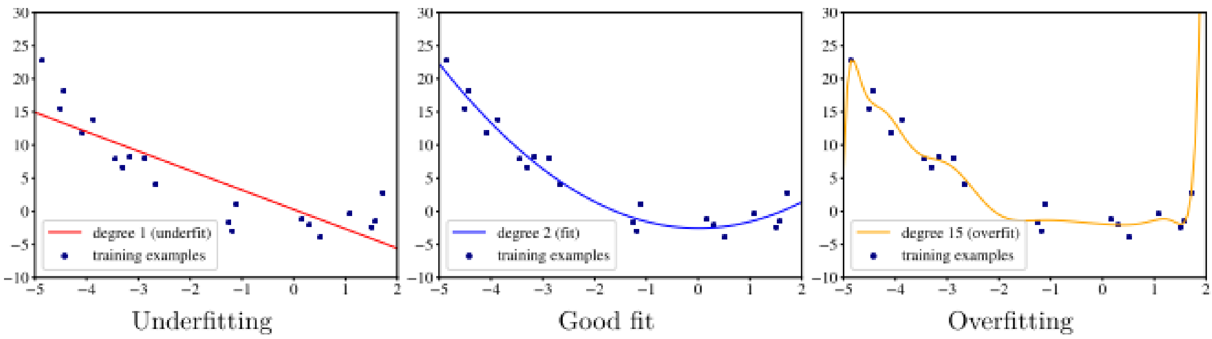


Figure 3.1: Examples of underfitting (linear model), good fit (quadratic model), and overfitting (polynomial of degree 15)

Source: Burkov, 2019: 11

Variance is concerned with the difference of model fits between the training set and the test set. A model with high variance is one that has been overfitted to the training data and follows the errors, or noise, too closely. The implications of this are that the model performs very well on the training data but performs poorly on unseen validation or test data - observations that weren't used in the algorithms learning process. Overfitted models are said to have 'high variance' due to the model's sensitivity to small fluctuations in the training set. The implications of high variance are that if the training set was resampled and the model was retrained, the structure of the new model would be different in comparison to the original model. Overfitted models, therefore, perform poorly on new data.

Reasons for overfitting may be that the model is too complex for the dataset or that the ratio of feature variables to the number of training observations is too large.

The relationship between bias-variance and the effect of model complexity is illustrated in *Figure 3.2*, which demonstrates how simple models suffer from bias resulting in underfitted models whereas models that are too complex for the dataset result in increased variance resulting in model overfitting.

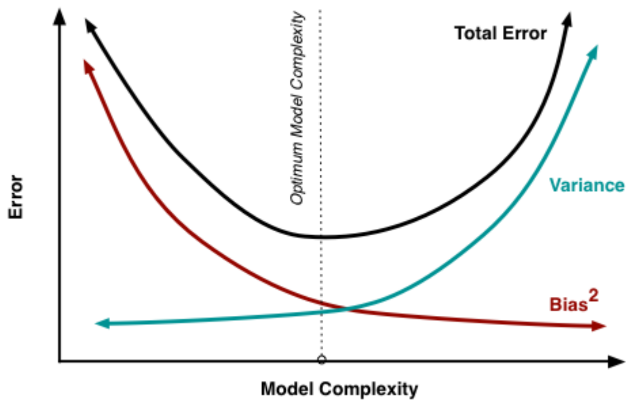


Figure 3.2: The bias-variance trade-off and the effect of model complexity

Source: Fortmann-Roe, 2012

3.2 Trade-off between prediction and inference

As discussed in the previous section, the complexity of machine learning models vary, which impacts the flexibility or shape of the model's function. Less complex models tend to be less flexible or restrictive, whereas more complex models tend to be more flexible.

There is a trade-off to be made between the level of model flexibility and the model interpretability. As flexibility increases, model interpretability tends to decrease. This impacts researchers' model selection decisions based on the research question trying to be solved.

Statistical learning questions can be categorised into two groups: predictions or inferences.

Prediction questions are concerned with choosing a model that most accurately represents the relationship between the features and the decision variable to accurately predict future output based on new inputs. Prediction questions are not concerned with the form of the model if the model provides the most accurate predictions. Purely prediction-based models are often referred to as 'black-box' models, where the models tend to be more complex and cannot be easily interpreted.

In contrast, inference questions are concerned with understanding the relationship between the features and the decision variable, how the output changes as a function of the input. Highly complex black-box models do not perform well for inference-based questions because the form of the algorithm is unknown. Less complex, simple restrictive models tend to be more interpretable, aiding in answering questions concerned with the inference paradigm.

It must be noted, however, that continually increasing model complexity in search of selecting a model that most accurately predicts the decision variable has the potential for overfitting, which in turn would result in a reduction of model accuracy.

3.3 Regression vs. classification tasks

Variables can be considered as either quantitative or qualitative (categorical). Quantitative variables consist of numerical values, whereas qualitative variables are non-numerical and instead take on values in one of *K* different classes, or categories (James, Witten, Hastie, *et al.*, 2013: 28). Numerical response values are regression problems whereas qualitative response values are classification problems. There are, however, exceptions to this such as logistic regression, which is used on qualitative two-class (binary) responses.

Classification problems try to predict a discrete class-label whereas a regression task predicts a continuous integer value. The tasks also differ in how the prediction accuracy is evaluated, there are numerous techniques available but are, however, beyond the scope of this study.

Classification algorithms build what is known as a decision boundary, a boundary separating the examples of different classes. The form of the boundary can be straight, curved, or more complex. *Figure 3.3*, is example of a decision boundary for two classes represented by the dashed line. The shaded areas indicate the region where test observations will be assigned depending on either of the two classes. The colour of the circles represents the actual class of the observations.

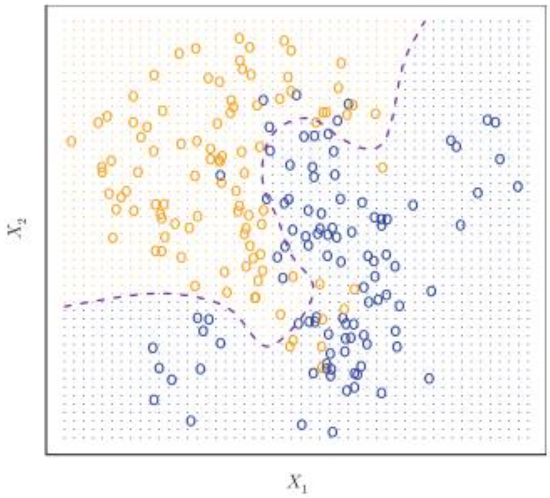


Figure 3.3: Decision boundary example

Source: James et al., 2013: 38

The decision boundary is important because it determines model accuracy, the ratio of examples whose labels are predicted correctly. The form of the decision boundary, computed algorithmically / mathematically based on the training data set, is what differentiates classification learning algorithms from one another (Burkov, 2019: 8).

The study is a classification task since the dependant variable, arrival QC score of table grapes, is an ordinal variable.

3.3.1 Types of supervised learning

Supervised learning approaches can either be model-based or instance-based. Model-based learning algorithms use the training data to create a model that has learned parameters, which are used to make predictions on new input data. Instance-based learning algorithms on the other hand, use the entire dataset as the model (Burkov, 2019: 13). The algorithms try to generalise a new case based on similar cases used in model training by means of a similarity measure. Model-based approaches used in the study consist of Multinominal Logistic Regression, Decision Trees, and Random Forests. Instance-based models consisted of k-Nearest Neighbours.

3.4 Models

Performing machine learning involves selecting a model that is trained on training data, consisting of input and response pairs, which can then process new input data to make output predictions.

3.4.1 Multinominal Logistic Regression

Logistic regression is a classification model used to predict the probability of an outcome (from two alternatives) by having the log-odds for the event be a linear combination of one or more independent variables (predictors) (James *et al.*, 2013). Multi-class logistic regression is a classification method that generalises logistic regression to a multiclass problem (>2 outcomes). Multinominal logistic regression uses a linear combination of observed features and problem specific parameters to estimate the probability of a particular value of the dependant variable. The optimal values for the parameters are determined from a training data set.

3.4.2 k-Nearest Neighbours

k-Nearest Neighbours' algorithm (k-NN) is a supervised learning method used for both classification and regression problems. It is a non-parametric statistical model, implying that the model does not assume that the sample data can be demonstrated by a probability distribution that follows a fixed set of parameters (e.g., normal distribution).

k is a user defined constant; a test-point is classified by assigning the label that is most frequent among the k training samples nearest to that point. The 'nearness' is measured by a metric known as the Euclidean distance; a line drawn between two points, in space, calculated by using the Pythagorean theorem (James *et al.*, 2013: 39). k-NN classification for multi-class predictors uses a plurality vote of the training labels to predict the new test label.

This can be explained visually by referring to *Figure 3.4*, where the new test-point, represented by the green circle, needs to be classified either as a blue square or a red triangle. If k=3 (observations within the solid-line circle), the three 'nearest' training-points to the test-point (green circle) are used for classification and the green circle is, therefore, assigned to be a red triangle. However, if k=5 (observations within the dotted-line circle), then the green circle will be assigned as a blue square, since that is the most frequent label of the five 'nearest' neighbours.

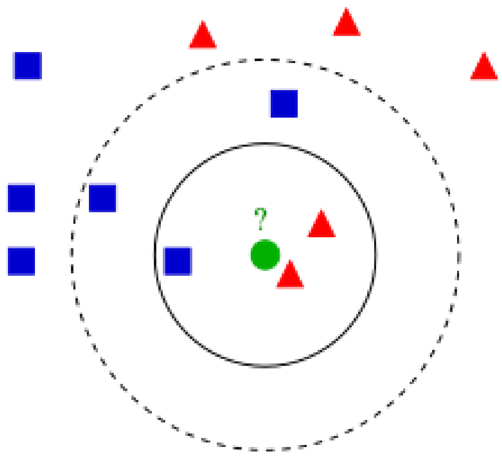


Figure 3.4: K-NN example

Source: Srivastava, 2014

k-NN is a good model for non-linear functions with many observations but does not perform as well on small datasets with many outliers. The optimal value of k is typically chosen through cross-validation, discussed in *section 3.7.*

3.4.3 Decision Trees

Decision trees (DT) are a supervised learning method used in machine learning. The approach can solve both regression and classification problems, referred to as either regression trees or classification trees. Again, classification trees are applied to data with a categorical decision variable.

The structure of a classification tree consists of the dataset recursively being split into smaller subsets according to the feature cardinalities until pure leaf nodes are left. This is when additional splitting no longer adds value to the prediction score (Pieterse, 2022: 72).

A tree analogy is used to describe the structure of a DT and consists of the following four components:

- **Root –** first node on the tree, consisting of the source set, which is split into subsets
- **Internal node** (split node / branch attachment) – points along the tree where the predictor space splits, creating a fork, based on a set of splitting rules.

- **Branch** – segment that connects the internal node to the terminal node.
- **Terminal node** (leaf node) – The final subset of a classification tree, determines the predicted class-label of the new observation.

The tree is displayed upside down and starts at the root decision node and ends at the terminal / leaf nodes. The leaf nodes are the endpoint of a branch, the final output of a series of decisions. In the classification setting, the terminal nodes (leaf) display the predicted class-labels of the categorical decision variable. The internal nodes display splitting decisions made according to the features. *Figure 3.5* demonstrates the structure of a typical decision tree with the elements labelled.

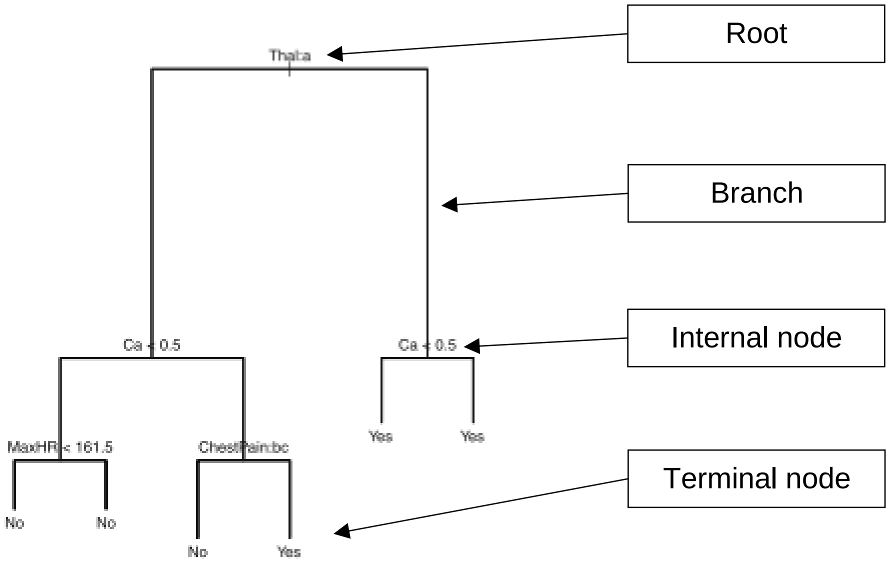


Figure 3.5: Example of a Classification Tree

Source: adapted from James et al., 2013: 313

The process of building a classification tree utilises a top-down approach making use of recursive binary splitting to successively split the predictor space, each split indicated by two new branches further down the tree. The cut-point splits the predictor space and can be determined by either using the classification error rate, Gini index or entropy. These techniques measure node purity - where the node contains predominantly observations for a single class (James *et al.*, 2013: 311).

The Gini index was used in this study for recursive binary splitting to assign observations to their most commonly occurring class. The Gini index is defined as “the measure of total variance across the number of response classes” (James *et al.*, 2013: 312), which measures node purity.

There are, however, shortcomings to the DT model, such as the suboptimal accuracy created by the recursive binary splitting, instability in the tree formation when adding additional training data and the expansive computational run time needed on deeper trees (Russell & Norvig, 2021). These issues are resolved in the Random Forest model discussed in *section 3.4.4*

3.4.4 Random Forests

Random Forests are what is known as an ensemble model, a collection of hypotheses that combine their predictions by averaging, voting, or by another level of machine learning (Russell & Norvig, 2021). Ensemble models allow for a reduction in bias and variance often present in more restrictive individual hypotheses - base models. The result is a model that is less likely to misclassify observations since the ensemble is made up of multiple binary classifiers trained on different subsets of the data, which are then aggregated using a majority voting system to classify a test point more accurately.

Decision trees tend to overfit on the training data due to the learned hierarchy used in the splitting process, and therefore, tend to underperform in prediction accuracy when compared to the Random Forest model.

A Random Forest model enforces variation due to the way it de-correlates the individual DTs within the ensemble. This is achieved by varying which features are chosen when making the split considerations of each DT by randomly sampling a subgroup of features and applying those to each DT (James *et al.*, 2013: 319). Each split is then only allowed to use one feature from the sample group. A new random sample of features are drawn at each split across the entire ensemble of DTs. The number of sample features drawn for classification problems is usually $\sqrt{n}$, if there are n features for the random forest model (Russell & Norvig, 2021).

This process avoids only selecting highly correlated features for each DT within the forest, therefore, de-correlating the trees avoiding overfitting of the model on the training data set.

3.5 Feature / variable selection

Often, not all the variables in a dataset are associated with predicting the response variable, therefore, the variables/ features need evaluating to select those that best describe the output variable. This process involves keeping features that improve model performance in favour of those that do not. The overarching feature selection method chosen, depends on whether the researcher has already considered the underlying learning model to be applied. Filter methods are performed using scores independent of the learning model, whereas wrapper and embedded methods are dependent on the model selected (Beraha, Metelli, Papini, Tirinzoni & Restelli, 2019). This study only made use of filter methods for feature selection.

3.5.1 Filter methods (Univariate statistics)

Filter methods find the intrinsic properties of the features, measured using univariate statistics instead of cross-validation . Feature evaluation is conducted outside of the model, with the subsequent model only consisting of those features that meet certain criterion (Kuhn & Johnson, 2013: 490). *Figure 3.6*, summarises which univariate statistical evaluation techniques should be applied depending on the variable relationship in question. The statistical evaluation techniques circled in *Figure 3.6,* are those that are applicable to this study.

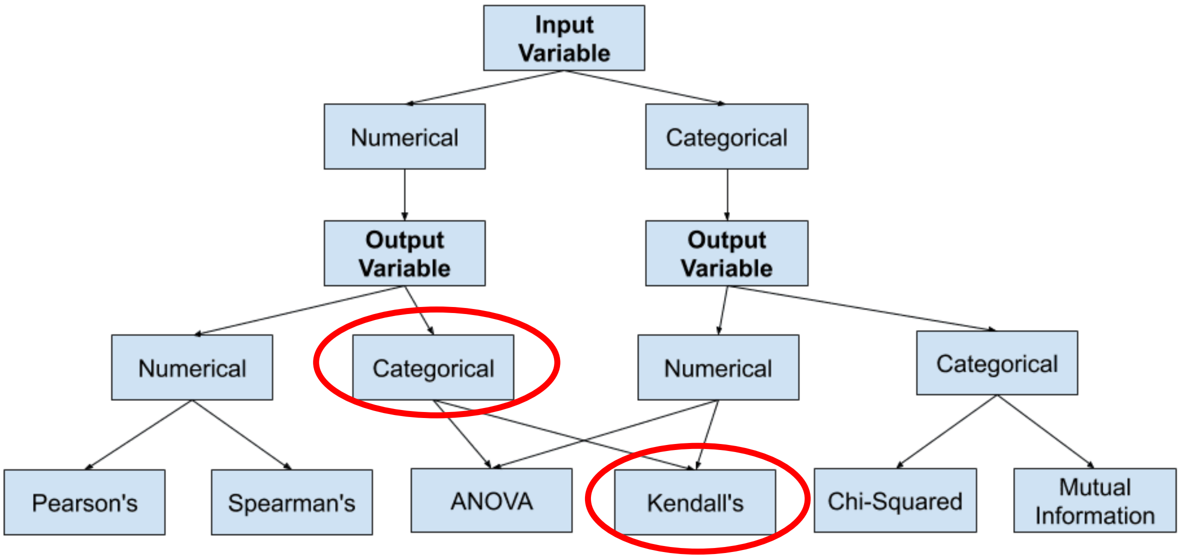


Figure 3.6: Univariate statistical techniques used for filter-based feature selection

Source: Brownlee, 2019

Correlation Coefficient

Correlation measures the linear relationship between two variables. The values of a correlation coefficient range from -1 to +1, where +1 indicate a strong positive correlation, -1 a strong negative correlation, and 0 when uncorrelated. Good features tend to be highly correlated, positively or negatively, to the target variable and should be uncorrelated in relation to other features. Adding two features that are correlated to one another adds additional noise to the model. There are three methods to check the linear relationship of two variables, namely Pearson's, Spearman's, and Kendall's correlation techniques (Ye, 2020).

The choice between which method to use when calculating the correlation between variables depends on a set of assumptions. Pearson's correlation is used when the data for the variable is 1) parametric (normally distributed), 2) follows a linear relationship, 3) the data type is continuous (interval or ratio), and 4) there are minimal outliers. Spearman's correlation relies on all the same assumptions except that it works on non-parametric (non-normally distributed) data as well as categorical (ordinal variables). Both Pearson's and Spearman's correlations measure the degree of the association between the linearly related variables.

Kendall's rank correlation is a non-parametric test that measures the strength of dependence between two variables, ordinal or continuous, commonly used when the sample size is small and has many tied ranks. It is used to test the similarities in the ordering of data when it is ranked by quantities based on pairs of observations, determining the strength of association.

Other filter feature selection techniques not used in the study consist of univariate statistical methods such as Chi-squared test of independence and ANOVA (analysis of variance) as well as Mutual information (information gain), a feature variable indexing method.

3.6 Evaluation methods (Model performance Assessment)

Performance metrics are a way to evaluate the predicted outputs of the model. The evaluation scores expose characteristics and capabilities of the model, which can be used to make improvements as well as provide insights into how the model performs for this context.

3.6.1 Feature importance

As discussed in *section 3.2*, as model complexity increases the interpretability decreases. This is true when comparing DTs to RFs, which are an ensemble of multiple DTs, resulting in a loss of interpretation.

Although RFs are difficult to interpret, an overall summary of the importance of each feature can be created by determining how much the performance of the RF would decrease if a specific feature was removed or rendered useless. If the change has a large effect on the model's performance, then the variable is important.

This process is implemented using the Gini index (discussed in *section 3.4.3*) for classification tasks, a measure of node purity. The feature importance is determined through the total amount that the Gini index is decreased by over a given feature, averaged across number of trees in the forest (James *et al.*, 2013: 319).

3.6.2 Confusion matrix

The performance evaluation of a classification model is based on the count of test records correctly or incorrectly predicted by the model, often displayed in a table known as a confusion matrix (Tan, Steinbach, Kumar, Karpatne, 2018).*Table 3.1* is an example of a confusion matrix where the number of observations for each predicted class (1 and 0) are counted. Correct prediction for each class is recorded in f_{11} and f_{00}, whereas false predictions are recorded in f_{10} and f_{01}.

Table 3.1: Example of a confusion matrix for a two-class problem

		Predicted Class	
		$Class = 1$	$Class = 0$
Actual Class	$Class = 1$	f_{11}	f_{10}
	$Class = 0$	f_{01}	f_{00}

Source: Tan et al., 2018: 149

For binary classification, the confusion matrix is summarised as follows:

True positives (TP): Predicted positive and are actually positive (f_{11}).

False positives (FP): Predicted positive and are actually negative (f_{01}).

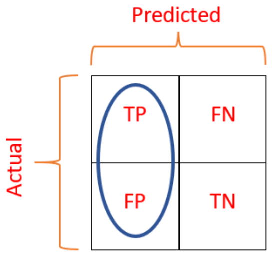
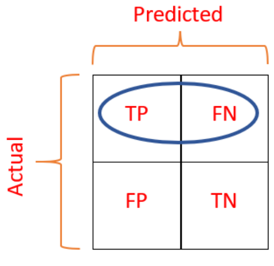
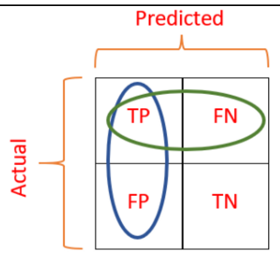
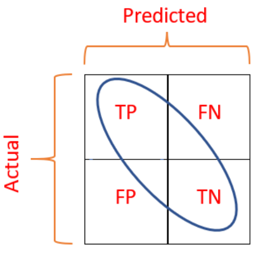
True negatives (TN): Predicted negative and are actually negative (f_{00}).

False negatives (FN): Predicted negative and are actually positive (f_{10}).

3.6.3 Performance metrics

A Summary of the equations, based off the confusion matrix, can be found in *Table 3.2*.

Table 3.2:Statistical analysis techniques for binary classification

precision	recall	F1 score	accuracy
			
$\frac{TP}{TP + FP}$	$\frac{TP}{TP + FN}$	$2 \times \frac{(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$	$\frac{TP + TN}{TP + FP + TN + FN}$

Source: adapted from Tuychiev, 2021

For multi-class classification problems all model performance metrics can be applied by treating the problem as a collection of binary classification problems by averaging the scores across the number of classes. The average scores are weighted based on the balance of observations for each class.

1. **Accuracy score**

   Accuracy score summarises the confusion matrix results into a single metric displaying the prediction accuracy. This is merely the number of correct predictions divided by the total number of predictions. Equivalently, the error rate can be used to determine the number of wrong predictions by instead using the number of incorrect predictions as the numerator.

The accuracy score metric is, however, flawed when the number of observations per class is imbalanced. In medical research, for example, the number of cancer patients observed tends to be low within the entire population. Of a sample of 100 patients, 10 might have cancer. If all 10 patients are undetected (false negative) the accuracy score would still be 90%, even though the model is clearly not performing accurately.

1. **Precision and recall scores**

The precision score is a ratio of the number of correct positive predictions to the overall number of positive predictions. The recall score is the ratio of correct positive predictions to the overall number of positive examples in the test set (Burkov, 2019: 15). The recall score is used to measure the occurrence of false negative predictions.

It is not possible to have both high precision and recall scores simultaneously, therefore, researchers tend to optimise the model in favour of one score over the other. Optimisation can be achieved in a number of ways such as by assigning higher weights to certain classes, hyperparameter tuning during model training, or by varying the decision threshold for the algorithm (Burkov, 2019; Tuychiev, 2021).

2. **F1 score**

As previously discussed, there is a trade-off between the precision (false positives) and recall scores (false negatives), which provide insights into false predictions (Murugan, 2022: 24). The F1 score is the harmonic mean of the precision and recall scores, ranging from 0 to 1. The metric indicates how well the model performs by focusing on all false predictions. The metric minimises false positive and false negatives in imbalanced classes and is, therefore, a preferred metric for datasets with imbalanced classes.

3.6.4 ROC curve

Two additional performance measures that can be derived from the confusion matrix are sensitivity and specificity. Model sensitivity is the rate at which the class of interest is predicted correctly for all samples (the proportion of positive observations predicted correctly) also known as the true positive rate (TPR). Sensitivity is calculated according to *Equation 5* (Tharwat, 2018).

$$sensitivity\ (TPR) = \frac{TP}{TP + FN} \tag{5}$$

Equation 5: Sensitivity

Source: Burkov, 2019: 17

Specificity is the converse statistic to sensitivity, described as the true negative rate (TNR). 1 – specificity is equal to the false positive rate (FPR), which is the proportion of negative observations that are incorrectly classified, calculated according to *Equation 6*.

$$1 - specificity = (FPR) = \frac{FP}{FP + TN} \tag{6}$$

Where:

TP = Predicted positive and are actually positive

FN = Predicted negative and are actually positive

Equation 6: 1 - specificity

Source: Burkov, 2019: 17

The receiver operating characteristic (ROC) curve is a graphical plot summarising the trade-off between the true positive rate (TPR) and the false positive rate (FPR) presenting a range of probable predictions obtained by varying the decision threshold. One of the key advantages of the ROC curve over the confusion matrix is that it is insensitive to skew class distribution and misclassification cost (Trung Thanh, Thi Anh Dao, Linh-Trung & Vu Le, 2017).

The decision threshold indicates where all values greater than the threshold are mapped to one class and all other values below are mapped to another class. Setting different thresholds changes the sensitivity and specificity of the model. This allows the researcher to tune the behaviour of the model based on the type of error more important to the classification task.

The two types of error are *false positives* (FP), predicting the event when there was no event or *false negative* (FN), predicting no event when there was in fact an event. The ROC curve can be used to find the threshold that balances the two concerns, improving the model's performance. *Figure 3.7*, illustrates the effect of changing the decision threshold and the effect that has on the two types of error.

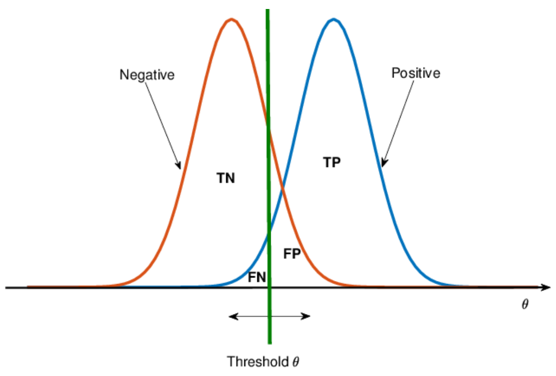


Figure 3.7: Graphical illustration of how varying the decision threshold affects the confusion matrix results

Source: Trung Thanh et al., 2017

The ROC curve is a two-dimensional graph that represents the TPR on the y-axis and the FPR on the x-axis. A set of different thresholds are used to interpret the changing TPR and FPR predictions for the positive class, and the scores are plotted in a line of increasing threshold to create a curve (Brownlee, 2021).

There are four important points along the ROC curve, illustrated in *Figure 3.8*.

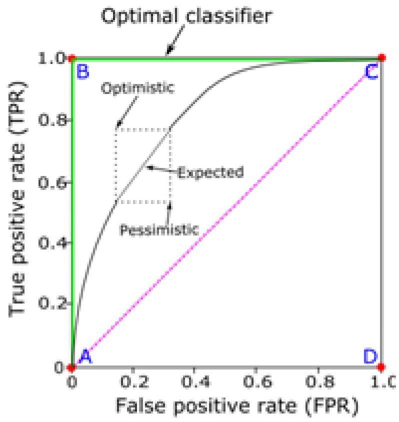


Figure 3.8: The ROC curve indicating the important points, the optimal ROC curve (green), and the no classifier line (pink)

Source: Tharwat, 2018

- Point A, in the lower left corner, (0, 0) represents a classifier where there are no positive classifications, while all negative classifications are correctly classified, TPR = 0 and FPR =0.
- Point C, in the top right corner, (1, 1) represents a classifier where all positive samples are correctly classified, while the negative samples are misclassified.
- Point D, in the bottom right corner, (1, 0), represents a classifier where all positive and negative observations are misclassified.
- Point B, in the top left corner, (0, 1), represents a classifier where all positive and negative observations are correctly classified. This point represents the perfect classification, also known as the ideal operating point.

The perfect ROC curve is represented in green, which rises vertically from (0, 0) to (0, 1) then horizontally to (1, 1). This curve perfectly ranks all positive observations relative to negative observations. The pink dashed line, plotted diagonally from point A to C, represents the 'random classifier' a model that is no better than chance.

3.6.4.1 Area under the curve (AUC)

The metric used to evaluate the ROC curve is the area under the ROC curve (AUC), an aggregate measure of performance across all possible classification thresholds. The higher the AUC, the better the classifier is at predicting the positive class. The AUC score is bound between 0 and 1. When the AUC = 1, the classifier is perfectly able to distinguish between positive and negative classes. A score where AUC = 0.5 signifies a classifier that is no better than chance, the random classifier. The AUC score effectively shows the researcher how well the model distinguishes between the different classes.

3.6.4.2 The AUC-ROC for multi-class classification

The ROC curve and AUC score are binary classification evaluation techniques, where there are only two classes. For multi-class classification tasks, such as this study, there are two methods available to apply the evaluation technique. There are the one-vs-rest (OVR) and one-vs-one (OVO) methods. The OVR method is applied in the study.

The OVR method compares each class individually against the rest. This is achieved by considering one class as the ‘positive class’, while the rest are considered as the ‘negative class’. This reduces the multiclass output into a binary classification. The process is repeated for the remaining classes, resulting in an ROC curve and AUC score for each class. A weighted average can then be used to develop an overall ROC curve an AUC score for the multi-class model.

3.6.5 Black-box models

As stated in *section 3.2*, complex black-box models selected for prediction accuracy suffer from limited interpretability. There are, however, a handful of techniques available devised to improve complex model’s interpretability, allowing for inferences to be made about the relationship between features and the output variable. Techniques consist of partial dependence plots (PD plots), individual conditional expectation (ICE) plots, as well as other agnostic techniques, which can be applied to a host of supervised learning methods. PD plots were used to draw inferences from the complex random forest model built in this study.

3.6.5.1 Partial dependence plots (PD plots)

PD plots help visualise the average partial relationship between the predicted response and one or more features (Goldstein, Kapelner, Bleich & Pitkin, 2015). PD plot the change in the average predicted values as specified feature(s) vary over their marginal distribution.

Simple 1-way PD plots, show how the model’s predictions depend on a single feature over a range of observations. At each value for the specific feature, the model is evaluated for all observations of the remaining features used to build the model, and the output is then averaged.

For multi-class classification tasks, such as the one investigated in this study, PD plots are created for each output class (e.g., green, amber, and red) where the probability of class adherence is displayed on the y-axis, compared to a range of observed values for the feature subset, displayed on the x-axis. This allows the researcher to see the type and direction of the relationship between a feature and a predicted class-label. Some relationships might be linear whereas others may be more complex, positively, or negatively associated.

There can, however, be pitfalls to PD plots if there is a strong correlation between the chosen feature and the remaining features. This can obscure the complexity of the modelled relationship.

3.7 Hyperparameter tuning

A hyperparameter is a value used to control the learning process of a model external to the training process. Models have default hyperparameters, which need to be optimised based on the dataset to control for over / underfitting – the model's flexibility (Burkov, 2019: 20). Overfitting occurs when the model adapts too much to the training dataset, losing predictive accuracy on new unseen data, whereas underfitting is the reverse.

The hyperparameters must be "tuned" by experimentally searching for the combination of values, which yield the best model. The tuning process can be applied to a validation set, a portion of the dataset that is set aside and not used in the training or testing procedures. For small datasets, portioning the data into three (train, validation, and test) sets would reduce the number of samples used in the model's learning process. A technique called cross-validation can instead be utilised, which splits the data into two sets, a train and test set, the validation set is no longer needed.

cross-validation is a resampling method that uses different portions of the training set to train and validate the model over multiple iterations. A round of k-fold cross-validation consists of partitioning the data into equally sized subsets, known as folds, which are then used to train and validate the model. To reduce variability, multiple rounds of cross-validation are performed using different folds (James *et al.*, 2013). The validation results are averaged over the rounds resulting in a more accurate estimate.

The tuning process traditionally employs a grid search method, an exhaustive searching process based upon a manually specified subset of hyperparameters (Zheng, 2015). A model for each set of hyperparameters is generated, trained, and tested on the cross-validation folds. The combination of hyperparameters that yields the highest score is then kept for testing the model. The process is illustrated in *Figure 3.9*, where the dataset is split into training and testing sets, followed by cross-validation , hyperparameter tuning, and model validation. Each model's performance is evaluated and the one that performs best is retained for testing and final evaluation.

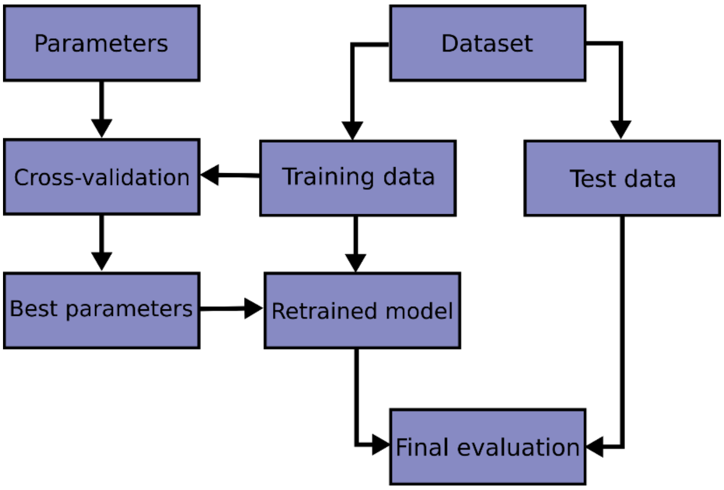


Figure 3.9: Model tuning and testing process

Source: Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Verplas, Passos, Cournapeau, Brucher, Perrot & Duchesnay, 2011

3.8 Conclusion

The machine learning literature review explored the different types of classification models based on the level of complexity, inference, and prediction accuracy required for the classification task. Various feature selection techniques are discussed and are to be used in combination with domain knowledge regarding table grapes. Model evaluation and tuning techniques are discussed, which aid the research in selecting the best performing model.

The modelling process undertaken is described in *chapter 6* and explains the workflow applied to build the classification models. The four models trained are evaluated in *chapter 7* with the best performing model being retained for evaluation.

Chapter 4: Research Design and Methodology

After identifying the research problem, the researcher determined what the appropriate method(s) were to follow when approaching the problem. This study adopted the 'research onion' (Saunders, Lewis & Thornhill, 2019) to provide structure to the research process (refer to *Figure 4.1*).

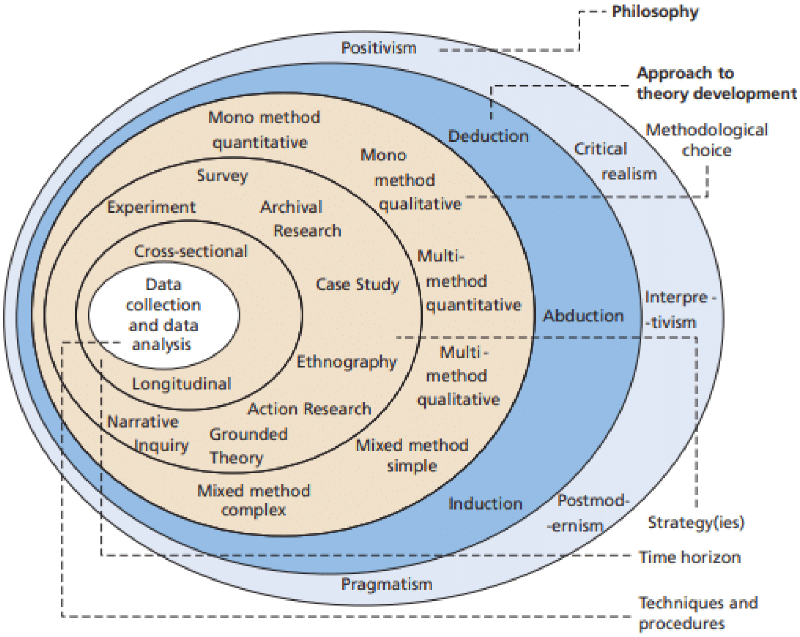


Figure 4.1: Research Onion diagram

Source: Saunders et al., 2019: 130

The research onion aids in providing a succinct summary of the important considerations required when undertaking research. The onion comprises of various layers, which funnel the research from the broad considerations to the specific. *Figure 4.2* gives an indication of the various categories that are relevant to this study throughout the research process.

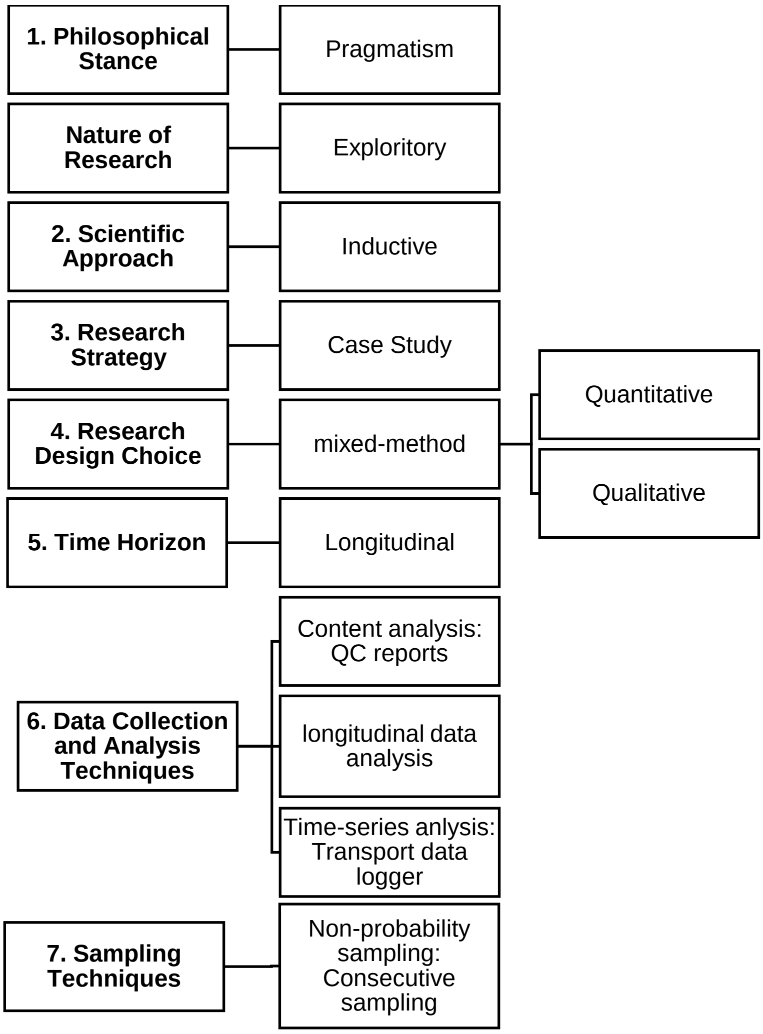


Figure 4.2: Categories identified for study research process

Source: adapted from Saunders et al., 2019

4.1.1 Philosophical stance

A paradigm is a theoretical framework made up of a set of basic beliefs, which define how the world works that guide the research. Paradigms influence the way knowledge is studied and interpreted within the specific discipline (Weaver, 2018). The pragmatist paradigm deals with facts that claim that the philosophical stance of the research is determined by the research problem (Žukauskas, Vveinhardt & Andriukaitienė, 2018). In addition, according to Alghamdi and Li (2013:2), pragmatism is a research stance that is not committed to any single system of philosophy and reality, as according to positivists, the world is ever changing as reality is actively created as individuals act in the world, therefore, giving positivist researchers freedom of choice. Pragmatism embraces a form of neutralism (the idea that philosophy is not prior to science but continues with it), which therefore, allows pragmatists to embrace a mixed method research stance (Weaver, 2018).

The pragmatism paradigm is applicable to this study as pragmatism rejects the idea that social enquiry using a single scientific method can uncover truths regarding the 'real world'. This study, therefore, used multiple data collection and analysis techniques to delineate truths about the research problem. This study's research design is mixed method in nature as quantitative temperature and climatic data was collected in collaboration with qualitative QC reports and nominal logistic data to establish the causes of current poor arrival quality of table grapes.

4.1.2 Exploratory research

This study followed an inductive research approach, where research guides the generation of new theories. According to Zikmund, Babin, Carr, *et al.*, (2013: 44) inductive research is when the logical process of creating a general proposition on the basis of observation of particular facts. This approach was chosen since there were no existing methodologies available to guide the study. *Figure 4.3* provides a summary of research design flow guiding this study. The figure aids in screening the alternative research avenues that were considered when determining the most applicable research methodology.

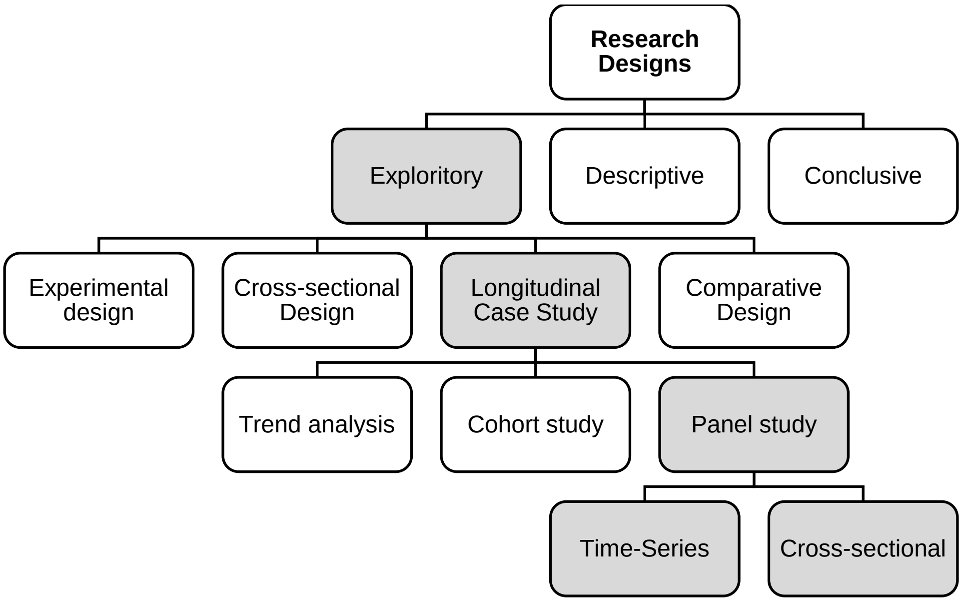


Figure 4.3: Research design flow

Source: Adapted from Saunders et al., 2019

4.1.3 Research Design

The research design is the plan that provides the structure that guides the researcher to address the research question(s) and research objective(s) (Deforge, 2010).

This study follows a case study research strategy. According to Bryman and Bell (2015), case study research consists of a detailed specific analysis of a single case where the study is concerned with the complexity and nature of the particular case in question. Put more simply, it involves examining a phenomenon within its specific context. This study is case specific since it analyses a South African table grape supply chain where the data gathered came from one company. The implications of a case study research strategy are that the results will have limited generalisability or *low external validity* indicating that results will have no wider inferences beyond the context of this study.

A mixed method research approach was utilised. According to Johnson, Onwuegbuzie and Turner (2007:120), mixed method research can be defined as "…a class of research where the researcher mixes and combines quantitative and qualitative research methods, approaches, concepts or language into a single study or set of related studies". There are various forms of mixed method research to distinguish between corresponding literature delineating the typology of each form. This study follows the three criteria laid out by Bentahar and Cameron (2015:8) in an article looking at various mixed method approaches. The three criteria consist of: 1) time dimension (the order of the methods), 2) emphasis dimension (dominant or equal methods status) and 3) mixing dimension (partially or fully mixed methods). (Refer to *Appendix D*).

This study followed a partially mixed sequential dominant status design (refer to P4 in *Appendix D*). Therefore, this study follows a strategy where quantitative data is the main data collection approach, but qualitative data collection preceded quantitative collection (qual → QUAN). Qualitative data acted as a preparation for the main phase, which is quantitative in nature (Bryman & Bell, 2015).

4.1.4 Time horizon

The time horizon for this study is longitudinal in nature as the data pertains to table grape quality collected at two different points in time. According to Bryman and Bell (2015:43-44), longitudinal design is used to map change and allows some insight into the time order of variables since data is collected on the same sample at least twice. This study can, therefore, be described as a longitudinal case because it attempts to evaluate how the quality of table grapes, supplied by one organisation, changed along the supply chain.

Panel data consists of observations of numerous phenomena, collected over several time periods for the same sample group (Lavrakas, 2008). This could be generated by pooling time-series observations across a variety of cross-sectional units.

The distinction between a cross-sectional panel data, also known as longitudinal data (Frees, 2004:2) and time series analysis is made clear when two principles are considered: 1) the number of cases measured $\boldsymbol{n}$ (the sample size) and 2) the number of occasions those cases are measured $\boldsymbol{k}$. A longitudinal data will typically have high $\boldsymbol{n}$ and low $\boldsymbol{k}$, whereas time series studies usually have low $\boldsymbol{n}$ and high $\boldsymbol{k}$.

The time horizon for this study is one harvest season where the sample size ($\boldsymbol{n}$) is made up of all shipping containers sent to the Europe by a single exporter. The sample consists of QC reports, collected at two nodes in the supply chain, and the transit temperature data for each container shipped. Time-series data analysis is applied to temperature data captured in ten-minute intervals throughout the shipping leg of the supply chain.

4.2 Data sources, types, and collection techniques

Data collection for this study was broken into two broad categories consisting of primary and secondary data sources. Primary data sources are defined as "first-hand accounts of information by the researcher, which is categorised as original research" (Primary vs. Secondary Sources, 2018). Secondary research, on the contrary, is described as information that was originally presented in another source, where the researcher did not participate in the current research (Primary vs. Secondary Sources, 2018). Secondary data was used to help describe the research context.

4.2.1 Primary data sources

Primary data sources for this study consisted of qualitative methods. Qualitative methods consisted of discussions, and semi-structured interviews. Information collected from these methods were exploratory in nature and had a subsidiary role.

4.2.2 Secondary data sources

Secondary data guided the research and literature review. The data that guided the research was quantitative in nature and stemmed from fives sources.

The first data source was data collected from temperature monitors called 'TempTale ®4 USB MultiAlarm'. The TempTale® devices are ambient air temperature monitors, which are placed inside a carton, amongst the grapes, so that the device experiences the same ambient conditions as the fruit within the reefer containers. The primary purpose of the TempTale® data is to act as an internal control measure against the various supply chain actors such as the shipping lines and cold stores. Each table grape container exported to Europe is required to have at least one temperature monitoring device per shipment (PPECB, 2018), which is enforced by the perishable product control board (PPECB), but benefits producers/exporters as the data generated acts as an indicator of who in the supply chain is liable for poor end of chain fruit quality. In this study, the TempTale® data was used to map and track the fruit temperature profiles, so that weak points where temperature breaks occur can be identified. The data can, therefore, be used as an indicator of how the well the cold chain performed.

The second and third data sources were the two sets of quality control (QC) reports generated for each shipment at two supply chain nodes: the packhouse and the European receiver. The reports give an indication of each shipment's general quality according to various quality indicators. The primary purpose of the QC reports generated in the packhouses is to act as a quality guideline for the exporters' marketing team when allocating the shipments to various global markets. Since fruit quality deteriorates overtime, it is vital to get an indication of the fruit quality before the produce has been allocated since the length of the voyage to each market varies greatly. When fruit arrives in poor quality, receivers issue claims against the exporters. These claims have a monetary implication since poor quality fruit has a shorter shelf life, might have to be reworked/packed, or could be a total loss. QC reports are, therefore, generated as proof of the end-of-chain fruit quality. For this study, each shipment's packhouse (beginning-of-chain) and European receiver (end-of-chain) QC reports were used as a quality benchmark to see how table grape quality changes along the supply chain.

The fourth data source consisted of historic climatic and meteorological data from weather stations located in the various geographic production regions. The data was collected by Automatic Weather Stations (AWS) owned by the South African Weather Service (SAWS), a public entity under the Ministry of Environmental Affairs (SAWS, 2022). The primary use of the data is to provide the aviation, marine and other industries with weather services as well as the fulfilment of international weather-related obligations. The data was provided by the Agricultural Research Council (ARC), a South African public science institution that conducts agricultural research.

The fifth and final data source consisted of logistical data generated by Company X for general business purposes. This dataset included other nominal variables that contextualised each shipment.

4.3 Analysis process

Quantitative content analysis was applied to the QC reports. Content analysis is an approach used to analyse text documents by quantifying the content in terms of predetermined categories in a systematic and replicable manner (Bryman & Bell, 2015b). The categories that were chosen worked as fruit quality indicators, which in turn were used to determine how the supply chain impacts overall quality. Content analysis requires the design of a coding schedule, the document into which the data relating to quality was entered, and a coding manual, which is a written statement of instruction to coders that specifies the categories to be used to classify the text based on a set of rules that define how the text is to be classified. The two frameworks help to ensure a transparent research method allowing for feasible replication and follow-up studies. Once the data has been recorded in the coding schedule, which were stored on Microsoft Excel spreadsheets, in-depth data visualisation techniques using *Tableau* were conducted.

The temperature monitoring device data underwent time-series analysis where line graphs were plotted using statistical programmes such as *Tableau*. The in-transit handling protocol, according to the PPECB, requires table grapes' pulp temperature to be set at minus 0.5°C (maximum pulp temperature at loading 1.5°C; carrying at minus 0.5°C for the full duration of the voyage) GT15 (PPECB, 2018). A deviation from handling protocol is considered a temperature break, which can be defined as, "every instance in which the temperature reading rises higher than 2°C or drops lower than -1.5°C for longer than 90 minutes" (Fresh Produce Exporters' Forum, 2016:105; Goedhals-Gerber, Haasbroek, Freiboth & Van Dyk, 2015; Freiboth, Goedhals-Gerber, Van Dyk & Dodd, 2013). By analysing the TempTale® data, the extent of cold chain temperature breaks was established. In-depth data visualisation techniques were achieved by using software such as *Tableau,* which allowed the analysis at the individual device level.

By using the data, generated from each source, in tandem, the cold chains' impact on table grape quality could be established.

Qualitative data was collected by means of measures mentioned above.

4.4 Sampling techniques

For this study, all shipments for the 2019/2020 season shipped by the exporting company made up the population. This study followed a case study design where data received from fruit sent to Europe made up the sample frame.

A non-probability consecutive sampling approach was applied to data available within the sample frame. A consecutive (total enumerative) sampling technique is one where every subject meeting the criteria for inclusion is selected until the required sample is achieved. For this study, shipments sent to European receivers that didn't generate satisfactory reports, reports that lack key quality indicators, shipments where temperature monitors malfunctioned, or receivers that only received a small number of shipments (which could distort the results) were all excluded. The sample size is, therefore, as large as the total number of subjects that qualify according to the criteria, which states that a sample size should at a minimum be 100 observations with a maximum of 10% of the total population, not exceeding 1000 observations (Zikmund *et al.*, 2013: 436). The sample for this study consisted of 467 observations to conduct the data analysis.

4.5 Statistical techniques

Statistics is the discipline concerned with collection, arranging, analysing, interpreting, and presenting data, i.e., the process of converting data into information. Statistics is broken down into two categories, namely descriptive statistics and inferential statistics (Saunders *et al.*, 2019).

4.5.1 Descriptive Statistics

Descriptive statistics is the collecting, organising and summarising of data using tables and graphs (Zikmund *et al.*, 2013: 484). Data can be summarised using bar graphs, histograms, line graphs, pie charts etc. These graphs can show the distribution of the data, the skewness, which impact inferential statistic decisions. Measure of central tendency, such as mean, median and mode can also be used to describe the data. There are also measures of variability, such as range, variance, and standard deviation available.

Descriptive statistics were especially useful for this dataset since data from five unique sources were combined for the first time. These data sources consisted of 1) climate data from the vineyard prior to harvest, 2) shed intake QC reports, 3) supply chain temperature data along the sea voyage, 4) arrival QC reports, and 5) nominal logistic data. Descriptive statistics aided in determining trends visually, which were then further explored. Descriptive statistics are usually the preliminary step before more formal inferential statistics are applied as descriptive statistics give insight into the dataset.

4.5.2 Inferential Statistics

Inferential statistics make use of statistical models based off mathematics to explain the relationship between variables using data drawn from the population in the form of a sample. A statistical model is a set of assumptions concerning the generation of observed data and similar data.

Inferential statistics used in this study consisted of correlation analysis to understand the relationship between variables, which further aided in feature selection for the machine learning component of this study.

4.6 Description of Variables

The final data frame used in the analysis process is described in *Appendix E*. Each variable is tabulated providing information on the variables name, the data type, variable type, a description of the variable, and an example. The variables and where they are obtained from are discussed in *chapter 5.*

4.6.1 Level of measurement

The variable scales consist of the four levels of scale measure, nominal, ordinal, interval, and ratio scales.

Nominal, also known as categorical, variables are used as labels for identification, which cannot be ranked, as the scaling is arbitrary, any value can be assigned. Nominal variables can be dichotomous also known as binary, where the variables consist of two categories.

Ordinal variables have ranked order with an equal distance across the range of values. The values can either be quantitative or qualitative in nature.

Interval and ratio values are purely quantitative in nature and contain properties of both nominal and ordinal scales, but they also capture information about the differences in quantities for a concept. The difference between interval and ratio scales is that ratio scales represent absolute quantities, whereas interval scales only possess relative meaning (Zikmund *et al.*, 2013: 297). Temperature is an example of an interval scale, where 0°C does not mean an absence of temperature but is merely a point on the temperature scale. In contrast, on a ratio scale 0 represents the absence of the concept, for example, 0 decayed berries found on arrival means decayed berries were absent or not found.

4.6.2 Statistical analysis of scales

Quantitative or numerical values can consist of all scale types, even nominal, such as in a binary setting where observations can be assigned with numerical labels (e.g., female=1, male=2) but those numbers have no mathematical meaning, they cannot be added together.

In mathematics and statistics, quantitative variables are categorised as either continuous or discrete depending on whether the value is obtained by measuring or by counting, respectively. The distinction affects the statistics that can be used on the two measure types. *Figure 4.4* summarises the distinction between the two types of measure.

Continuous variables can take on a set of uncountable values because there are infinitely many values between two measurements, often ratio scales.

Discrete measures can only take on one of a finite number of values. A discrete measure does not represent intensity of measures, only membership. Nominal and ordinal scales are discrete measures.

Interval scales can be treated as both discrete and continuous depending on what the scale is measuring. Whole numbers are discrete and fractional numbers are continuous.

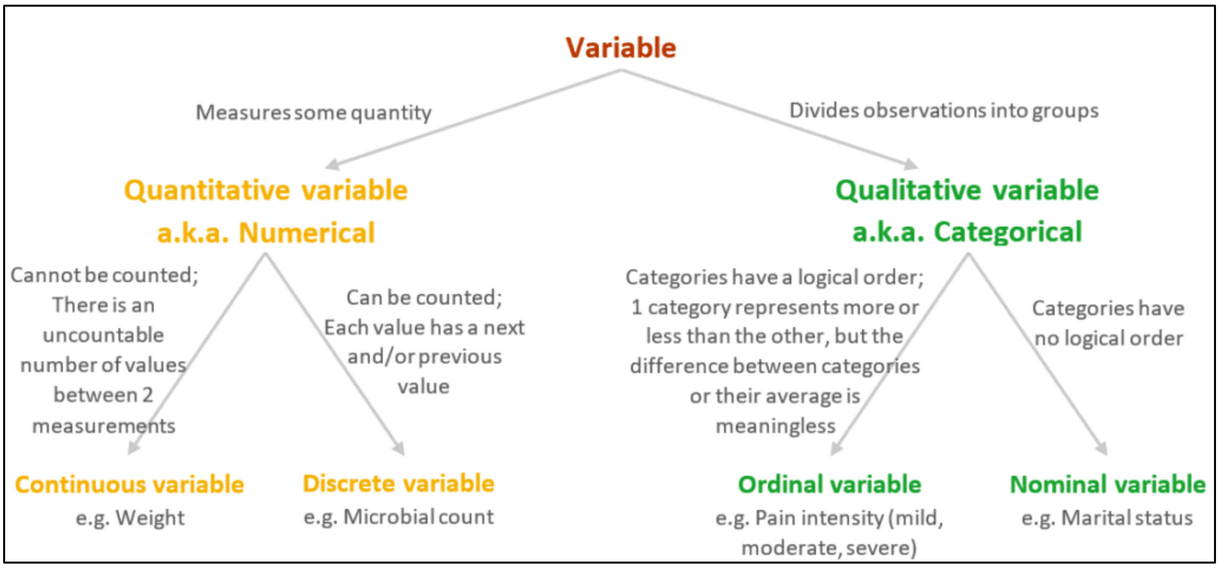


Figure 4.4: Summary of scales and measures

Source: Choueiry, 2022

4.7 Ethical Issues

The research undertaken in this study presented limited ethical issues since there is no need for questionnaire survey data in this study. There is, therefore, minimal ethical risk relating to the study due to the absence of human participants, with no personal information needed. The only risk relates to industry information. To ensure protection and anonymity, pseudonyms are used when referring to farms, packing and cooling facilities, marketing agencies, shipping lines, brands and retailers. Further precautions were taken in the form of non-disclosure agreements between the researcher and all the various stakeholders to ensure informed consent was granted, so that no privacy rights are infringed upon.

4.8 Research Validity and Reliability

Questions regarding validity and reliability are used to evaluate the quality of the research conducted.

Reliability is concerned with whether or not the data collection techniques and analytical procedures would reproduce consistent findings if the study were to be repeated on another occasion by a different researcher (Bryman & Bell, 2015: 41). The question is concerned with whether the measures are stable and consistent, resulting in the same outputs if data were to be recaptured. For this study, variables collected by sensor's, weather stations and temperature loggers in the containers, are reliable and stabile due to calibration and consistency in production of the hardware, therefore, ensuring repeatability of the data captured. Variables relating to table grape quality may suffer from the lack of inter-observer consistency due to the level of subjective judgment involved in the capturing of ordinal scale quality data. This is overcome in some variables by the capturing of continuous ratio data measuring the count of quality defects.

Research validity is concerned with the integrity of the conclusions generated by the research (Bryman & Bell, 2015: 42). Validity is categorised into a couple of types, namely construct (measurement) validity, internal validity, and external validity.

Construct or measurement validity questions whether the construct measures what it claims to measure. Many of the variables used in this study are based on standardised scales, such as the Celsius scale, which measures temperature, or degrees Brix measuring the sugar content of an aqueous solution etc. These empirical scales have been developed and tested in numerous studies confirming the validity. However, the ordinal scales developed by quality controllers have a more ambiguous measurement of validity as it is a rating scale of the extent to which an adverse quality occurrence is observed. The likelihood that true extent is accurately represented is unknown. A study can both have construct validity and not reliability if the construct correctly measures the phenomenon but is incorrectly captured or interpreted.

Internal validity refers to the degree to which there is a causal relationship between the variables chosen. This questions if the study truly explains the relationship, ruling out alternative explanations for other potential effects causing the response. This study is exploratory in nature, trying to determine which features, in what combination, have the greatest impact on the arrival quality of table grapes, the dependent variable. Multivariable analysis can, however, suffer from internal validity in terms of omitted variable bias or errors-in-variable bias due to imprecise measurement.

External validity is concerned with representativeness, whether the findings of the study can be generalised across the population or to other settings. There are threats to this evident in the study, firstly regarding the sample. The sample received is not representative of all the markets serviced by the export company, it is not randomly drawn from all table grapes exported globally or even from South Africa. The sample also only consisted of one season's product. This may influence the aptitude on the treatment interaction due to the overstated effect of certain independent variables, which could have been avoided if a multi-season dataset had been studied.

4.9 Limitations

This section presents and discusses gaps/ limitations for this research study. These potential limitations are addressed again in the final chapter where future research avenues are suggested to overcome the limitations mentioned.

The limitations for this study consist of the lack of financial data and the limited representativeness of the sample.

4.9.1 Lack of financial data

The implications of the results are limited as there was no financial data accompanying the reports. Sales price data for fruit presented in the report would have contextualised the extent to which poor scores effect income. Retail programs work on a fixed price agreed pre-season, whereas prices for fruit sold on consignment would look at the average pooled price according to pack week and product specifications (e.g., variety, pack, size). Pooling is the practise of sharing a price by all producers. The pricing method is applied to average the market value of a crop over the course of a pooling period, where the price reflects average price for the period (weekly) of market sales activities compared to the price risk inherent in fixing a price on a given day (Sengupta, 2012).

Financial data would add perspective to the implications of more or less severe claim scores. Therefore, the data set used can only determine if there was a claim but not the extent of the claim. From a producer's perspective, all their fixed/sunk costs per hectare to produce the crop have been spent by the point of harvest. It is, therefore, in their best interest to maximise the cartons packed per hectare at an additional, relatively low, variable cost. *Table 4.1* shows the farm gate[1] cost breakdown of each component applicable in the production of one hectare of table grapes. Costs further down the chain would be subtracted from the price received by the producer on the account sale received. Refer to *Appendix F*, where costs such as customs clearance, depot transport and cold storage are subtracted from the final foreign currency return.

Table 4.1: Hex River production cost per hectare breakdown

Production cost per hectre		
Fuel & Oil	FC	1.3%
Fertiliser	FC	2.3%
Water	FC	3.8%
Chemicals	FC	4.4%
Electricity	FC	8.5%
Hired transport	FC	35.1%
Management & Labour	VC	7.4%
Packaging & Marketing	VC	36.4%

Source: Adapted from SATI, 2021

To maximise profits, producers need to pack the maximum number of cartons while still receiving a farm-gate price above the total carton production cost. Due to the perishable nature of the commodity, there are significant risks associated with packing the entire crop yield and exporting the produce to the premium market channels as the likelihood that all the produce

[1] Farm gate price: Ex-works price less cooling and cold storage costs prior to containerisation.

would arrive in good condition is slim, not to mention the irreparable damage to the exporter / importer relationship.

By integrating financial data to the predictive quality scoring algorithm, marketers would be able to effectively allocate fruit to maximise returns across the quality spectrum. For example, packaged fruit with a poor future quality outlook would be sent to a less decerning market channel, where the sales price received is still greater than the break-even price per carton. This method would allow greater efficiency and less waste in the system while still maximising producer returns.

4.9.2 Sample representativeness

A drawback of the non-probability sampling approach, such as the one applied in this study, is the lack of sample representativeness. This sample selection error means that the data used for analysis does not necessarily adequately represent the entire population, negatively impacting the quality of the findings (Zikmund *et al.*, 2013: 81).

4.9.2.1 Effect of the Covid-19 pandemic

Due to the Covid-19 pandemic, the South African government declared a state of disaster and ordered a nationwide lockdown starting on Friday, 27 March 2020 (Government Gazette, 2020). This declaration imposed immediate local and international travel restrictions on citizens, in the hope to decelerate the spread of the respiratory disease. The implications on this study were that quality controllers, individuals employed by Company X, had to cease with fruit inspections at packhouses effective immediately, resulting in a loss of data for fruit harvested and packed beyond the date that the lockdown was implemented. The sample population made available to the researcher, therefore, is not representative of the entire 2019/2020 harvest period, meaning the results of the study are not generalisable to late season table grapes.

4.9.2.2 Information bias

To accurately forecast arrival quality, the input data needs to be as representative as possible. In the case of this study, the input data drawn from the shed quality reports suffer from potential information bias due to successive selection bias. Selection bias is when sample data collected for a study is not representative of the population, leading to a systematic error in the outcome of the study results (Nunan, Bankhead & Aronson, 2017). When grape harvesting occurs, pickers avoid suboptimal bunches when picking. The bunches are selected according to a broad set of subjective criteria influenced by the fluctuating product specifications required for each market. Further selection bias occurs at the next step in the chain, where the clipping teams trim out any unwanted berries, thinning out the bunches, removing any potentially decayed berries/ decay causing berries. By the time the random samples are drawn for the packhouse QC reports, the product quality is not normally distributed since all suboptimal product has been discarded.

4.10 Conclusion

This chapter discussed the major elements for the design and methods in this study. The study is an experimental case study design following a mixed method research approach, therefore, consisting of both qualitative and quantitative variables. The data analysed in the study consisted of secondary data from five unique sources, namely 1) climate data prior to preharvest, 2) shed intake QC reports, 3) cold chain temperature data, 4) arrival QC reports, and 5) nominal logistic data. A non-probability consecutive sampling technique was applied and resulted in a dataset of 467 observations. The data was analysed using both descriptive and univariate statistics to understand the data and to aid model building.

Chapter 5: Data Pre-processing

In this chapter, the data extraction, preparation and understanding phase is discussed. As mentioned in the methodology, the data obtained had another primary purpose and was gathered from five independent sources. The format of each consisted of reports, both Excel and PDF, .ttv files and a tabular dataset readable in Excel. This meant that the data first needed to be extracted, reformatted, and joined before any data analysis could take place.

The data sources consisted of reports gathered on the same product at different points along the table grape supply chain, identified in *Figure 5.1*.

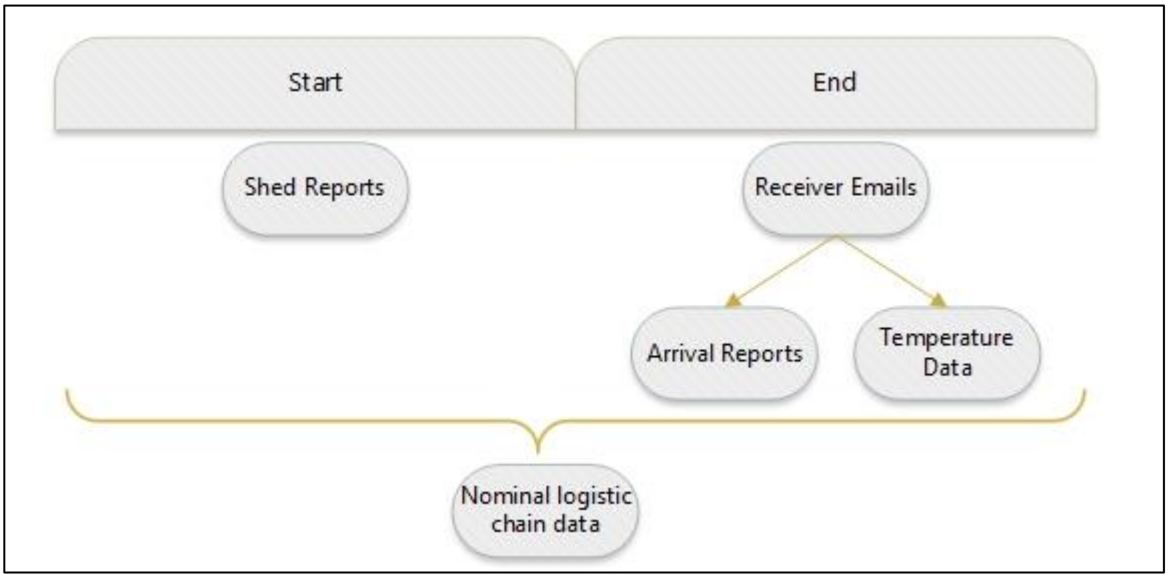


Figure 5.1: Data source collection points in the supply chain

Source: Author's own

5.1 Data Extraction

Shed-quality reports' (reports generated to indicate the post-harvest quality/condition of table grapes at the start of the supply chain) native file format is generated in a non-tabular text layout, saved as MS Excel workbooks. Each file was created per farm; variety; and date. The format resembles that of a PDF file, with fixed-layout text. The choice of format relates to the primary use of the reports, which was to give marketers' insight into the fruit quality to appropriately allocate the fruit according to the programs available. An example of the data and format can be seen in *Figure 5.2*.

ID	Product		Detail		Lug Inspection		Brix	Carton Weight	External Quality		QC Score	General Comments	Tolerance Comments
Grading: GREEN													
1	**Farm Name** **Commodity** **Variety Name**	[redacted] FARMS GRAPES GRAPAES	**Size** **Mark** **Pack** **Target Market**	XL [redacted] 4.5 KG EUROPEAN	**Berry Temp** **Stem Condition**	20.20 FRESH	14.50	4.68	**Berry Size** **Browning** **Colour Uniform** **Decay per Carton** **Loose Berries** **Marks** **Split Berries**	18.00 N GOOD 0.00 6.00 6.00 0.00	GREEN	Good Quality. . Weights Are Too Light - But It Was Fixed Immediately. . Only One Pallet Was Packed With Underweight Boxes.. Berry Size Is Borderline. .	Optimal Temperature No Decay
Grading: RED													
2	**Farm Name** **Commodity** **Variety Name**	[redacted] FARMS GRAPES GRAPAES	**Size** **Mark** **Pack** **Target Market**	L [redacted] 10X500G EUROPEAN							RED	The Grapes Look Good.. The Individual Punnets Are Too Light.. The Weights Per Table Was Adjusted Immediately.. Two Pallets Is Packed Already..	Underweight punnet

Figure 5.2: Shed report example

Source: Jooste, 2020

The data needed to be extracted and reshaped into a wide format for analysis. To create the data set, the data had to be manually extracted from each report by copying and pasting it into sequential rows in an Excel workbook.

Arrival quality data came in the form of reports comprising of a single grower and variety per shipping container (load unit), meaning multiple reports per container were exported. Of the reports received, 762 were useable, the remainder were from clients using reporting companies with different formats, rating scales etc. According to Bryman and Bell (2015:188), the larger the sample size the lower the sampling error, improving the precision. However, there is a diminishing return of precision as the sample size grows to 1000 observations. Therefore, the reports in other formats were disregarded as the sample size was already large enough.

Arrival reports were received in email communication between the exporting company and the foreign clients. The quality reports therein were configured as PDF (portable document format) files. This format is often used for files that need to be easily shared and/ or printed, without being modified. PDF files are standardised as ISO 32000 containing flat text formatting and images in a fixed-layout format. An example of an arrival quality report can be found in *Appendix G*. To extract the text, an optical character reader/ recognition (OCR) program was used. OCR is an electronic conversion of an image based on printed text into a machine-encoded text. This is a form of text mining known as information extraction (IE), which is the process of automatically extracting structured information from semi-structured machine-readable documents.

The OCR program extracted the PDF documents' plain text and saved it as text file format (.txt), ANSI (American National Standards Institute) encoded due to the fixed width encoding format. These files are readable by Windows Notepad.

Once all the PDF files were converted to the text format, a Visual Basic for Applications (VBA) macro was used to simultaneously copy the text of the files into successive columns in Excel.

VBA is Microsoft's event-driven programming language, which allows one to build user-defined functions (UDF's) for automating processes. VBA is built into most Microsoft office applications, such as MS Excel, MS Word, etc. The macro code used was found on the Microsoft community forum and can be seen in *Appendix H*.

Once in MS Excel, the data had to be cleaned and wrangled to transform it from a raw state into a more appropriate form, ready for data analysis. This is a key step and assures quality and useful data.

Within the email communication mentioned above, temperature data files were available. These files were saved as .ttv files (TempTale® Manager Desktop Data), a format associated with Sensitech. These files are created from two types of temperature recorders listed below:

First, USB T4 (TempTale® 4), a conventional temperature monitoring device that must be retrieved and downloaded once the shipment has arrived at the destination. To extract the temperature/time data, a USB Interface Plus Reader, seen in *Figure 5.3*, must be used. Second, Temptale RF (Radio Frequency), a wireless device that automatically downloads when passing an RF Gateway reader located within the arrival destination cold store. The data recorded is the same as the T4, namely temperature/time data.

Figure 5.3: Temptale® 4 - USB Interface Plus Reader

Source: "TempTale® 4: USB Interface Plus Reader", n.d.

To extract the data from the .ttv files, the files needed to be uploaded to Sensitech's software application, ColdStream. When the recorders are physically inserted into the shipments, the recorders serial numbers are captured by the exporter. The relevant shipment information is then sent through to Sensitech to be uploaded to their database. ColdStream acts as a platform to access this database. To extract the data from the .ttv files, each file must be individually uploaded to the site so that the system can cross-reference the serial numbers with those active in the system. Once uploaded and cross-referenced, the temperature data can be downloaded to a customisable format like MS Excel.

5.2 Climate data

Climatic data was obtained from the Agricultural Research Council (ARC), a South African public science institution that conducts agricultural research. The institution has multiple campuses, administrative offices, laboratories, and research farms located across the country with the primary mandate to conduct research, drive research and development, drive technology development and transfer (dissemination).

ARC has also been mandated to manage and maintain National Public Goods Assets (NPGA). These assets include natural resources management focused on biosystematics and integrated pest and weed management, soil, climate, and water. Applicable to this research is data regarding agro-meteorological network, databanks, and information systems (Agricultural Research Council (ARC), 2019:20).

ARC provided climatic data from weather stations within the five geographic regions. This data corresponds to 39 farms from the other data sources. The distribution of farms per region can be seen in *Table 5.1*, the proportion of farms per region were as follows: Hex River (56%), Worcester (23%), Trawal (10%), Piketberg (5%), and Robertson (5%). The general geographic location of each region can be seen in *Figure 5.4*, *Figure 5.5* and *Figure 5.6*. All five regions are in the Western Cape Province of South Africa.

Table 5.1: Climatic data regional distribution

regions	**Farms per region**	**Station Name**
Hex River	22	Hexvallei PP
Worcester	9	Veldreserwe
Trawal	4	Klawer
Piketberg	2	Piketberg: Pools - Ideal_Hill
Robertson	2	Robertson PP

Source: Author's own

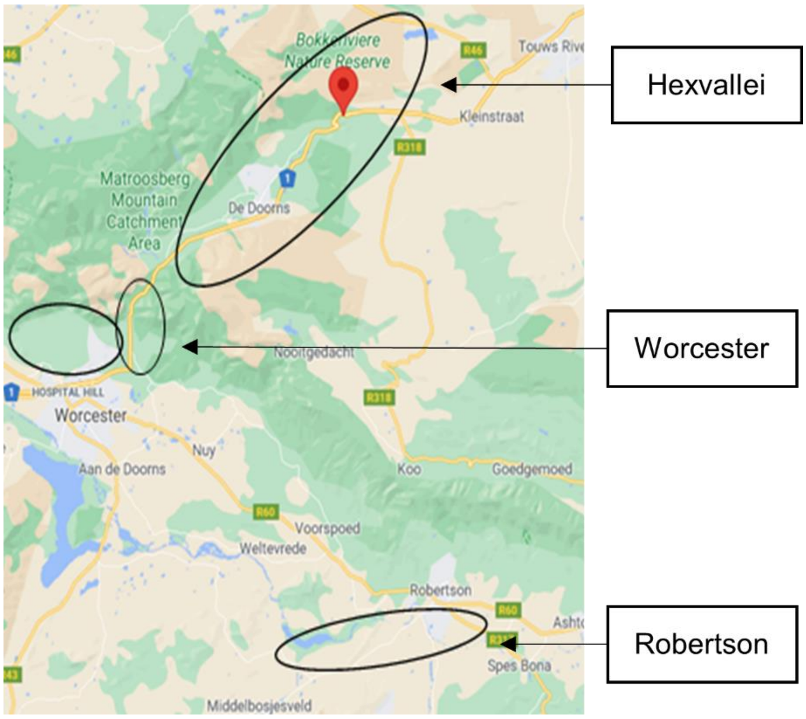


Figure 5.4: Production Regions

Source: Author's own

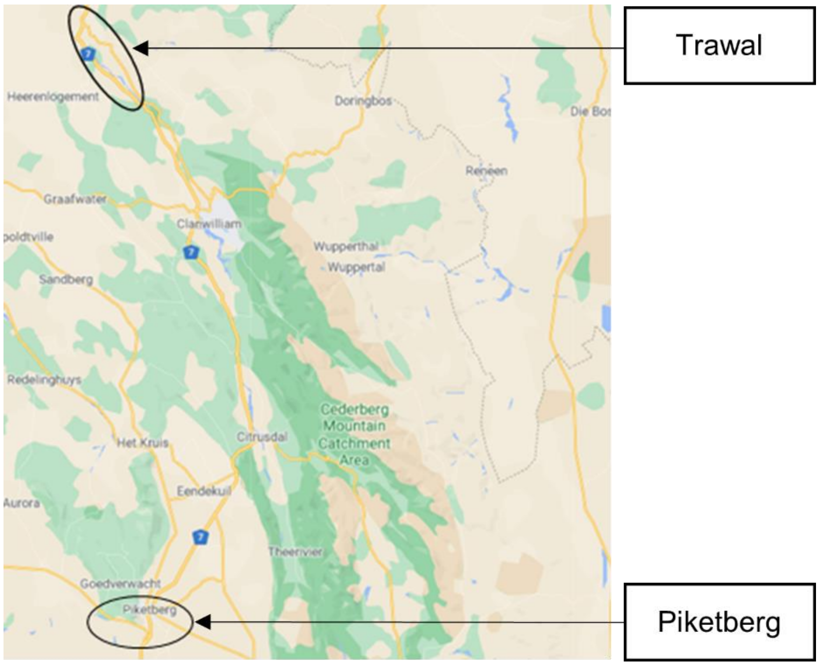


Figure 5.5: Production Regions

Source: Author's own

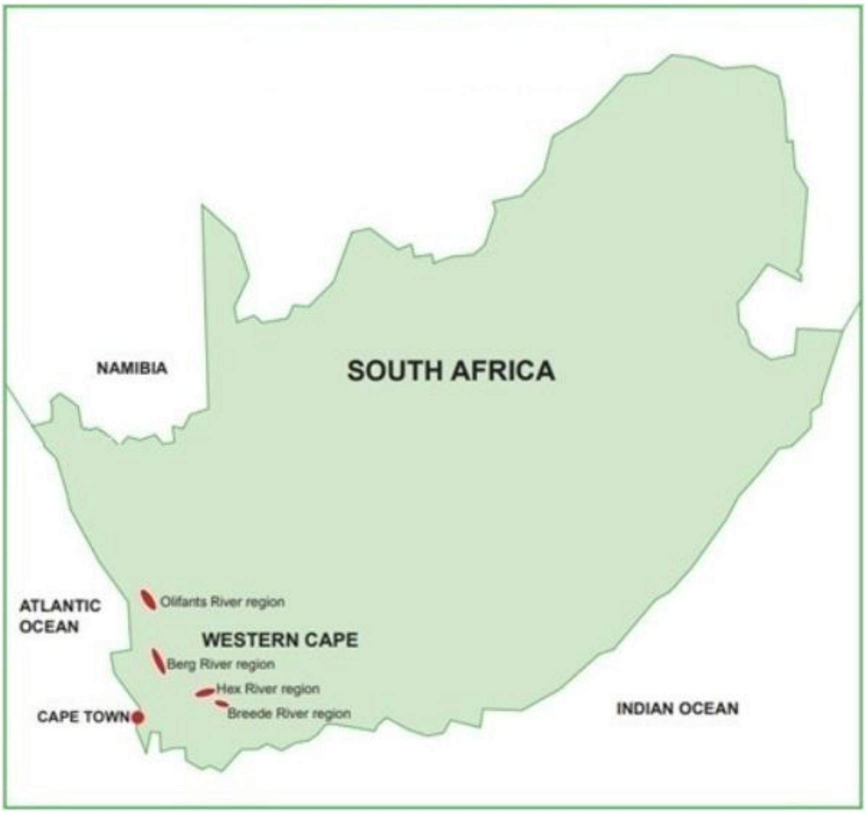


Figure 5.6: Western Cape production regions

Source: EXSA webpage, 2021

The data received consisted of seven climatic indictors recorded daily for the period 01/12/2019 to 30/04/2020 (5 months). The weather stations data is classified as time-series, a sequence of observations indexed in a time order over a set interval used to track change (Dix, 2021). To analyse the data alongside that obtained from the other data sources, the climatic data needed to be converted to a cross-sectional format.

According to the study by Hill *et al.*, (2017:87) which used logistic regression to identify the best climatic predictors of *Botrytis* decay in grapes, it found that total rainfall calculated for the last fortnight (14 days) prior to harvest is a good predictor of *Botrytis* growth and future decay. The study's results found that the predictor is significant for every *Botrytis* severity threshold and predictor combination tested, confirming producer perceptions that rainfall during this period prior to harvest leads to a higher severity of decay compared to dry weather. This study built on the sentiment that weather conditions the fortnight prior to harvest have a significant impact on future quality by creating variables that reflect this view. The variables calculated for the research can be found in *Table 5.2.* To join the climatic data to quality reports, two join clauses were used, 1) weather station region to production region, and 2) weather date time stamp to harvest date.

Table 5.2: Weather variables calculated from raw weather data, variable name, and description

Variable	Description
VIN_sum.Rain_LF (mm)	Total rainfall for the past 14 days
VIN_ET0_LF (mm)	Total evapotranspiration for the past 14 days
VIN_av.RH_LF (%)	Mean daily relative humidity for the past 14 days
VIN_av. Tn_LF (˚C)	Mean daily min temperature for the past 14 days
VIN_av. T_LF (˚C)	Mean daily temperature for the past 14 days
VIN_av. T (˚C)	Mean daily temperature (av. of min and max temp.)

Source: Author's Own

5.3 Data wrangling

Data wrangling is defined by Kandel, Heer, Plaisant and Kennedy (2011:272) as a process of iterative data exploration and transformation that enables analysis. The process consists of mapping data from a "raw" format into one more appropriate for downstream purposes with the goal of ensuring quality and useful data. Data wrangling is often the most tedious and time-consuming aspect of analysis. *Figure 5.7* summarises the broad steps involved in the data wrangling process and how they link to research in a general sense.

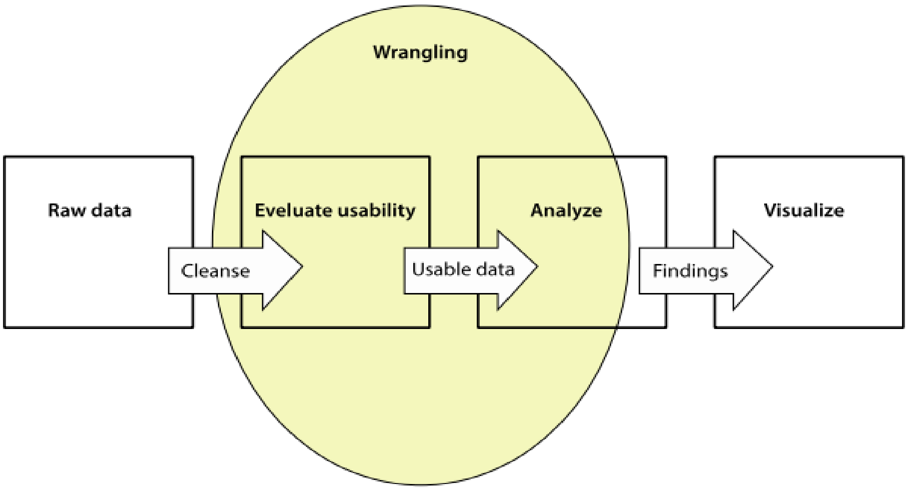


Figure 5.7: Data wrangling process

Source: Vanhamäki, Heinonen, Manskinen and Kälviäinen, 2017:1830

5.3.1 Joins and relationships

Two types of "Joins" were used to merge the tables into one complete dataset for analysis. A summary and description of each can be seen in *Table 5.3*. Data joining is necessary to combine data from multiple sources to perform the desired data analysis. Data joins are accomplished by linking two or more data tables according to a common variable, also known as a “join clause”. A join clause is a data field that is shared between both tables, using an equality operator (=), which matches rows of the same values. For example, two tables with a “container number” field can be joined by saying “container number = container number”, resulting in rows with the same IDs being aligned in the output table.

Table 5.3: Data join types used

Join Type (Venn diagram)	**Description**
Inner  $A \cap B$	Result table contains values that have matches from both tables. When a value doesn't match across both tables, it is dropped entirely.
Left  $A \cup B$	Result table contains all values from the left table and corresponding matches from the right table. When a value in the left table doesn't have a corresponding match in the right table, you see a null value in the data grid.

Source: adapted from “Join Your Data”, 2021

Venn diagrams, also known as set diagrams, show all the hypothetical potential logical relationships between a fixed collection of sets. The diagrams are used in multiple fields of study, including set theory, logic, probability, statistics, and computer science (Menzel, 2009). From a set theory perspective, the table joins can be defined using the following notation. An inner join known as an *intersection* of two sets, denoted $A \cap B$, is the set of all objects that are member of both sets *A* and *B* (Bagaria, 2019). From a data science perspective, the output table will display all data from both joining tables if the join clauses match.

From a set theory perspective, the second Venn diagram would refer to the *set difference* of two tables as well as the *intersection*, $A \cup B$. This describes a left join’s resultant output (joined table) where all the *intersection* data from set *A* and *B* is visible. The mismatching data from Table *A* will have ‘null’ values in the fields brought over from Table *B*. An example of how the left Join works can be seen in *Figure 5.8*.

Table A
Key_field (1)
1
2
23
79
108
122
1526

Table B
Key_field (2)
1
20
23
34
54
78
159

Table AB	
Key_field (1)	Key_field (2)
1	1
2	(null)
23	23
79	(null)
108	(null)
122	(null)
1526	(null)

Figure 5.8: Left Join (Tableau) example

Source: adapted from Data Flair, 2021

Five data tables were joined to get the complete dataset.

A) Intake quality shed reports (681 rows)
B) Arrival quality reports (827 rows)
C) Logistical nominal data (11 000 rows)
D) Recorder temperature data (135 rows)
E) Climate Data (760 rows)

The four data tables were joined as follows. Table A, B, and C were linked according to an inner join, so that all report data, either from intake or arrival, with no corresponding match would be removed. The output of the three joined tables (Table ABC) was then linked by a left join to Table D. A left join was chosen as there was a lack of voyage temperature data and by selecting a left join, the amount of good data excluded was minimised compared to if an inner join had been chosen. The data flow can be seen in *Figure 5.9*, with each original data table in a different colour. The final output data table consisted of 467 rows of unique data points.

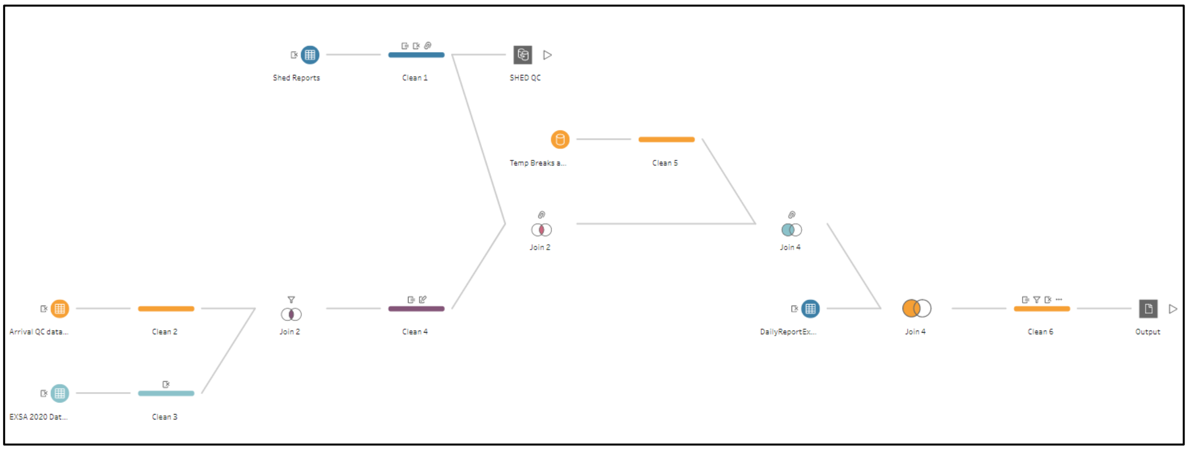


Figure 5.9:Tableau Prep Builder - Data flow

Source: Author's own

Table E, the climate data was joined to the final out data table by means of an inner join. The common key variable between the two data tables was the region, discussed in *Table 5.1*.

5.3.2 Duplicated values

Once the final output had been exported, all duplicated values were removed. Duplicated values were present due to the absence of exact fields in both data sets. If values such as the pallet ID or packing date of the sample carton drawn were available on the quality reports, it would have allowed a perfect row synchronisation between the two data tables, instead the output is an approximation. Join clause key fields (fields available in both tables) consisted of the container number, variety, farm code, and packhouse code. This level of detail is sufficient as shed reports were not conducted to a more precise level of detail either.

5.3.2.1 Adjustments

Intake shed report numerical data needed to be adjusted so that the findings would be comparable. All the discrete counts of berries with quality issues needed to be adjusted depending on the carton weight. This is important, otherwise the data would skew toward larger carton sizes. Larger cartons hold more fruit, thus, increasing the likelihood of the higher total count of poor-quality berries found in those cartons. In the table grape industry, a 4.5kg carton is the industry standard size and was, therefore, chosen as the base from which the adjustments were made. *Table 5.4* displays the conversion factors according to the various packaging configurations present in the data. The berries counted of each quality-indicator was divided by the conversion factor.

Table 5.4: Table grape carton conversion factors

		OUTER DIMENSIONS							
pack	**long_desc**	**length**	**depth**	**height**	**Std. pack**	**gross_ mass**	**net_ mass**	**Cartons/ pallet**	**std_conv (B04I_4.5KG)**
B04I	4.5kg – Loose	400	300	120	B04I	5	4.5	180	**1**
B04M	10 X 400g – Punnets	600	400	90	B04I	5	4	120	**0.8889**
B05P	10 X 500g – Punnets	600	400	90	B04I	5.5	5	120	**1.1111**
B04P	4kg - Loose	400	300	120	B04I	4.5	4	180	**0.8889**
B08I	8.2kg – Loose	500	400	127	B04I	9	8.2	102	**1.8222**
B09D	9kg - Loose	600	400	120	B04I	9.5	9	90	**2**

Source: Jooste, 2020

5.3.2.2 Common scale

Data extracted from the arrival quality reports, pertaining to the characteristics of the fruit, consisted of ordinal data on different scales. To make the data comparable, a common scale needed to be created. The condensed findings of all the reports can be seen in *Table 5.5*.

Table 5.5: Common Scale for Ordinal Data

1	not found	0%	not found
2	single piece	0-2%	very light extent
3	occasional	2-5%	light extent
4	some	5-10%	light to moderate extent
5	frequent	10-20%	moderate extent
6	much	20-25%	moderate to serious extent
7	unacceptable	25%+	unacceptable

Source: Author's own

All text data was categorised according to the above table and was given a value on a 7-point Likert scale, seen in the left-most column in *Table 5.5*.

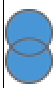
5.3.3 Voyage temperature data

Temperature data comprised of .csv files accessible by way of MS Excel. Each file comprised of one container's data. The data within contained three columns, namely: point no., date/time, and Sensor 1: Ambient Temperature (°C). The temperature data, recorded by the loggers, were set to 10-minute intervals, meaning 144 data points were captured per day. The average voyage lasted between 20 and 28 days depending on whether the container went into stacks early or late or if there were severe winds, i.e., wind speeds/gusts in excess of 40 km/h, in the Port of Cape Town causing delays to the actual time of departure. The number of rows/intervals captured per recorder, therefore, ranged between 2880 to 4032 rows per file.

A total of 135 recorders, each corresponding to a unique container number, were available from the email data source provided by Company X. The reason for a lack of temperature data along the voyage compared to arrival QC report data is due to the recorders not always being retrieved (cannot be found) by the foreign QC teams, or for those recorders that were recovered, only the downloaded graph was made available and not the .ttv file needed for data extraction.

To perform the desired analysis, the data available in each .csv file needed to be combined. The data tables were joined via *Tableau Prep Builder*, using the union join type. A description of a union between two or more sets is described in *Table 5.6*.

Table 5.6: Data join types used

Join Type (Venn diagram)	**Description**
Union  $A \cup B$	A union represents the entirety of two or more data sets. This type of join yields all elements that exist in all the data sets being joined. Data union works best when tables have the same number of fields with matching names and data type.

Source: adapted from "Join Your Data", 2021

The temperature data is classified as time-series, a sequence of observations indexed in a time order with data points consisting of successive measurements form the same data source over a set time-interval used to track change (Dix, 2021). This data type contrasts with the other cross-sectional data source, i.e., a snapshot at one instant compared to a sequence of many observations.

To make use of the data, it needed to be transformed. As stated in the research methodology, *Section 4.3*, "A deviation from the handling protocol is considered a temperature break, which can be defined as, 'every instance in which the temperature reading rises higher than 2°C or drops lower than -1.5°C for longer than 90 minutes' (Fresh Produce Exporters' Forum, 2016:105; Goedhals-Gerber, Haasbroek, Freiboth & Van Dyk, 2015; Freiboth, Goedhals-Gerber, Van Dyk & Dodd, 2013)". This definition was used as the basis for recording the frequency distribution of temperature spikes and breaks. A spike constituted one observation, for a single ten-minute interval, where the temperature recorded was outside the required range. A temperature break constituted nine sequential ten-minute intervals where the recorded temperature was outside the required range. The frequency, or count of occurrence, of temperature spikes and breaks were recorded for each container.

5.4 Data Cleaning

If multiple quantitative data samples were collected in an intake shed report, the sample data was averaged. For example, sample 1: 10 berries; sample 2: 7 berries, therefore, the data point is 8.5 poor-quality berries found for that quality indictor in that row of data.

5.5 Conclusion

In conclusion, the process to create a usable dataset to determine the factors that lead to poor table grape arrival quality consisted of many steps, involving data being extracted from five unique sources. The data was then cleaned and joined using tools such as *Tableau Prep Builder,* which resulted in a dataset of 467 unique observations. The final dataset that resulted for all the pre-processing steps can be seen in *Appendix E*. Each variable is tabulated providing information on the variable's name, the data type, variable type, a description of the variable, and an example.

The following chapter explains how the dataset and which specific variables were used to build four unique machine learning models to predict the arrival class-labels of table grapes exported to Europe.

Chapter 6: Modelling

This chapter focuses on the practical workflow of the modelling process in terms of the design decisions, methodology and steps.

The data pre-processing steps were discussed in the previous chapter. The process starts in this chapter with the model input preparation (§*6.1*), which expands on the feature selection and processing techniques applied. Model development (§*6.2*) focuses on the constructing, training, iterating, and testing of the models. Lastly, Model performance evaluation in the following chapter (§*7.3*) explains the approached used to assess the model performance and prediction on the test data.

Figure 6.1 is a schematic that indicates the steps and subprocesses undertaken in the modelling process. The order of steps undertaken are denoted by the bold chain (text and arrows) in the diagram, from *Start* to *End*. The colour block groupings in the schematic correspond with the successive sections for this chapter.

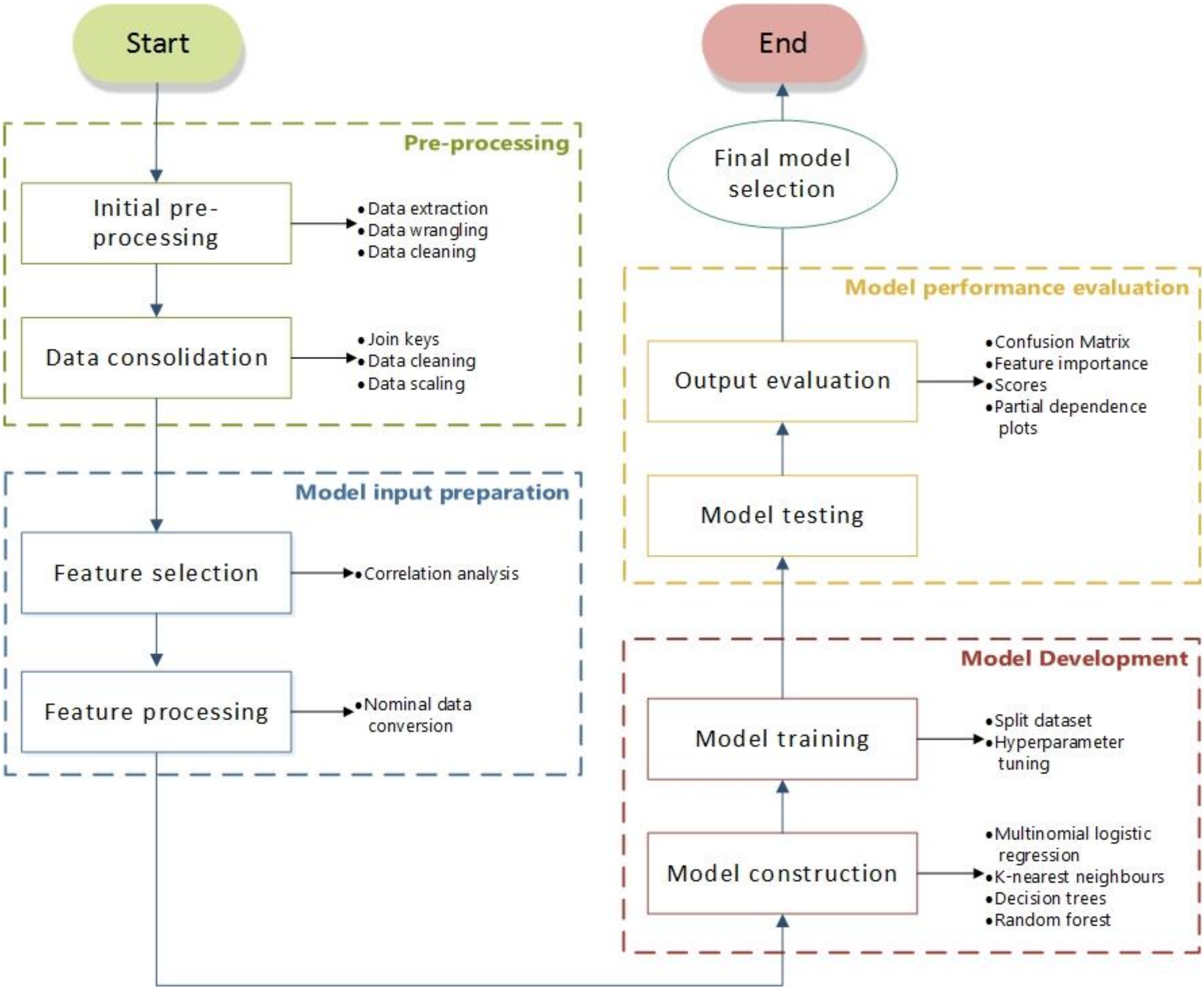


Figure 6.1: Modelling approach

Source: Author's own

6.1 Model input preparation

The data preparation step of the modelling methodology entails transformation and removal of feature attributes and training samples. This is an important step before model training to remove and (or) fix erroneous, misleading, missing data samples.

It is also an important step before features are used for model training and testing in the creation of the classification models.

6.1.1 Feature selection

Features are all the variables present in a dataset. Feature selection, therefore, refers to techniques implemented to evaluate all the features to identify the most useful ones for response variable prediction. A feature is deemed useful based on the degree it informs the model about the desired output (Murugan, 2022:22). One strategy for feature selection is the use of inferential statistics by means of confidence values to determine correlations between individual features and target variable(s). The in-depth analysis is discussed in detail in *section 7.2*, of the following chapter.

A total of 9 features, listed in *Table 6.1*, were selected for model building from the 42 available features. These features showed the strongest correlation to the response variable and were, therefore, chosen. These features concur with literature regarding variables that influence table grape quality. The features consisted of the table grape varieties (V6), packaging variables (V9 and V10), table grape sugar levels (˚Brix) at harvest (V20), in field climatic variables (V21 to V25), and packhouse QC variables (V32 and V45).

Table 6.1:Input Features selected for model building

REF#	FIELD NAME	DATA TYPES	VARIABLE TYPE	EXAMPLE
V6	NOM_variety	CATAGORICAL	NOMINAL	CSS
V9	NOM_pack	CATAGORICAL	NOMINAL	B05P
V10	NOM_inv_code	CATAGORICAL	NOMINAL	UL
V20	VIN_Brix readings - AVERAGE	NUMERICAL	CONTINUOUS	19.8
V21	VIN_sum.Rain_LF (mm)	NUMERICAL	CONTINUOUS	0.51
V23	VIN_av.RH_LF (%)	NUMERICAL	CONTINUOUS	0.5998
V25	VIN_av. T_LF (˚C)	NUMERICAL	CONTINUOUS	21.33
V32	IN_Decayed berries*	NUMERICAL	DISCRETE	0
V45	IN_QC Score	CATAGORICAL	ORDINAL	2

Source: Author’s own

The response variable the models will predict can be seen in *Table 6.2*, which is the arrival quality score (V63), and consists of three class-labels: green, amber, and red.

Table 6.2: Output (response) variable to be predicted

REF#	FIELD NAME	DATA TYPES	VARIABLE TYPE	EXAMPLE
V63	AR_QUALITY SCORE	CATAGORICAL	ORDINAL (1-3)	2

Source: Author's own

6.1.2 Feature processing

Categorical variables (features V6, V9 and V10) needed to be encoded before modelling could take place. There are various approaches available; label and one-hot encoding were used for this study.

One-hot encoding was applied to feature V6, the variety feature. Due to the number of unique variety variables found in the dataset, a ranking or 'bin' system was applied to reduce the number, otherwise creating 'noise' in the model. The varieties were ranked based on the perceived long-term quality or storability of each variety. Two industry experts were consulted and agreed on the rank order displayed in *Table 6.3*.

Table 6.3: Table grape varieties ranked according to perceived long-term quality

RANKING	VARIETY NAME	VAR	OBSERVATIONS	OBS. PER GROUP
1	SWEET JOY	I17	39	59
1	SABLE	SGS	2	
1	SWEET CELEBRATION	I75	18	
2	CRIMSON	CSS	210	213
2	LA ROCHELLE	LAR	2	
2	ALPHONSE	ALP	1	
3	SCARLOTTA	SGE	7	37
3	RED GLOBE	RGB	3	
3	SWEET GLOBE	I10	25	
3	ALISON	ALI	2	
4	MELODY	MLY	12	58
4	ARRA 13	ATN	1	
4	SUGAR CRISP	SCF	1	
4	AUTUMN ROYAL	ATR	15	
4	JOYBELLS	JBS	21	
4	ADORA	S34	2	
4	AUTUMN CRISP	S35	6	
5	THOMPSON	THS	8	13
5	SUNRED SEEDLESS	SRS	2	
5	RALLY	RAL	2	
5	FLAME	FMS	1	
6	TAWNY	TAW	7	87
6	REGAL	RGT	7	
6	SUNDANCE	SDC	5	
6	SWEET SAPPHIRE	IF6	68	

Source: Author's own

One-hot encoding the new variety data created six additional features, where the data for each feature contained either a 1 or 0 (true or false) binary value, implying that for the observation, the variety category was either present or not present. This method of encoding was applied to all 467 observations from the data frame.

Label-encoding was applied to the external packaging (V9) and inventory / internal packaging (V10) features respectively. The process of label-encoding assigns a numerical label that corresponds to the unique categorical label.

6.1.3 Handling missing values

Models cannot perform on datasets with missing values. To overcome this, missing values were filled by using the average of the feature column.

6.2 Model development

The model development phase encompasses the steps associated with constructing, training, iterating, and testing the models.

6.2.1 Model construction

Four models were developed side-by-side to investigate which type of model performed best for the classification task. The models consisted of:

- Multinominal logistic regression
- K-nearest neighbours
- Decision tree classifier
- Random forest classifier

Models were developed in Python using the open-source Jupyter notebook programming software. The *scikit-learn* machine learning toolset was used for model building. The *pandas* and *seaborn* Python libraries were used for data wrangling and visualisation. Python modules used and at what stage of the workflow, can be found in *Table 6.4*.

Table 6.4: Python packages installed for model construction

Stage in workflow	**Package names**
Data pre-processing	pandas
	numpy
	from sklearn.impute - KNNImputer
	from sklearn - preprocessing
	from sklearn.preprocessing - OrdinalEncoder
	seaborn

	matplotlib.pyplot
Models	from sklearn.linear_model – LogisticRegression
	from sklearn.neighbors - KNeighborsClassifier
	from sklearn.tree - DecisionTreeClassifier
	from sklearn.ensemble - RandomForestClassifier
Hyperparameter tuning	from sklearn.model_selection - GridSearchCV
Model Analysis	from sklearn.metrics - accuracy_score
	from sklearn.metrics - classification_report
	from sklearn.metrics - recall_score
	from sklearn.metrics - confusion_matrix
	from sklearn.inspection - partial_dependence, plot_partial_dependence

Source: Author's own

The goal of a model is to be able to generalise the prediction beyond that of the dataset it was trained on. Cross-validation was, therefore, used to split the dataset into equally sized groups known as 'folds' for model training and testing. Five-fold cross-validation was used where the dataset was randomly split into five folds: {F1,F2...,F5} each containing 20% of the training data. The process works by means of training five models, each with a different combination of folds, four folds being kept for training and the fifth for validation. For example, folds F1 to F4 for training with fold F5 being kept as a validation set. The process is iteratively repeated with each of the five folds getting a turn to be a validation set. Finally, the performance measures are averaged across the five iterations to estimate the capability of the algorithm for the classification problem. The performance measure for this problem is the prediction accuracy of the arrival quality scores of table grapes.

Model training and optimisation

Model parameter tuning was required to better tune the models to the prediction problem. The tuning process employed a grid search method, which searched through each set of model-specific hyperparameters. The hyperparameters chosen were based off the combination that yielded the highest score on the training data set.

The grid searched and best hyperparameters per model are listed in *Table 6.5-Table* 6.8.

Table 6.5: Multinominal Logistic Regression

Hyperparameter grid	1. Algorithm to use: ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'] 2. Number of iterations: [100, 1000, 10000]
Best parameters	1. Algorithm to use: ['saga'] 2. Number of iterations: 100

Source: Author's own

Table 6.6: k-Nearest Neighbours

Hyperparameter grid	1. Leaf size: [1 to 5], increments of 1 2. Number of neighbours: [1 to 15], increments of 1 3. weights: ['uniform', 'distance'], 4. algorithm: [‘auto’, ‘ball_tree’, 'kd_tree', 'brute']
Best parameters	1. Leaf size: 1 2. Number of neighbours: 5 3. weights: 'uniform' 4. algorithm: ‘auto’

Source: Author’s own

Table 6.7: Decision Tree Classifier

Hyperparameter grid	1. Tree depth: [5, 10, 20] 2. Maximum number of leaf nodes: [5 to 11], increments of 2
Best parameters	1. Tree depth: 5 2. Maximum number of leaf nodes: 5

Source: Author’s own

Table 6.8: Random Forest Classifier

Hyperparameter grid	1. Number of estimators: [50 to 300], increments of 50 2. Tree depth: [5 to 50], increments of 5 3. Min number of samples required for a split: [2, 4, 8, 16] 4. Min number of samples required for a leaf node: [1, 2, 4] 5. Number of features to consider for the best split: ['auto', 'sqrt'] 6. Criterion for split quality: ['gini', 'entropy']
Best parameters	1. Number of estimators: 50 2. Tree depth: 20 3. Min number of samples required for a split: 8 4. Min number of samples required for a leaf node: 1 5. Number of features to consider for the best split: 'sqrt' 6. Criterion for split quality: 'gini'

Source: Author’s own

Cross-validation scores were used to select the best performing class of model and their hyperparameters and are discussed in *section 7.3*.

6.3 Model performance evaluation

The different model performance was evaluated based on the cross-validation scores computed on the training dataset.

The final model was developed based on the best cross-validation scores and hyperparameters and trained on the entire training dataset. The model performance is evaluated on the predictions from the test set and can be found in *section 7.3.1* onward.

6.4 Conclusion

The modelling process followed a succinct repeatable methodology, outlined in the workflow (*Figure 6.1*). Each task executed was described in a sufficient level of detail for the research to be clear, and for the result to be justified. The feature selection process briefly mentioned and the results from the modelling process implemented are discussed in the following chapter.

Chapter 7: Data Analysis

This chapter is dedicated to graphically illustrating, analysing and summarising the data collected for this study. The analysis process consisted of three levels: descriptive statistics, inferential statistics, and model evaluation. The initial descriptive analysis guided the inferential analysis, which in turn, guided the machine learning modelling and evaluation. The analysis process was conducted using two software packages, which consisted of Tableau® and Anaconda. Navigator's Jupyter Notebook for Python coding and graphing.

The chapter analyses the relationships between the different combinations of variables to understand how each relates to one another and ultimately impacts the arrival quality of exported table grapes. Thereafter, the most relevant variables were selected and used to build models that predict the arrival (end of chain) quality of table grapes. The best performing model was selected for in-depth evaluation.

7.1 Descriptive statistics

Descriptive statistics are used to describe characteristics of the dataset. According to Zikmund, Babin, Carr, *et al.* (2013: 484), "descriptive analysis is the elementary transformation of raw data to describe characteristics such as central tendency, distribution and variability". Other techniques include mean, median, range, variance, and standard deviation. All these methods provide a summary describing basic properties of a dataset.

7.1.1 Financial implications of arrival quality scores

Not all fruit that is poorly scored will necessarily have a financial impact on the agreed sales price. A financial impact, or 'claim', can be seen as any additional costs subtracted from the previously agreed sales prices due to the arrival condition of the packaged fruit not being presentable to retail customers. There are varying degrees of financial impact, with negligeable re-sorting being on one end of the spectrum and a dumped consignment (known as a 'total loss') on the other.

The additional costs associated with a claim are due to the added labour required to sort and remove bunches/punnets in poor condition. Aside from these costs, a lower total sum would be paid by the importer/trader since a complete consignment was not delivered, occasionally forcing the importer to procure supplementary fruit on the open market at a higher price. An example of a punnet that would be removed in the arrival sorting process can be seen in *Figure 7.1,* a punnet of white seedless grapes with a visibly decayed berry and dry stems. This punnet would be separated from the rest of the consignment and would be reworked or even dumped/destroyed.

Figure 7.1: Punnet of white seedless table grapes being evaluated during inspection

Source: Author's own

Exporters have created a grading system to manage their income expectations based on the arrival quality report scores. A green will have no claim, an amber a potential claim (QP), and red a quality claim (QC), which is a definite quality (condition) related claim. QP's can either become a claim or will have no financial impact. The QP grade is used to flag the consignment prior to the payment. The data is reconciled once account sales are received, either adding or removing claim grades where applicable.

Figure 7.2 demonstrates the quality scores relative to claim probability. The figure shows the percentage of arrival reports according to the original arrival report scores (scale from 1 – 6, 11-point Likert scale) visible on the x-axis and colour shows details about arrival quality score. The data is grouped according to claim type data received from the exporter, indicating if there was a financial claim or not.

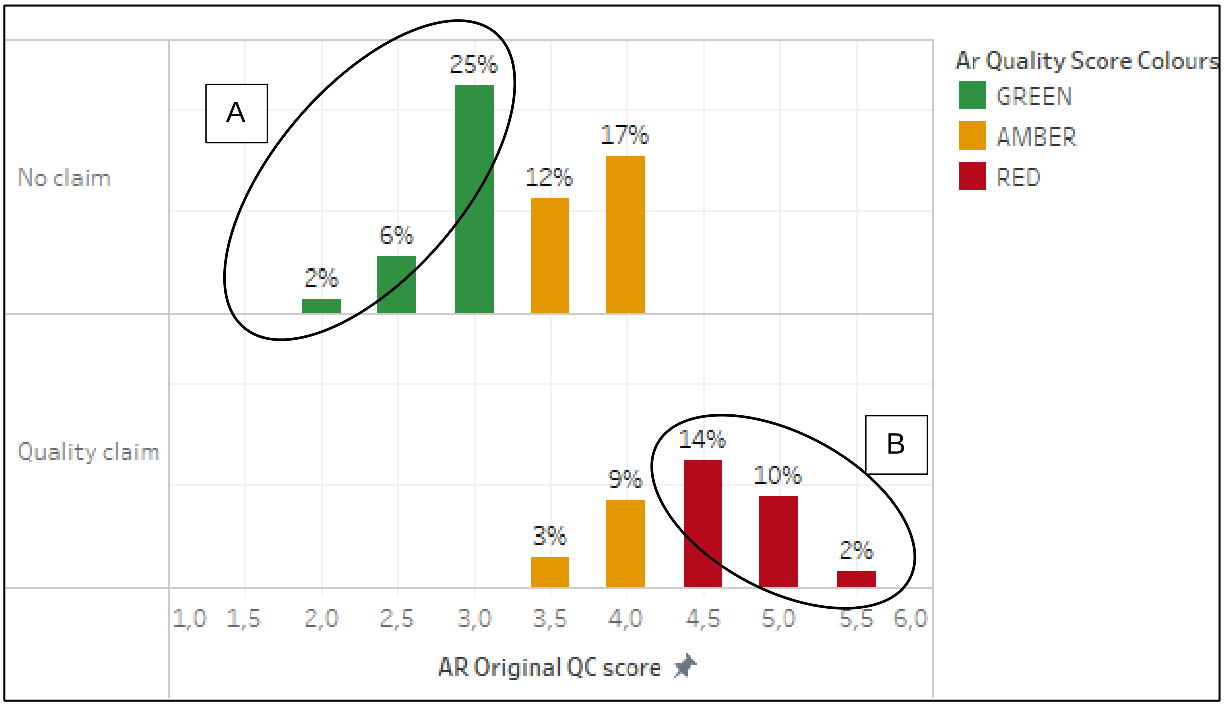


Figure 7.2: Financial implications based on report grading scale

Source: Author's own

Two filters were applied to the data. Firstly, data smoothing was employed to any bars with less than 5 report incidences (5/467 ≈ 1%) to remove outliers. Secondly, all insurance claim data was removed, consisting of 2 observations. Insurance claims are instituted when the integrity of the temperature-controlled cargo is damaged along the chain. These claims are instituted against a third-party carrier, such as the shipping line, port authority or inland distribution provider.

Common causes of insurance claims are reefer breakdown, reefer not being plugged-in, container impact damage, setpoint changed in-transit, etc. (Yarwood, 2021). Quality reports of fruit with insurance claims will score poorly but will not have any financial impact as the insurer carries the additional costs. *Appendix B* is an example of fruit that arrived in poor condition due to an incident along the supply chain. The fruit was still salvageable but received a sales price below the average pool price for the week. An insurance claim was instituted by the exporter for the income difference between the actual and potential plus the decay and repacking costs.

Insurance claims were, therefore, removed as including the data would misrepresent the extent to which report scores correlate with the likelihood of a financial claim.

Figure 7.2 demonstrates the quality reports accuracy in predicting the likelihood of financial repercussions downstream in the supply chain. A good score (green) between 1 and 3, caption *A*, will most likely have no financial claims, whereas a poor score (red) between 4.5 and 6, caption *B*, will most probably have additional costs resulting in a claim. Amber scores represent uncertainty but from this dataset, over 70% of amber scored reports resulted in no claim. Even though amber scored fruit is not of the optimal condition, it is still acceptable as there are no additional costs. Sixty two percent (62%) of reports were sound with no claim, while the remaining 38% resulted in a financial claim.

7.1.2 Content analysis of "issues" in arrival reports

As stated in the research methodology, content analysis was used to extract data from the arrival quality reports, an example of a report summary can be seen in *Figure 7.3*.

Grapes, Crimson Seedless, 5.00 kg cartons (10 x 500 grs punnets)

Number	Reference	Grower(s)	Packhouse(s)	Brand(s)	Cat	Decay	Issues	Storage	Summary	Size (quantity)	Score	Color
1425091-01-01	1400024955				1	occasional a single decayed piece found	Decay	short	The consignment is fit for a careful marketing.	r (480)	3 / 4 ø 4	

Grapes, Joybells, 5.00 kg cartons (10 x 500 grs punnets)

Number	Reference	Grower(s)	Packhouse(s)	Brand(s)	Cat	Decay	Issues	Storage	Summary	Size (quantity)	Score	Color
1425091-01-02	1400024955				1	much found during inspection	Decay, First impression, Texture/consistency	none	We advise to repack the consignment prior to marketing. Marketing problems are to be expected.	r (1080)	4 / 5	

Grapes, Sweet Celebration, 5.00 kg cartons (10 x 500 grs punnets)

Number	Reference	Grower(s)	Packhouse(s)	Brand(s)	Cat	Decay	Issues	Storage	Summary	Size (quantity)	Score	Color
1425091-01-03	1400024955			Generic	1	frequent found during inspection	Decay, Texture/ consistency	none	Marketing problems due to the presence of decay are to be expected.	r (960)	4 / 5	

Figure 7.3: Example summary of an arrival QC report

Source: Author's own

Content analysis was used to determine what the "issues" were that led to green, amber, and red class-labels scoring. Data was captured on the "issues" noted from the in the arrival QC reports (circled in *Figure 7.3*) and were recorded in variable V66 (AR_ISSUES). Unique values were counted and grouped based on the colour class-label for the observation. The results can be seen in *Appendix I*, and are summarised in *Figure 7.4*.

	Ar Quality Score Colours (grou..		
AR_ISSUES (group)	GREEN	AMBER	RED
None	31%		
Decay		31%	13%
Decay / Shrivelling		2%	10%
Berry drop / Decay		1%	1%
Grand Total	32%	41%	27%

Figure 7.4: Summary of content analysis

Source: Author's own

The content analysis results indicate the major cause of amber and red arrival reports is decay, accounting for 41% of amber and 27% of red reports respectively . The reports do not specify the type of decay, i.e., *Botrytis cinerea, Alternaria* (*Alternaria alternate*), or *Penicillium*, etc. but the general preventative measures are the same for all post-harvest fungal infections.

7.1.3 Fruit age and condition

As stated in the literature review, table grapes are a non-climacteric fruit, meaning that the bunches are harvested at optimal maturity and start to decline in freshness from that point onward. Table grapes, therefore, have an age dependant deterioration rate (Chabada, Damgaard, Dreyer, *et al.*, 2014) meaning that the condition of the fruit worsens over time. Consequently, the data should be correlated so that older fruit should have worse quality scores and fruit with a lower total age should have better scores.

Figure 7.5, displays the total fruit age (days) at inspection (arrival inspection date – pack date), grouped into age buckets ranging from less than 21 days to greater than 36 days. The number of arrival reports in each bucket is displayed on the x-axis, and colour shows details about arrival quality score. Each bar is labelled by the colour band's percentage of total number of reports per age group.

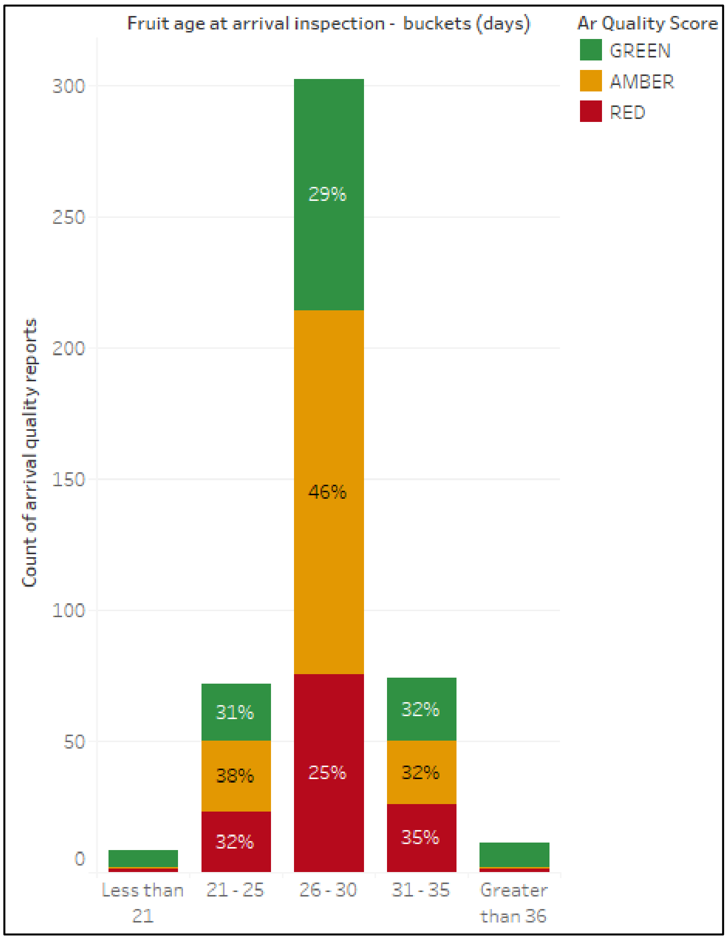


Figure 7.5: Fruit age at inspection

Source: Author's own

From the histogram it is noticeable that the total voyage length is roughly normally distributed with most reports occurring between 21- and 35-days, for 448 of the 469 reports (95.5%), the remaining outliers being 2 standard deviations from the mean. The implications of the distribution of fruit age at inspection compared to arrival QC scores are discussed further in the following chapter.

7.1.4 Cold chain descriptive statistics

One hundred and thirty-five (135) temperature monitors were retrieved for 323 of the 467 arrival quality reports. The total spread of arrival scores were 90 green observations (28%), 153 amber observations (47%), and 80 red observations (25%). The arrival quality scores in terms of the total temperature breaks observed are illustrated in *Figure 7.6*, a histogram of the total number of temperature breaks for the 135 temperature recorders.

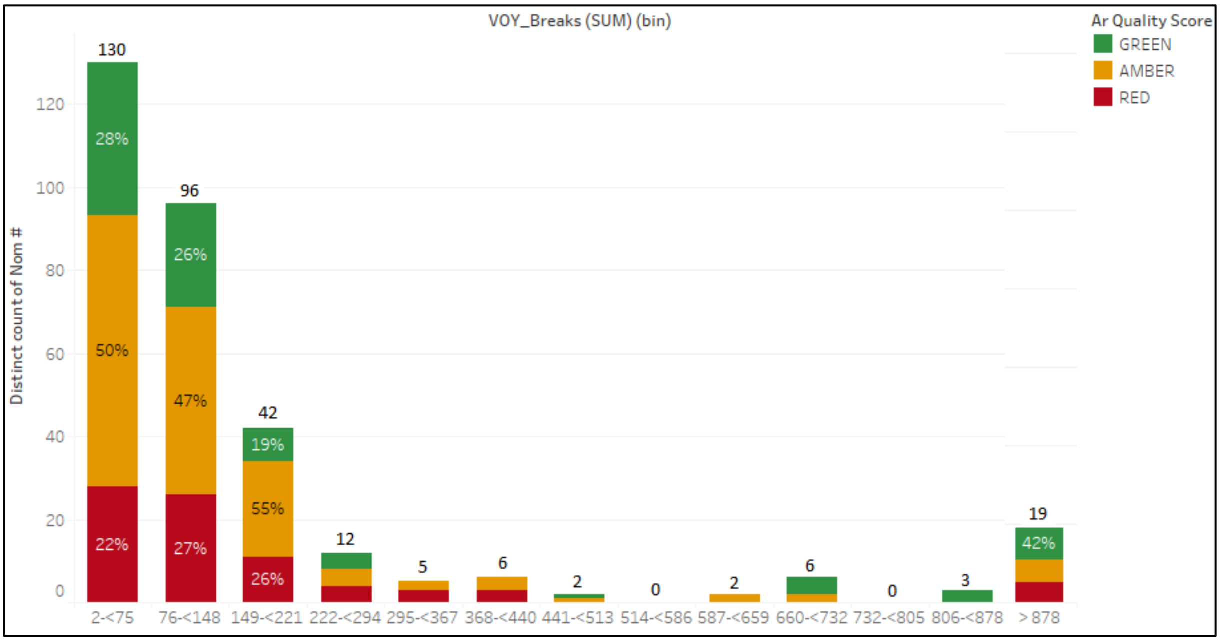


Figure 7.6: Histogram of the total temperature breaks according to the arrival quality scores

Source: Author's own

The number of histogram bins are determined by the square root of the number of monitors, $\sqrt{135}$ = 11.61, rounded up to 12 bins. The bin width was determined by subtracting the minimum value from the maximum and then dividing that by the number of bins (Bin width = (878 - 2) ÷ 12 = 73). Eight hundred and seventy-eight (878) was used as the maximum value since the outliers at the upper end of the temperature breaks distorted the histogram. Therefore, 878 was used. All outlier values greater than 878 were grouped in an overflow bin, resulting in 13 bins used for the histogram. The bin width resulted in 73 temperature breaks per bin. A summary of the histogram can be seen in *Table 7.1*.

The temperature break data is skewed to the right, also known as a positively skewed histogram, with most observations occurring in the bins where there were fewer temperature breaks recorded along the cold chain. The mean number of temperature breaks was 231, whereas the median was 92 breaks for the 135 temperature recorders. This indicates that poor performing outlier containers distort the average.

Table 7.1: Temperature break histogram summary

135	*# of monitors*
13	*# of bins*
878	*max*
2	*min*
73	*points per bin*

Bin	*Frequency*	*Cumulative %*
75	63	46.7%
148	33	71.1%
221	16	83.0%
294	7	88.1%
367	0	88.1%
441	2	89.6%
514	1	90.4%
587	0	90.4%
660	1	91.1%
733	2	92.6%
806	0	92.6%
878	1	93.3%
>878	9	100.0%

Source: Author's own

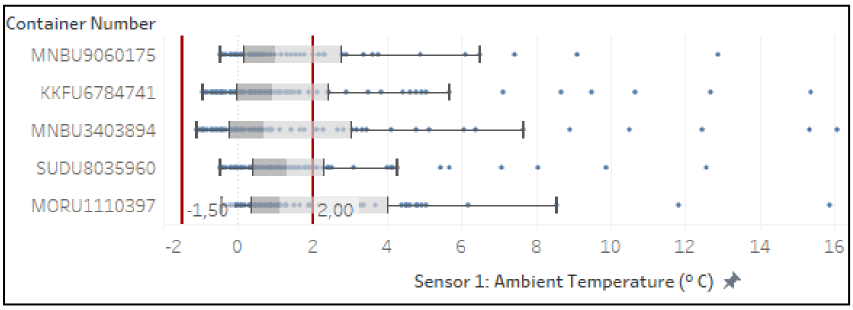
Figure 7.6 also demonstrates that the arrival quality scores are not related to the total number of temperature breaks, since the spread of the three colour scores are constant across each bin displayed in the histogram. Therefore, the supply chain effects do not seem to be visible 20 to 30 days after the time of harvest.

7.1.4.1 Characteristics of cold chain performance

Box-and-whisker diagrams are used to compare the five best performing (*Figure 7.7*) and the five worst performing (*Figure 7.8*) cold chains ranked according to the total number of temperature breaks. All 135 box-and-whisker plots can be found in *Appendix K*, as well as the number of temperature spikes and breaks for each container, found in *Appendix J*.

The red lines indicate the acceptable temperature range within the shipping container during the voyage proposed by previous cold chain studies, see *section 2.7.3.2,* where a temperature break is defined as "every instance in which the temperature reading rises higher than 2°C or drops lower than -1.5°C for longer than 90 minutes (Fresh Produce Exporters' Forum, 2016:105; Goedhals-Gerber, Haasbroek, Freiboth & Van Dyk, 2015; Freiboth, Goedhals-Gerber, Van Dyk & Dodd, 2013)".

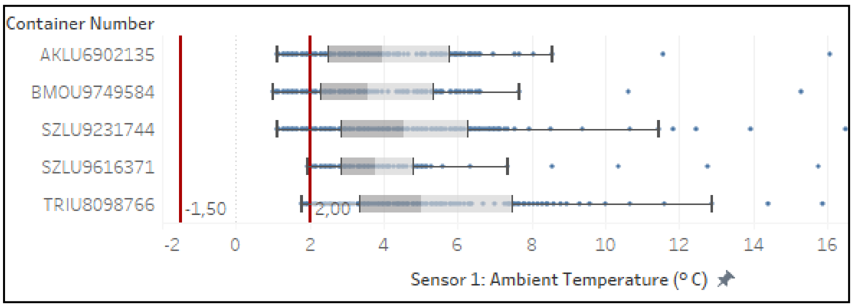
The figures indicate that the interquartile range (IQR), where 50% of the data is found, for containers that performed well are mostly within the acceptable temperature range, whereas containers that performed poorly had an IQR above the acceptable temperature range. The IQR spread, or length, for containers that perform well tended to be shorter compared to the poorly performing containers, implying that the temperature samples for the poorly performing containers displayed more variability with larger standard deviations from the mean. The whiskers, which are the extremities of the samples, are similar for both satisfactory and poor performing containers, with a lack of symmetry, where the upper bound whiskers tended to be longer than those of the lower bound, indicating that all container temperature data is skewed to the right.



Tot. spikes	Tot. breaks
16	2
20	4
21	4
28	4
21	5

Figure 7.7: The five best performing cold chains

Source: Author's own



Tot. spikes	Tot. breaks
2479	2451
2486	2461
2865	2843
3050	3014
4031	3998

Figure 7.8: The five worst performing cold chains

Source: Author's own

It must be noted that container number *SZLU9616371*, in *Figure 7.8*, displayed a symmetric temperature profile. It is, therefore, possible that the set-point temperature for that container was set at a higher incorrect temperature.

Spikes are also of relevance as a sudden increase in the air temperature while the fruit pulp temperature remains constant, can result in condensation of the moisture within the air of the container. The negative effects of condensation are discussed in *section 8.4.1*.

7.1.5 Comparing rainfall and QC scores

As stated in the literature review, mature table grapes are notably susceptible to decay caused by fungal infection during and after periods of high rainfall and relative humidity (Hill *et al.*, 2017). During the 2019/2020 harvest season, the Hex River region experienced unseasonal summer rainfall during the month of January. The average annual rainfall of Worcester South Africa, the nearest large town – 25km away, is 410 mm (World Wide Travel Organisation, 2021). Actual daily rainfall results for the harvest period can be seen in *Table 7.2*. The rainfall results were obtained from the Agricultural Research Council (ARC), demonstrating the two days of rain that fell on the 18th and 19th of January 2020, totalling 77.47mm, almost 20% of the regions' total annual precipitation. De Doorns has a semi-arid Mediterranean climate, characterised by hot dry summers and cold wet winters, therefore, any volume of rainfall above 10mm in a summer month is considered unseasonal.

Table 7.2: Hex River (De Doorns) January 2020 rainfall

Date	Week	Rain (mm)
12/01/2020	Wk 3	0
13/01/2020	Wk 3	0
14/01/2020	Wk 3	0
15/01/2020	Wk 3	0
16/01/2020	Wk 3	0
17/01/2020	Wk 3	1.27
18/01/2020	Wk 3	38.61
19/01/2020	Wk 4	38.86
20/01/2020	Wk 4	0
21/01/2020	Wk 4	0
22/01/2020	Wk 4	0.25
23/01/2020	Wk 4	0
24/01/2020	Wk 4	0
25/01/2020	Wk 4	1.78
26/01/2020	Wk 5	0.51
27/01/2020	Wk 5	0

Source: Nel, 2020

The graph seen in *Figure 7.9* displays the arrival scores and rainfall for each pack week for farms located within the Breede River district. The left y-axis shows the count of arrival reports and right y-axis shows the average sum of fortnightly rainfall prior to the pack date. The x-axis shows pack week. Each bar is segmented by colour according to arrival quality score and is labelled by the colour segment's percentage of that bar. The sum of fortnightly rainfall prior to pack date is displayed as a continuous line where there is a value of y for each value of x, calculated by averaging the fortnightly sum of rainfall for each observation of fruit packed within that week.

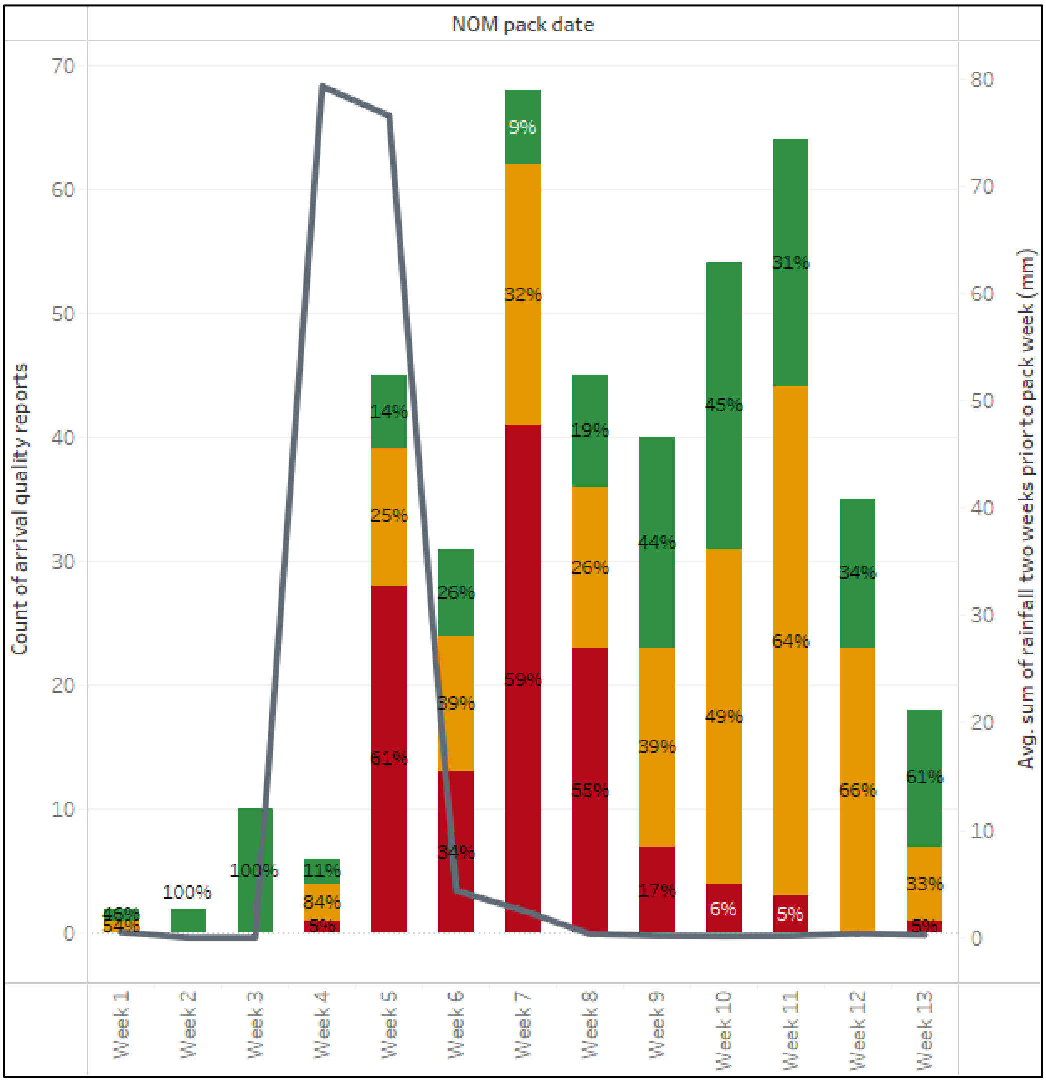


Figure 7.9: Rainfall prior to harvest compared to arrival quality scores

Source: Author's own

As stated above, the majority of the rain fell on the 18th and 19th of January, the weekend separating week 3 from week 4. Due to the rain, producers packed minimal volumes for week 4. There are a couple of reasons for this, namely the picking teams would have gotten wet from the residual water collected within the vine canopies, but the primary reason behind the break was so that producers could allow the probable physiological damage to evolve so that it could be better identified and managed within the packhouses. The rainfall variable VIN_sum.Rain_LF (mm) - Total rainfall for the past 14 days, described in *Table 5.2*, is the sum of rain that fell in the last fortnight (two-weeks). The reason as to why the line graph sharply slopes downward from week 5 to 6 is due to the fact that the start of week 6 is greater than 14 days from the heavy rainfall on the 18th and 19th, whereafter minimal rain fell for the remainder of the harvest period.

7.1.5.1 Arrival quality 5 weeks post-rain

From *Figure 7.9*, it seems that rain has an impact on arrival quality of fruit packed up to 5 weeks post-rainfall (week 4 to 8). Under closer inspection of that 5-week period, arrival quality scores seem to be variety dependent. With further inspection, visible in *Figure 7.10*, it is noticeable that different varieties harvest periods varied according to three distinct groups: early-, mid-, and late-season varieties.

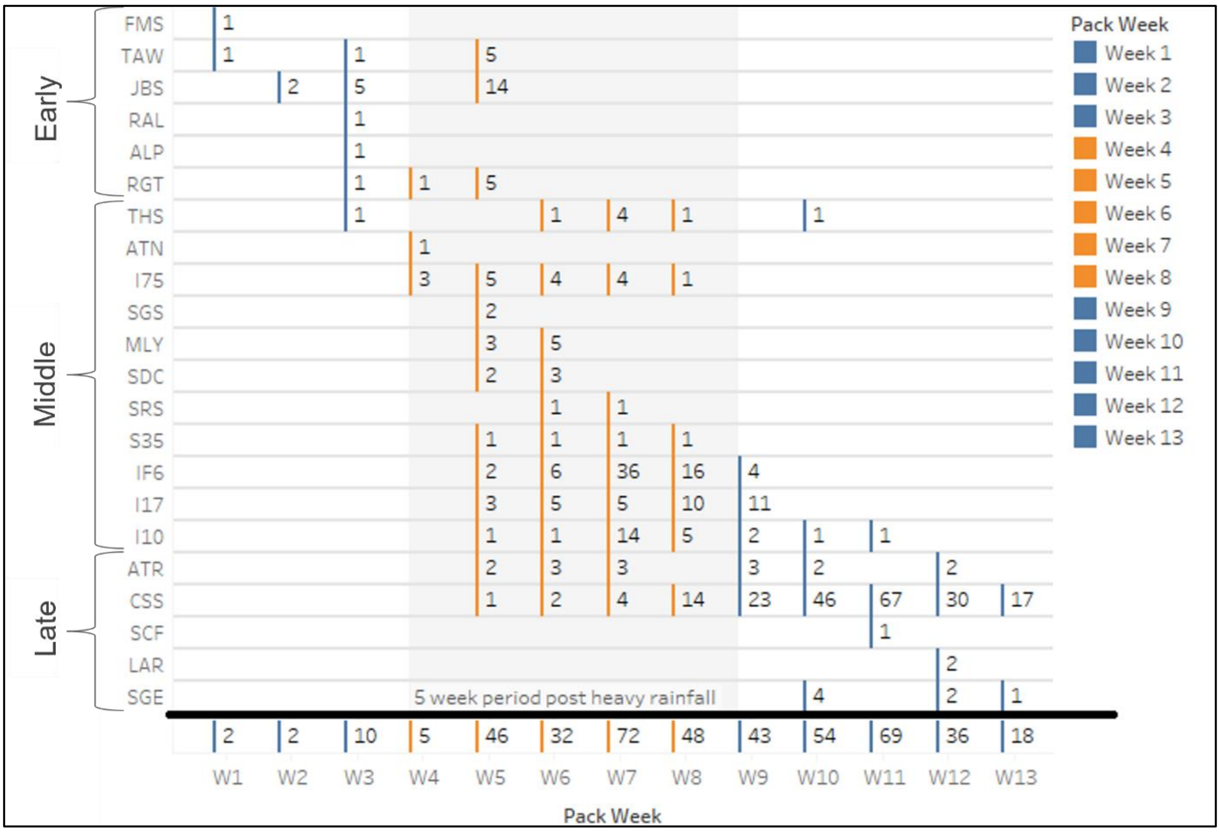


Figure 7.10: Gannt chart of the Breede River region's weekly pack according to variety

Source: Author's own

The reason for the spread in harvest period is two-fold:

Firstly, different varieties ripen at varying stages of the region's harvest period and were, therefore, at varying phenological stages at the time of rainfall. It is a deliberate choice to plant a range of varieties and is done to stagger the harvest as it reduces the producers labour burden and allows for an extended period of supply to global markets.

Secondly, different production units within the area have varying microclimates due to geographic factors affecting the growing conditions. One of the main factors affecting phenological development of a single grape vine is air temperature, expressed as heat accumulation or growing degree days index (GDD) (Verdugo-vásquez *et al.*, 2017). For example, a farm in the shadow of a mountain might have a comparatively later season, expressed across all varieties, due to comparatively less direct sunlight as well as lower average daytime temperatures. This results in delayed phenological development from bud break, through to the vine dropping its leaves in autumn.

Due to these factors, the harvest period of any variety specified will span a 4+ week period, as the data used in this research is comprised of 39 farms, each with its own unique microclimate.

Varieties most severely impacted by the unseasonal rainfall can be seen in *Figure 7.11*, which displays the number of reports per week, the proportion of reports according to the adapted three quality colours (green, amber, red), and percentage of reports according to each quality score colour for each variety harvested within the 5-week post-rainfall period.

The various varieties harvested for the period are visible on the y-axis. The x-axis is split into two: 1) displaying the pack weeks, and 2) displaying the number of arrival reports per variety. Colour and percentage show the proportion of reports according to arrival quality score for each variety.

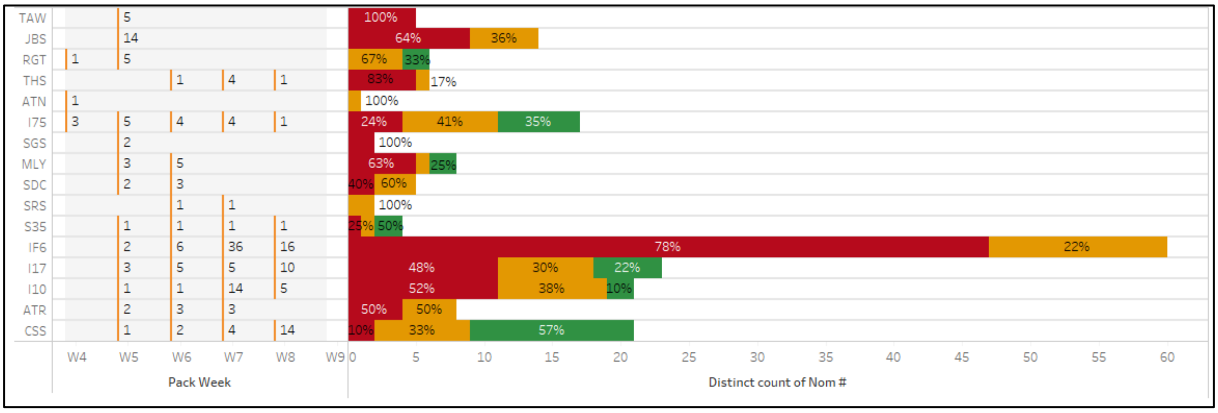


Figure 7.11: Varietal specific arrival quality scores for grapes picked within the 5 weeks post-rain period

Source: Authors own

According to the percentage of red arrival quality reports, the varieties that faired worst post-rain, were TAW (Tawny), IF6 (Sweet Sapphire®), JBS (Joybells Seedless), and MLY (Melody). Varieties that faired best, according to the percentage of green arrival reports, were CSS (Crimson Seedless), I75 (Sweet Celebration®), S35 (Autumn Crisp®), and RGT (Regal Seedless).

These results are, however, not representative as there is systematic sampling error in the form of sample bias present (Zikmund, Babin, Carr & Griffin , 2013:391). Due to the nature of the study design and the limited access to data, the results found on the impact of rain on the arrival quality of different table grape varieties are not generalisable as selection bias was apparent. This bias is made visible in *Table 7.3*, which shows the number of arrival quality reports for the eight varieties mentioned above. The general rule is that a sample size should at a minimum be 100 observations with a maximum of 10% of the total population, not exceeding 1000 observations (Zikmund *et al.*, 2013: 436). For these results, none of the variety samples are close to 100 observations (reports), therefore, the results are not conclusive and no generalisable statements can be made regarding the effect of rainfall on each of the specific varieties observed.

Table 7.3 Number of arrival reports per variety harvested during the post-rainfall period

		Arrival Quality Scores			
General arrival quality	Variety	GREEN	AMBER	RED	**Grand Total**
Resilient	CSS	31	13	2	**46**
	I75	6	7	4	**17**
	RGT	2	4		**6**
	S35	2	1	1	**4**
Vulnerable	IF6		13	51	**64**
	JBS		5	9	**14**
	MLY	2	1	5	**8**
	TAW			5	**5**

Source: Author's own

7.1.6 Packaging decisions

The packaging variables for the dataset consisted of the internal (primary) packing and external (secondary) packaging. The primary packaging comes into direct contact with the product and is the level of packaging that ends up on the retail store shelves. The secondary packaging holds several primary packages (Saghir, 2004). The secondary packaging or carton is the smallest granular unit of measure used in this study.

The two packaging variables are denoted as V9 (NOM_pack) and V10 (NOM_inv_code), which are referred to as internal (V10) and external packaging (V9) from this point onward.

The relationships between the two packaging variables and the arrival quality are visible in *Figure 7.12*, which compares the proportion of quality scores for each packaging type.

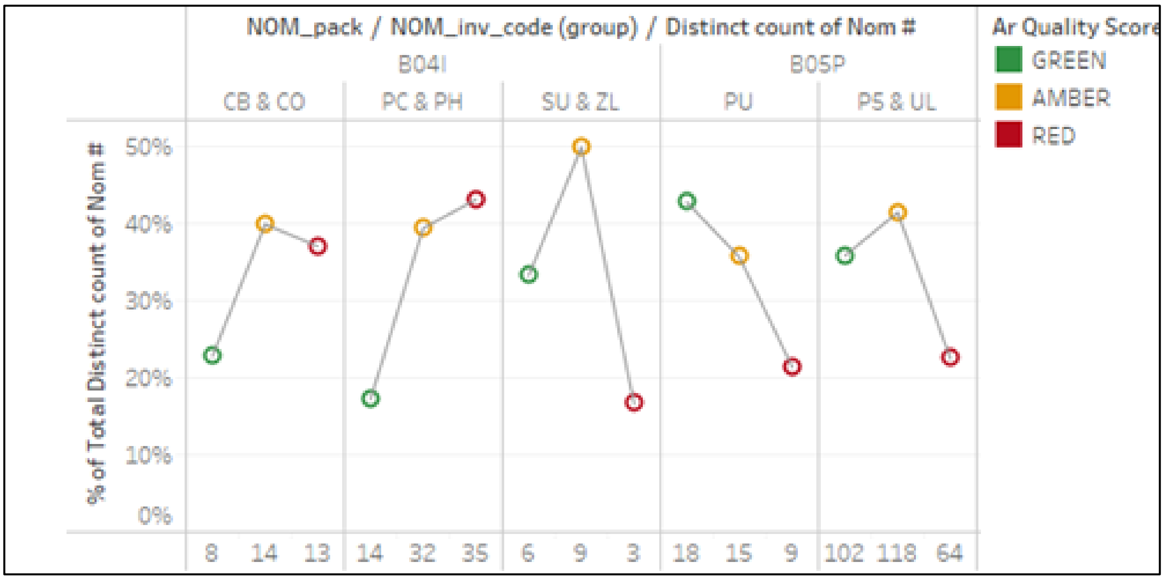


Figure 7.12: The proportion of arrival qc scores (green, amber, red) according to the external packaging and inventory (internal packaging) types

Source: Author's own

The two external packaging classes in *Figure 7.12* constituted 460 of the 467 observations from the dataset, which were B04I (134) and B05P (326) respectively. The internal packaging was subdivided into five groups based on packaging similarity, however, the distribution for each tended to relate to the type of external packaging to which it was applied.

As discussed in *section 7.1.4*, fruit packed post-rainfall from weeks 4 to 8, was negatively impacted by the pre-harvest climatic conditions, leading to an increased level of post-harvest decay. *Figure 7.13* displays the proportion of arrival QC scores based on the external packaging type. Two filters were applied, panel A only displays fruit packed in weeks 4 to 8, whereas panel B displays the entire seasons excluding those weeks.

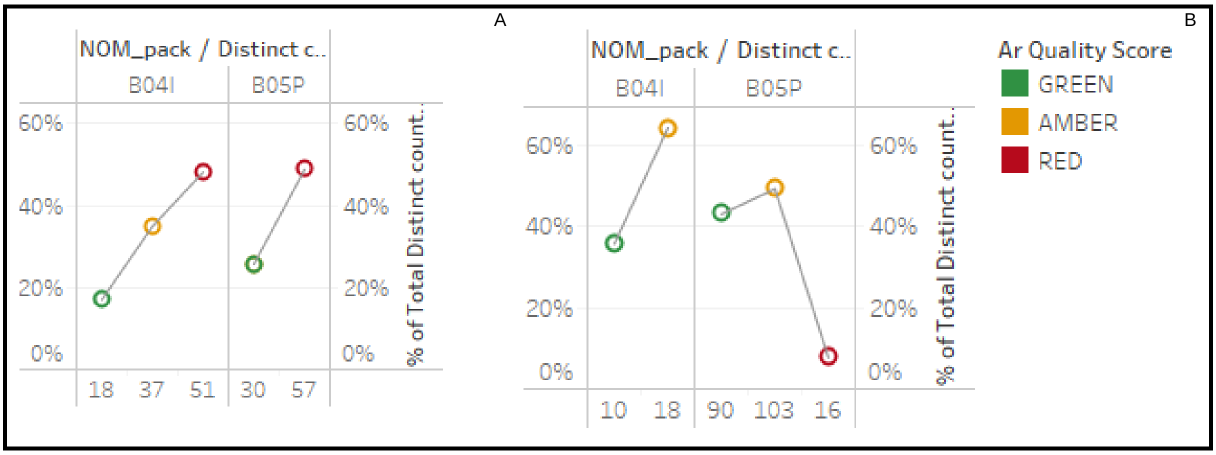


Figure 7.13: The proportion of arrival qc scores (green, amber, red) according to the external packaging filtered according to pack weeks (A) Week 4 to 8, (B) entire season excluding weeks 4 to 8

Source: Author's own

Isolating the rainfall period better illustrates the packaging decisions' comparative effect on table grapes that have undergone different pre-harvest conditions. The effect of the rainfall is evident in panel A, with roughly 50% of all grapes packed scoring red on arrival.

The effect of the packaging decisions is further discussed in *section 8.1*.

7.2 Inferential statistics

Inferential statistics are used to make inferences to the larger population based on statistical analysis applied to data gathered on the sample ("Inferential Statistics", 2022; Zikmund *et al.*, 2013: 507). Empirical testing usually involves inferential statistics, which tests hypothesis involving one (Univariate statical analysis) or more variables (bi / multivariate statistical analysis). This study applied correlation statistics to determine the inferences between variables.

Correlation coefficient is a single number ranging for -1 to 1. In absolute value terms, a correlation of 1 describes a perfect correlation, implying a change in one variable is directly associated with an equivalent change in the other variable. A correlation value of 0 indicates no meaningful association between the two variables. A corelation can either be positive or negative and indicates the direction of the association (Wheelan, 2014). The degree of association between variables can be described as follows:

Strong association +/- $0.7 \leq |r| \leq 1$

Moderate association +/- $0.3 \leq |r| \leq 0.7$

Weak association +/- $0 \leq |r| \leq 0.3$

There are three methods to check the linear relationship of two variables, namely Pearson, Spearman, and Kendall correlation techniques (Ye, 2020). Pearson's correlation is the most used of the three and is applied to this study's dataset.

To determine the relationship between nominal variables, a Cramér's *V* correlation statistic was applied.

7.2.1 Cramér's *V*

A Cramér's *V* was chosen as it allows the measurement of intercorrelation between two nominal variables or higher.

A Cramér's *V* measure was conducted for both the external (NOM_pack) and internal packaging (NOM_inv_code) variables to determine the level of association for each in relation to the arrival quality scores. The results for external packaging were 0.119 and 0.132 for the internal packaging. The scores indicated a weak correlation between packaging and arrival quality.

A Cramér's *V* was also conducted for the nominal variety variable (NOM_variety) and the arrival quality scores. The Cramér's *V* for the two variables scored 0.435, which indicates a medium level of association.

7.2.2 Pearson's correlation

A pairwise correlation matrix, seen in *Table 7.4*, was performed to select robust features. This is done to detect highly correlated variables (features), which add complexity to a model, increasing the chance of overfitting, while adding no beneficial information.

Table 7.4: Pearson's correlation matrix

		V51	V52	V53	V54	V57	V58	V59	V60	V61	V63
	Y / X	AR_Soft berries %	AR_Shrivelling	AR_Packing/handling damage	AR_Marks	AR_Decayed berries*	AR_Bruising	AR_Berry drop	AR_Berry drop weight (grams)	AR_Berry cracking	AR_QUALITY SCORE
V18	VIN_Brix readings - MIN	-0.12	-0.14	-0.13	0.06	-0.18	0.07	**-0.21**	-0.02	-0.16	**-0.24**
V19	VIN_Brix readings - MAX	-0.10	-0.02	-0.16	-0.01	-0.15	0.11	-0.16	0.02	-0.11	-0.16
V20	VIN_Brix readings - AVERAGE	-0.11	-0.06	-0.11	0.00	-0.16	0.13	**-0.23**	-0.02	-0.11	**-0.21**
V21	VIN_sum.Rain_LF (mm)	**0.24**	-0.09	0.04	0.06	**0.31**	-0.17	0.10	-0.05	**0.23**	**0.20**
V22	VIN_ET0_LF (mm)	-0.12	**0.28**	0.12	-0.07	-0.03	-0.14	0.11	0.01	-0.02	0.04
V23	VIN_av.RH_LF (%)	**0.26**	-0.10	0.06	0.14	**0.30**	-0.14	0.05	-0.04	0.17	**0.23**
V24	VIN_av. Tn_LF (˚C)	0.00	0.06	0.00	0.11	-0.08	0.02	-0.09	-0.05	0.05	-0.04
V25	VIN_av. T_LF (˚C)	-0.17	0.12	-0.02	0.00	**-0.23**	0.13	-0.06	-0.02	-0.12	-0.12
V26	VIN_av. T (˚C)	-0.04	0.16	0.04	-0.09	0.05	0.04	0.10	-0.04	-0.03	0.09
V27	IN_Split berries*	**0.28**	0.02	-0.04	0.10	0.08	-0.08	-0.05	0.01	0.01	0.08
V28	IN_Soft berries*	0.11	**0.21**	0.03	-0.06	0.00	-0.05	0.02	**-0.20**	0.03	0.13
V29	IN_Slip skin*	0.14	0.07	0.02	0.00	0.10	-0.07	0.00	-0.17	0.04	0.11
V30	IN_Neck splits*	0.13	0.09	0.00	0.05	0.08	-0.14	0.09	**-0.21**	0.04	0.11
V31	IN_Marks*	-0.10	-0.08	0.04	-0.01	-0.03	0.02	0.02	0.01	0.04	-0.04
V32	IN_Decayed berries*	0.10	**0.23**	0.05	0.01	**0.23**	-0.13	-0.02	0.02	-0.05	**0.24**
V33	IN_Berry drop*	0.00	0.17	0.04	-0.03	0.09	-0.09	0.08	0.04	-0.03	0.15
V46	VOY_Spikes (SUM)	-0.06	-0.04	0.00	-0.01	-0.04	0.05	-0.05	0.00	0.00	-0.03
V47	VOY_Breaks (SUM)	-0.06	-0.04	0.00	0.00	-0.04	0.04	-0.05	0.00	0.01	-0.03
V48	VOY_Fruit age at QC*	-0.15	-0.04	-0.09	0.07	-0.02	-0.01	-0.04	-0.02	-0.12	-0.05
V45	IN_QC Score	0.16	0.09	-0.03	-0.02	0.14	-0.14	0.05	-0.10	0.04	**0.20**

Source: Author's own

The correlation matrix displays all *x* input feature variables in the vertical column, coloured blue. All *y* output variables (response) are seen in the horizontal row, coloured yellow. The relationship between each input and response is displayed where the two features intersect.

7.2.2.1 Pre-harvest climatic variables

Climactic input variables (denoted with the prefix 'VIN') V21 – V26, described in *Table 5.2*, display the greatest number of correlations out of all the input variables available. The two most important climactic input variables are V21, (VIN_sum.Rain_LF [mm]), which measured the cumulative rainfall (mm) for the two weeks prior to harvest, and V23, (VIN_av.RH_LF [%]), which measured the average relative humidity (%) for the two weeks prior to harvest.

7.2.2.2 °Brix at harvest

Input variables V18 – V20 from the correlation matrix indicated a negative correlation between °Brix and arrival quality score. This implies that un-ripe grapes are more likely to arrive in poor quality compared to optimally ripe grapes with higher °Brix readings. The relationship is further discussed in *section 8.2*, in the following chapter.

7.2.2.3 Supply chain impact

Variables that measure the cold chain performance consisted of input variables V46 to V48 (denoted with the prefix 'VOY'). As stated in *Section 5.3.3,* data consisted of the frequency, or count of occurrence, of temperature spikes and breaks recorded for each container (V46 & V47), as well as the total fruit age from time of harvest to arrival quality surveyance (V48), as seen in *Table 7.4*. The correlation matrix indicates no correlation between the supply chain variables and the arrival quality scores.

7.3 Model evaluations

As stated in this chapter's (data analysis) introduction, the descriptive and inferential statistics applied aided in model and feature selection decisions. *Chapter 6* describes the modelling process, and the results are discussed below.

Once the machine learning models were tuned and trained, the model with the best performing cross-validation score was selected for further evaluation, discussed in the following sections.

The Grid-search best scores show the mean cross-validation scores for each type of model trained. The cross-validation -scores can be seen in *Table 7.5*.

Table 7.5: Cross-validation scores for each model trained

Logistic Regression	0.482486
K-Nearest Neighbours	0.520216
Decision Tree	0.498775
Random Forest	0.611351

Source: Author's own

According to the cross-validation score results, the more complex random forest classifier yielded the highest score, making it the optimal model. The RF classifier was, therefore, retained for further evaluation.

7.3.1 Multi-class confusion matrix

A multi-class confusion matrix was extracted for the Random Forest model's test set results. The confusion matrix for each class-label (green, amber, and red) is interpreted according to *Figure 7.14*, which indicates how the confusion matrix is to be interpreted depending on the class of interest.

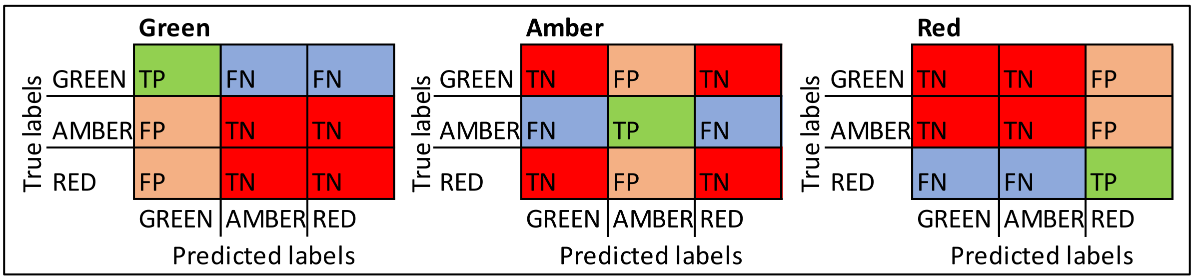
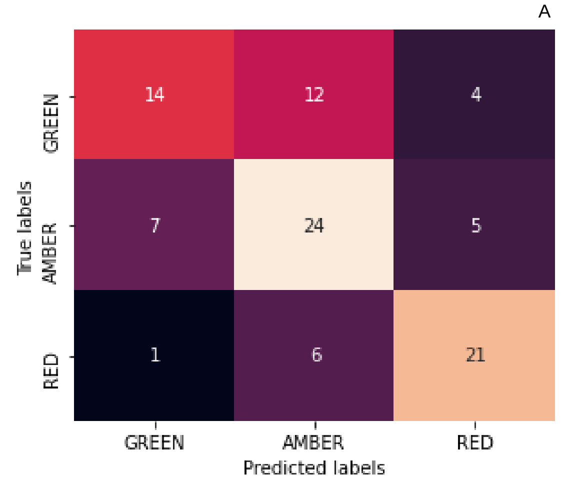


Figure 7.14: Interpretation of a confusion matrix for a 3-class classification task

Source: Author's own

Figure 7.15 shows the confusion matrix for the RF classification model (panel A), and panel B summarises the results of how successful the model was at predicting the three class-labels compared to the true labels for the test set. There were 94 observations used for model testing.



B

	GREEN	AMBER	RED
TP	14	24	21
	24+5+6+21 =	*14+4+1+21 =*	*14+12+7+24 =*
TN	56	40	57
FP	*7+1* = 8	*12+6* = 18	*4+5* = 9
FN	*12+4* = 16	*7+5* = 12	*1+6* = 7
	94	**94**	**94**

Figure 7.15: multi-class confusion matrix from the categorical response variable according to the Random Forest model

Source: Author's own

The results in panel B of *Figure 7.15*, indicate that red class-labels were correctly predicted most often (83% true to 17% false), then green label (74% true to 26% false), and amber labels with the largest proportion of false predictions (68% true to 32% false). This shows that the model performs best for predicting red labels and worst for amber labels.

More in-depth analysis of the confusion matrix is discussed in the following section.

7.3.1.1 Performance metrics

A classification report is a collection of performance metrics used to measure the quality of predictions from the classification model. For a three-class classification problem, the random chance of correctly predicting the class is a third (33.3%) and is, therefore, the baseline to compare the results against.

The report shows the classification metrics precision, recall and f1-score on a per-class basis as well as the overall scores for the model, indicated in the weighted average row. *Table 7.6* summarises the performance metrics for the random forest classifier.

Table 7.6: Classification report

	precision	recall	f1-score	Support
Green	0.64	0.47	0.54	30
Amber	0.57	0.67	0.62	36
Red	0.7	0.75	0.72	28
Accuracy			0.63	94
weighted avg.	0.63	0.63	0.62	94

Source: Author's own

Note that the support column in *Table 7.6*, indicates the predicted observations distribution for each class from the test set. This indicates how balanced or skewed the dataset is. The test set for this study is balanced with the proportion of observations for each class roughly equal to a third, used to calculate the weighted average for each performance metric.

The precision score indicates how well the model correctly predicted that an observation was actually in that class. According to the classification report, the model struggled to correctly predict amber observations, with a score of 0.57. This implies that there are many false positives (FP) for the amber class.

The recall score indicates ratio of correct positive predictions to the number of actual positive observations for each class. The model had many false negatives (FN) when predicting the green class, resulting in a score of 0.47. False negatives occur when it is predicted that the observation isn't in that class when it actually is.

The F1 score effectively shows the model's total prediction error for each class. A perfect F1 score would equal 1, where there are no FP and no FN predictions for that class. The green class predictions have the most errors, with an F1 score of 0.54.

7.3.2 ROC curves and AUC scores for each class

As stated in *section 3.6.4*, the receiver operating characteristic (ROC) curve is a two-dimensional graph that plots the trade-off between the true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis according to changes in the decision threshold.

The ROC curve is evaluated by calculating the area under the ROC curve (AUC), which is an aggregate measure of performance across all possible classification thresholds. The AUC score effectively shows how well the model distinguishes between the different classes.

Figure 7.16 is a graphical illustration of the ROC curve for each of the three class-labels for the response variable (green, amber, and red), and the AUC scores for each class using the one-vs.-rest evaluation technique.

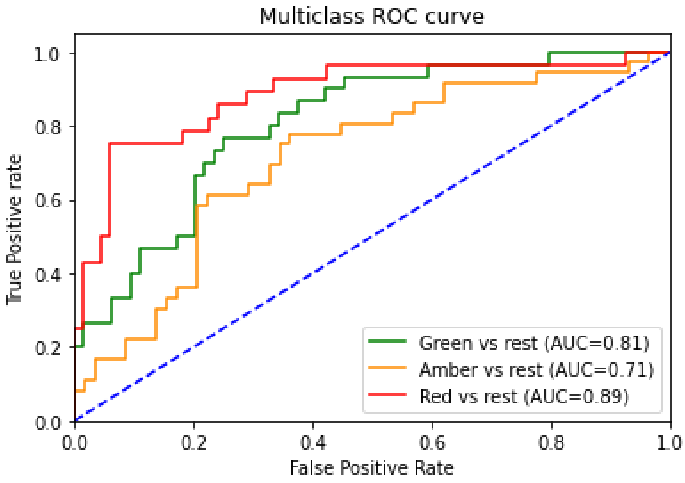


Figure 7.16: ROC curves and AUC scores for the three classes

Source: Author's own

The RF model's AUC scores indicate that the model is best at predicting red class-labels, with an AUC score of 0.89, followed by the green class at 0.81, and performed worst at predicting the amber class, at 0.71.

To understand the AUC scores, the red class AUC score is used as an example. The score of 0.89 implies that there is an 89% probability that a randomly selected red-class observation will actually be assigned to the red class instead of being assigned to one of the other two classes.

The weighted average AUC score for the RF model was 0.802, indicating that the model performs quite well at distinguishing between classes, however, there is still room for improvement.

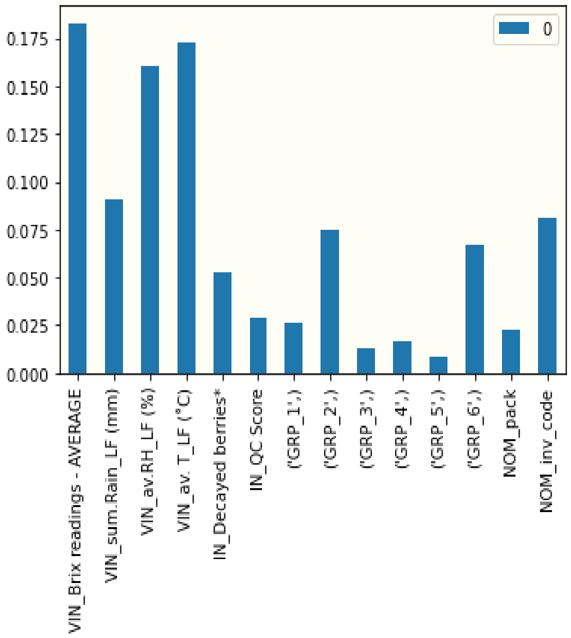
7.4 Understanding the model

As stated in *section 3.6.1*, complex ensemble models, such as the RF classifier, suffer from limited interpretability in terms of understanding the relationship between the feature inputs and the response output. Two methods, which are feature importance and Partial Dependence Plots (PD plots), are employed to describe how the model came to making the predictions.

7.4.1 Feature importance

The importance of each feature is determined by the mean decrease in the Gini index, averaged across all trees in the RF model. The metrics are determined by iteratively removing each feature to measure the impact on the model's prediction performance.

The feature importance for the RF model, based on the test-set results, can be seen in *Figure 7.17*.



REF#	Features	Importance %
V20	VIN_Brix readings - AVERAGE	18.3%
V25	VIN_av. T_LF (˚C)	17.3%
V23	VIN_av.RH_LF (%)	16.1%
V21	VIN_sum.Rain_LF (mm)	9.1%
V10	NOM_inv_code	8.1%
V6-2	('GRP_2',)	7.5%
V6-6	('GRP_6',)	6.7%
V32	IN_Decayed berries*	5.3%
V45	IN_QC Score	2.9%
V6-1	('GRP_1',)	2.6%
V9	NOM_pack	2.3%
V6-4	('GRP_4',)	1.7%
V6-3	('GRP_3',)	1.4%
V6-5	('GRP_5',)	0.8%
		100.0%

Figure 7.17: Most important features according to the RF model

Source: Author's own

1. **Preharvest climatic conditions and ˚Brix reading at harvest**

   The feature importance for the climatic variables measure were V21 with 9.1%, V23 with 16.1%, and V25 with 17.3%. The three variables accounted for 42% of the model's feature importance. The three variables measure different climatic conditions for the two-week period pre-harvest.

   The most important feature for the RF model was V20, the average ˚Brix measured at harvest, with a 18.3% importance score. The relationship between the feature and the model is discussed in the following section.

2. **Table grape varieties**

   As stated in *section 6.1.2*, the nominal categorical features needed to be encoded for model training and testing. This was done for the variety feature by ranking all table grape variety observations according to the perceived long-term quality or storability, resulting in six ordinal groups. A feature was created for each group, labelled GRP_1 to 6.

Each of the six variety features do not seem to be important to the model on their own, but when viewed as a whole, accounted for 20.7% of prediction performance. The most important variety subgroups consisted of groups 1 (2.6%), 2 (7.5%), and 6 (6.7%). The varieties within the three most important groups can be seen in *Table 7.7*.

Table 7.7: varieties present in the three most important variety groups according to the RF model

RANKING	VARIETY NAME	VAR	OBSERVATIONS	OBS. PER GROUP
1	SWEET JOY	I17	39	59
1	SABLE	SGS	2	
1	SWEET CELEBRATION	I75	18	
2	CRIMSON	CSS	210	213
2	LA ROCHELLE	LAR	2	
2	ALPHONSE	ALP	1	
6	TAWNY	TAW	7	87
6	REGAL	RGT	7	
6	SUNDANCE	SDC	5	
6	SWEET SAPPHIRE	IF6	68	

Source: Author's own

It must be noted that the varieties Crimson (210) and Sweet Sapphire (68) constituted 60% of all observations for the entire dataset (467). This might have been the reason for the higher level of feature importance for groups 2 and 6 respectively.

3. **Post-harvest QC features**

Features that measured the post-harvest quality were the packhouse QC score (V45) with 2.9% importance, and count of decayed berries (V32) with a 5.3% importance. This could indicate that either the factors that impact long term quality cannot be resolved by improved packing standards, or that the current packhouse QC measures need improving which might yield better prediction accuracy scores.

The packaging features (V9 and V10) had a 10.3% importance score.

7.4.2 Partial dependence plots

Partial dependence (PD) plots were graphed based off the model that yielded the best prediction score, the Random Forest (RF) model. As stated in *section 3.6.4*, from the literature, complex models, such as the RF classifier, offer the researcher limited understanding into the relationship between the features selected and the response variable. PD plots are a technique used to show how each feature is related to the outcome class, with all other features held constant (Berk & Bleich, 2013: 538).

One-way PD plots were created per features according to the probability of class-label (green, amber, and red) adherence for the predicted response variable, the arrival table grape quality score. The vertical y-axes denotes the probability of class adherence, whereas the x-axes displays observations of the chosen feature. Data deciles are indicated as tick marks on the x-axis. A decile rank arranges the data in order from lowest to highest and is done on a scale of one to ten, where each successive number corresponds to ten percent of the observations. One-way PD plots for the RF model can be seen in *Figure 7.18*, *Figure 7.19*, and *Figure 7.20* respectively.

7.4.2.1 PD plots for the average ˚Brix, total rainfall two weeks prior to harvest, and average RH two weeks prior to harvest

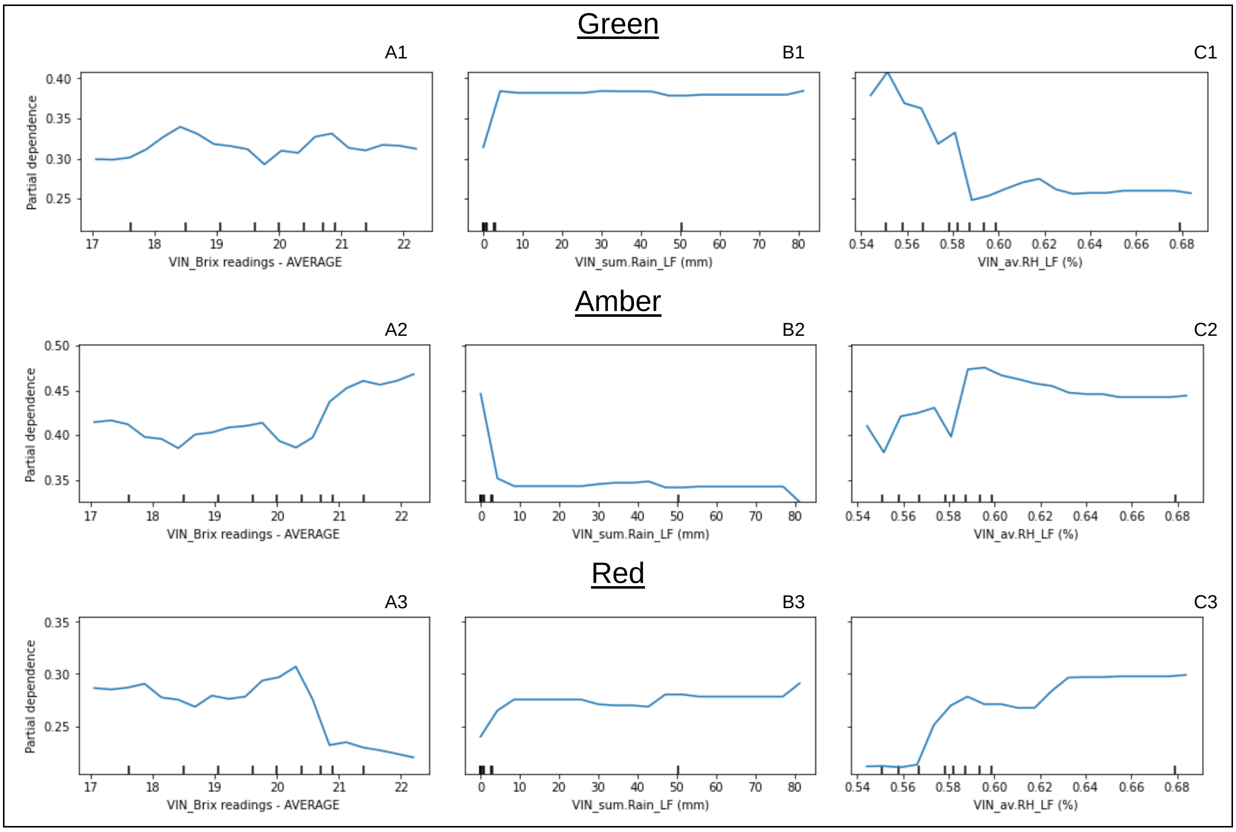


Figure 7.18: Partial dependence plots for three features: average ˚Brix at harvest (A1-3), total rainfall two weeks prior to harvest (B1-3), and the average RH two weeks prior to harvest (C1-3)

Source: Authors own

1. **Average ˚Brix at harvest**

According to *Figure 7.18*, the optimal ˚Brix for green scores was around 18.5˚ (A1), for readings with greater degrees seemed to have little effect on the probability of green class predictions. The probability of an amber score increases for ˚Brix greater than 20˚ (A2), whereas the probability of a red score decreases for ˚Brix greater than 20˚ (A3). The relationship between ˚Brix and quality scores seems to be positive for amber scores, and negative for red scores.

2. **Total rainfall two weeks prior to harvest**

Total rainfall prior to harvest is displayed in plots B1, B2, and B3 in *Figure 7.18*, and demonstrated a confounding relationship toward predicted arrival quality scores. The PD plots displayed a positive relationship between rainfall for green and red scores, and a negative relationship for amber scores. The relationship between rainfall and arrival quality agrees with the literature for amber and red predictions, but is contradictory for green score predictions.

It must be noted that eight of the nine decile tick-marks for plots B1, B2, and B3 are tightly grouped around 0 on the x-axis. This means that 80% of observations in the test set had little to no rainfall two-weeks prior to harvest. This possibly led to a distortion of the relationship due to the influence of outlier observations.

3. **Average RH two weeks prior to harvest**

The average Relative Humidity (RH%) two weeks prior to harvest is displayed in plots C1, C2, and C3 in *Figure 7.18*. The relationship between RH and predicted arrival quality scores found a negative relationship for green predicted class-labels and a positive relationship for both amber and red classes. The findings imply that as RH increases, the probability of a negative arrival score increases, i.e., amber, or red class-labels. The general relationship agrees with the literature in that high RH creates the ideal environment for pathogen germination, most likely leading to the presence of decay post-harvest, in turn leading to poor arrival table grape quality.

7.4.2.2 PD plots for the average daily temperature two weeks prior to harvest, the number of decayed berries found in the packhouse quality reports, and the packhouse intake quality score

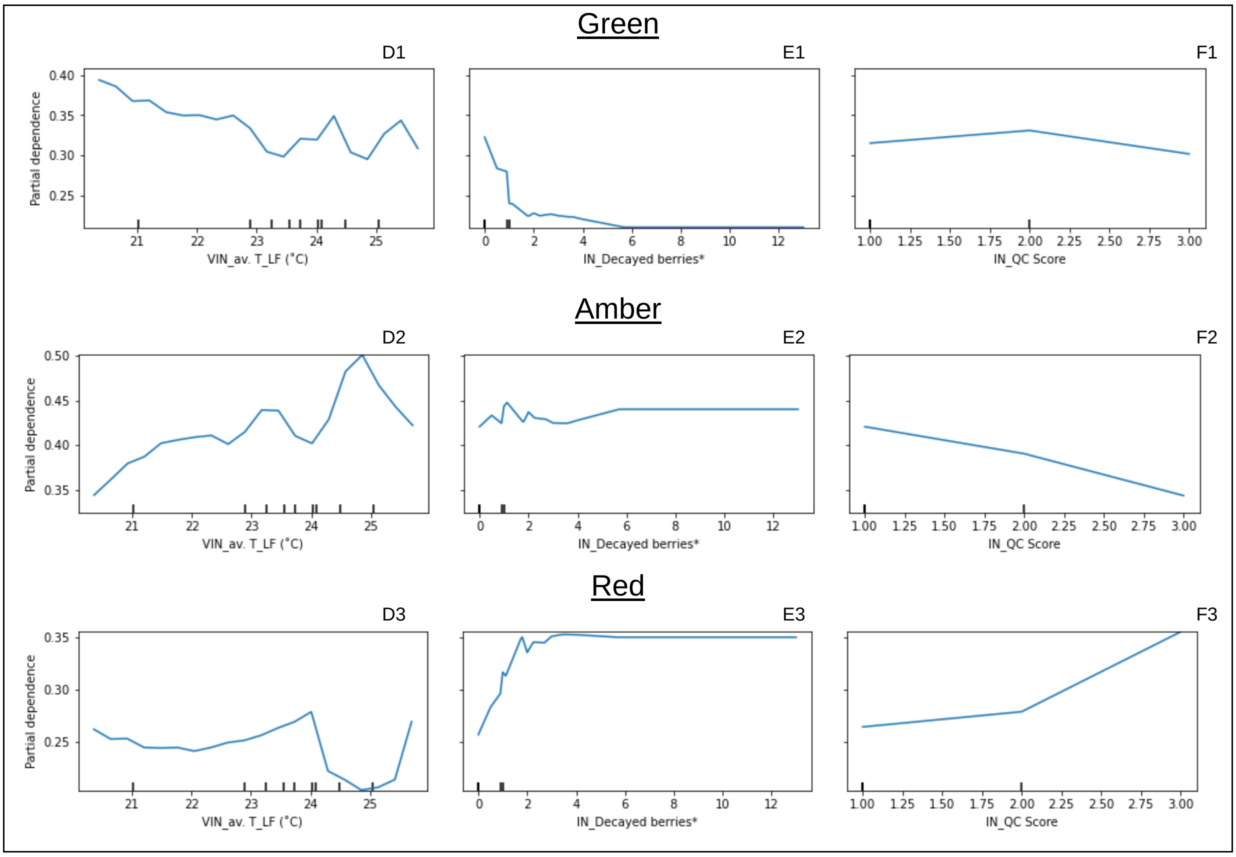


Figure 7.19: Partial dependence plots for three features: the average daily temperature two weeks prior to harvest (D1-3), the number of decayed berries found in the packhouse quality reports (E1-3), and the packhouse intake quality score (F1-3)
Source: Author's own

4. **Average temperature two weeks prior to harvest**

The average Relative Humidity (RH) two weeks prior to harvest is displayed in plots D1, D2, and D3 in *Figure 7.19*. Considering all the line graphs together, it seems that as average temperature two weeks prior to harvest increases, the probability of a negative arrival score increases, i.e., amber, or red class-labels. This relationship is especially true for amber class-label predictions (D2), where the probability increased by 5% for temperature averages from 21 to 23˚C, and the prediction probability increased by another 10% for average temperature between 23 and 25˚C, where half of all test observations were present. For average temperatures greater than 25˚C the probability of a red class-label prediction increased by 12% (D3), in contrast to the other two class-labels (D1 and D2). It must be noted that only 10% of observations were greater than 25˚C.

5. **Number of decayed berries found in the packhouse quality reports**

The effect of the number of decayed berries found in packhouse QC reports on arrival quality score class-labels is displayed in plots E1, E2, and E3, in *Figure 7.19*. Increasing the number of decayed berries found, had a negative relationship toward the probability of green class-labels (E1), no effect on amber class-labels (E2), and a positive effect on the probability of red class-labels (E3). The feature's effect was most pronounced for the probability of the red class-label, which increased by 8%, as the number of decayed berries increased from zero to two.

6. **Packhouse quality control score**

The packhouse QC report score feature is a categorical variable based on a three-point Likert scale to rate quality and condition of the packed table grapes prior to cold storage. The class-labels are the same as the response variable (green=1, amber=2, red=3).

The effect of the packhouse QC score feature on the response variable's predicted class-label is displayed in plots F1, F2, and F3, in *Figure 7.19*. The packhouse QC score feature seemed to have a small effect on the green predicted responses, seen in F1, but the relationship seemed to be more pronounced for amber and red response class-labels. The PD plot indicated a negative relationship for the probability of amber class-label outputs (F2), with a probability reducing by 3% from green (1) to amber (2) and an additional 4% reduction from amber (2) to red (3) inputs. The probability of red output class-labels displayed a positive relationship to the input QC scores (F3), where the probability increased by 3% from green (1) to amber (2) intake scores and an additional 7% increase in probability if the input score changed from amber (2) to red (3).

7.4.2.3 PD plots for the three most important table grape variety binary groups

The table grape variety feature was split into six groups according to the perceived long-term storability of each variety, visible in *Table 6.3*. Varieties included in group 1 consisted of Sweet Joy, Sable, and Sweet Celebration. Group 2 varieties were Crimson, La Rochelle, and Alphonse. Group 6 varieties were Tawny, Regal, Sundance, and Sweet Sapphire.

The binary features consisted of a 1 if the variety from that group was observed, or a 0 if absent. PD plots for the three most important variety groups (Group 1, 2, and 6) can be seen in *Figure 7.20*.

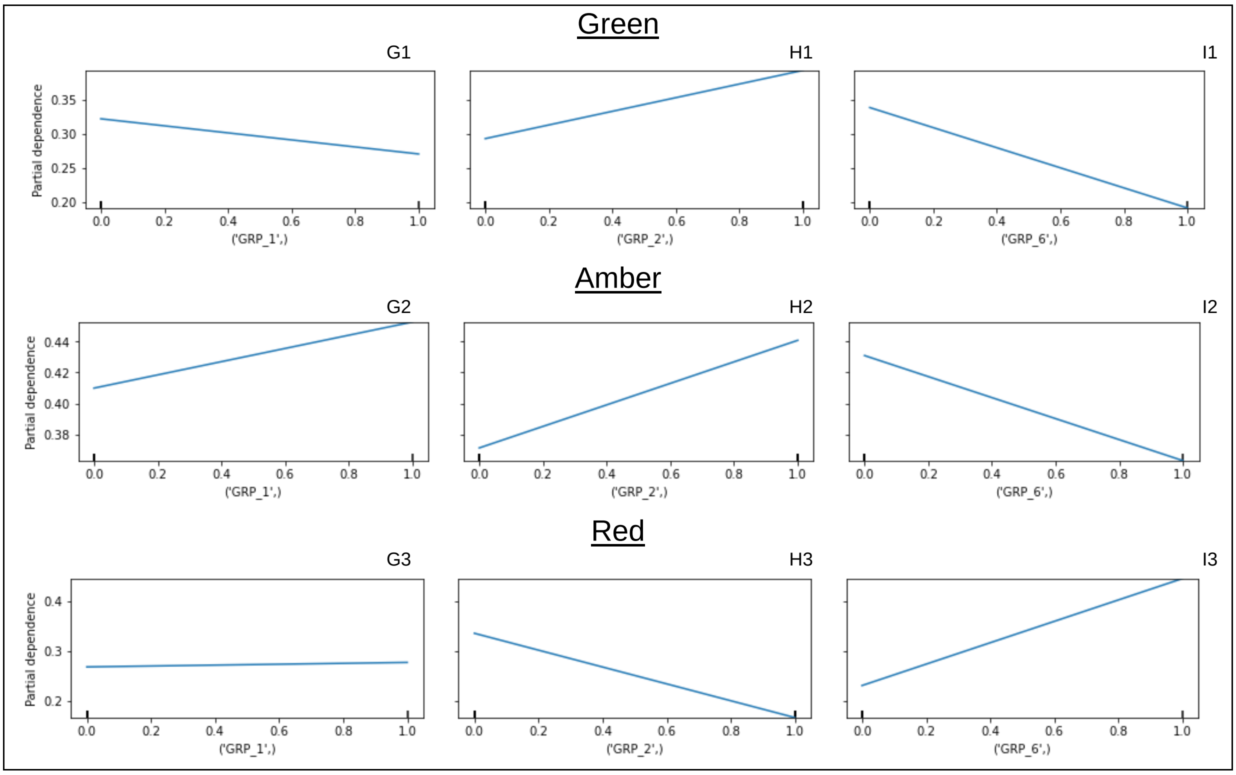


Figure 7.20: Partial dependence plots for three features: Table grape variety group 1 (G1-3), group 2 (H1-3), and group 6 (I1-3)

Source: Author's own

7. **Three most important table grape variety groups**

   The effect of the table grape variety on the predicted response variables' class-labels can be seen in *Figure 7.20*, (G1-3) for group 1, group 2 (H1-3), and group 6 (I1-3).

   Group 1 varieties, displayed in G1-3, had a modest impact on the response class-labels, with a negative relationship to green label predictions, positive to amber, and no relationship to red label predictions. The implications of the PD plots are that the group 1 varieties might not be the best varieties due to the negative relationship toward green class-labels but are still relatively storable due to the results of G3, having no relationship to the probability of red class-labels.

Group 2 varieties, displayed in H1-3, had a strong relationship to the response class-labels, with a positive relationship to green label predictions, positive to amber, and negative relationship to red label predictions. This indicates that group 2 varieties have good long-term storability and are more likely to arrive in a good condition (green class-labels).

Group 6 varieties, displayed in I1-3, had a strong relationship to the response class-labels, with a negative relationship to green label predictions, negative to amber, and positive relationship to red label predictions. Group 6 varieties had the strongest prediction probabilities across all three class-label outcomes, with a 12% decrease in green labels, 7% decrease for amber labels, and 20% increase in probability for red class-labels depending on whether or not group 6 varieties were observed.

7.5 Research reliability and validity

As stated in *section 4.8*, research reliability and validity are used to evaluate the quality of the research conducted (Bryman & Bell, 2015 and Zikmund *et al.*, 2013).

Reliability is concerned with whether the findings are replicable if the same collection techniques and analysis procedures were employed.

Test-retest reliability is concerned with whether there is consistency in the measures used. The methodology applied in the modelling process is well established and has been used in a range of studies from various fields. The study is exploratory and frames table grape quality as a classification task. The reliability of the classification tasks results can be assessed by comparing the predicted output to the actual output. The Random Forest model reliably predicted the red class-label output but underperformed for the amber and green class, visible in the classification report in *Table 7.6*.

Research validity is concerned with the integrity of the conclusions generated by the research (Bryman & Bell, 2015: 42). The types of validity evaluated are internal validity, external validity, and content validity.

Content validity is concerned with whether the measures and variables used in the research cover all the content in the underlying construct (what is trying to be measured). The construct being studied are the causes of poor arrival quality of table grapes. This construct is broad, which is why the study is exploratory in nature. Through the application of machine learning to predict the arrival quality, features were selected for modelling. If the features selected are not related to the response variable, the prediction probability would be no better than chance. The model evaluation discussed in *section 7.3*, indicates content validity for the measures and variables selected.

Internal validity refers to the extent to which the independent variables accurately produce the observed effect in the dependent variable. This is achieved by the application of inferential statistics to determine the degree of intercorrelation between variables. Variables which exhibited a degree of association to the dependent (response) variable, were used in the modelling process for the classification task. The Random Forest classifier performed relatively well when predicting the arrival class-label based off the features selected, therefore, the study has internal validity.

The study possibly lacks external validity as only one season was used in the data analysis process, limiting the generalisability of the results beyond the dataset. The 2020 grape season suffered from unseasonal rainfall, contributing to poor arrival quality of the table grapes. It is unknown if the effect of some variables is overstated, limiting the study's generalisability.

7.6 Conclusion

This chapter illustrates how the three different analysis techniques applied (descriptive, inferential, and machine learning) build on one another to give a more complete understanding of the relationships between the variables. The implications of the findings are described in the following chapter.

Chapter 8: Interpretation of Results

The results for various levels of data analyses utilised in *Chapter 7* are brought together in this chapter to provide a holistic interpretation. The interpretation is focused on explaining how and why variables relate to the poor arrival quality of table grapes.

Through the interpretation and discussion of the results, conclusions and recommendations can be made.

8.1 Packaging decisions

There are two potential reasons for the difference in the spread of QC arrival scores based on the packaging decisions. Firstly, the effect packaging has on packhouse behaviour and secondly, the effect which market supply levels have on arrival QC scores.

1. **Packaging's effect on packhouse behaviour**

   According to *Figure 7.13*, fruit packed in punnets (B05P) had a higher proportion of green arrivals across all pack weeks. The reason for this might stem from packhouse processes and a misperception of pack and product specification standards. Packing punnets (B05P) is a painstaking, tedious process, which requires attention to detail when clipping out rotten berries. In contrast, B04I cartons are packed far more quickly, where large bunches are often put directly into the carton without cutting them in two, to check if there are decayed berries within the cluster. In summary, the perception of packhouse managers and packers alike, is one that prioritises punnet (B05P) packing. This insight corresponds with the average number of decayed berries found in the packhouse QC reports, where B05P had an average of 0.3, whereas B04I had an average of 0.6 decayed berries found per carton sampled.

A portion of table grapes packed during the 2020 season arrived in the Netherlands between weeks 8 and 13. These dates coincided with the spread of the Covid-19 pandemic in Europe. According to Dutch news sources, from 12 March 2020 and onward, all public events where there were more than one hundred people, were cancelled. Three days later, on 15 March 2020, all restaurants and cafés were closed indefinitely in the Netherlands (NOS Nieuws , 2020). These public health containment measures effectively shut down the fresh fruit wholesale market.

Wholesale fruit is sold directly to the food service channel, consisting of restaurants, catering, cafés, and smaller fruit markets. This market channel constitutes roughly 25 to 30% of the fresh fruit and vegetable supply (OECD, 2020: 8). Part of this volume was rerouted into supermarket retail programs for at-home consumption. Unfortunately, the fresh produce supply chain was not designed for such large volumes to be rerouted within the timeframe that produce remains fresh. The effect of the pandemic is discussed below.

2. **Effect of market supply levels on arrival QC scores**

Exporters are of the opinion that when a market is under pressure, importers' QC standards become more stringent. The rationale is that importers explain the low sales price to be a result of poor arrival quality instead of an over-supplied / inflexible market. Fedeli (2019: 148), also referred to this phenomenon by stating that receivers are incentivised to portray the product in the worst possible light to pay a discounted price. This scenario played out in the 2020 season where the wholesale market collapsed due to the containment measures implemented to curb the spread of Covid-19.

This is a potential explanation for the greater proportion of B04I packaging red arrivals compared to the B05P packaging. B04I is a 4.5kg loose packaging used in the wholesale market, characterised by an open display carton where consumers can touch the fruit. This type of packaging would have become undesirable in the Covid-19 context, possibly causing importers to be more stringent when QC'ing B04I consignment arrivals. It must be noted, however, that this is merely an opinion held by exporters, there is no evidence to support this theory.

As stated above, the spread of the Covid-19 pandemic destroyed the wholesale market for the 2020 season. The response to the pandemic is visible in *Figure 8.1*, which illustrates how the proportion of B04I (wholesale market product) to B05P (retail product) arrivals changed month on month. The number of observations is indicated within the bars.

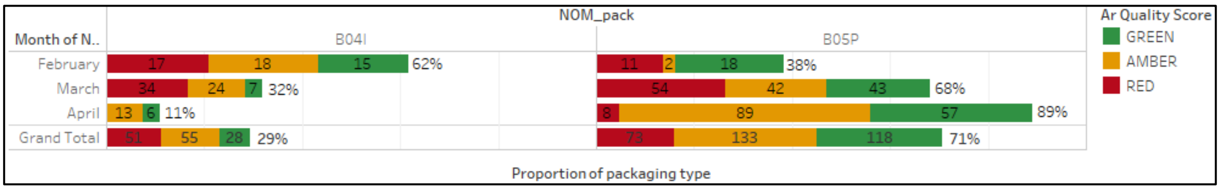


Figure 8.1: The proportion of monthly arrivals according to packaging type and quality score

Source: Author's own

From *Figure 8.1*, the change in the ratio of packaging type (B04I to B05P) is noticeable and corresponds to the outbreak of Covid-19. The February arrival ratio was 60:40, March 30:70, April 10:90 for B04I compared to B05P. This indicates that the exporters adapted their product offering based on consumer preference changes due to the Covid-19 pandemic.

8.2 Measure of grape maturity (ripeness) at harvest

As mentioned, table grapes are a non-climacteric fruit, meaning that it does not continue to ripen once harvested (Burger, 2000). The implications of this physiological process is that fruit harvested immature will have a poor eating quality, which will not improve, whereas over-ripe fruit at the point of harvest is prone to stem desiccation and *Botrytis* decay (Burger, 2000:123).

Table grapes are considered ripe when then TSS (Total Soluble Solids – °Brix), the measure of sugar content within the juice of the grape berry, is around 18 – 20 °Bx, variety specific*. Sugar levels continue increasing the longer the fruit remains on the vine. According to the feature importance metric, extracted from the RF classifier, variable V20 (VIN_Brix readings – AVERAGE) the average °Brix at harvest, is the most important variable used for the prediction of arrival quality, visible in *Figure 7.17*. As discussed in *section 7.2.2.2*, there is a negative correlation between the °Brix variable and arrival quality score, implying that unripe grapes (low °Brix) negatively affect the likelihood of sound arrival quality.

The Pearson's correlation presumes a linear relationship between variables, but the relationship between °Brix and arrival quality is most likely a quadratic relationship, where °Brix measures at harvest that are either too low (<16°) or too high (>23°) will both negatively impact the probability of sound table grapes on arrival. This is displayed in *Figure 8.2*, which indicates that both low and high ˚Brix readings result in high QC scores, implying poor quality on arrival. The lowest point on the figure is most associated with sound, green class, arrivals.

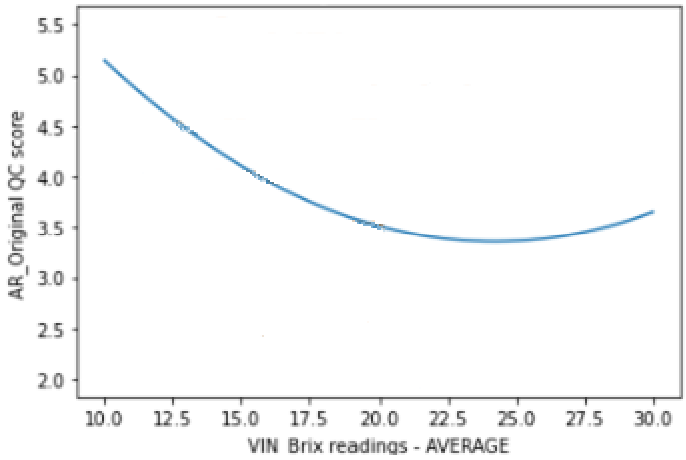


Figure 8.2: ˚Brix vs. Original QC arrival score

Source: Author's own

The likely explanation for the findings is that unripe fruit, with low sugars, will have a poor eating quality but the negative condition characteristics associated with overripe fruit will be reduced, whereas overripe fruit (with high ˚Brix readings) will more likely suffer from expected negative condition characteristics. The effect of fruit maturity is discussed further in *section 8.2.2.*

As noted in the study's limitations, *section 4.9.2.1*, due to the Covid-19 outbreak in South Africa, all packhouse quality controlling was suspended for fruit harvested in April. This resulted in a lack of late season data points, which are usually characterised with grapes being harvested over-ripe, with high ˚Brix readings. From the one-way PD plots, visible in *Figure 7.18* panel A, there appeared to be a positive relationship for amber class and a negative relationship for red class predictions in relation to increasing the ˚Brix. This relationship is also visible in *Figure 8.2*, where the quadratic relationship for high ˚Brix readings is not quite as pronounced as one would expect.

The implications may have led to a spurious relationship between high ˚Brix readings and arrival quality, which would have been more accurate if sampling bias had not occurred.

8.2.1 Climatic conditions pre-harvest

As discussed in *Section 2.10.2*, the ideal germination conditions for fungal pathogens, specifically *Botrytis cinerea*, are wet, warm, and humid conditions with infection occurring at open wounds in the berry skin. Open wounds can occur due to heavy rain when the turgor pressure within the grape becomes too great, resulting in splitting/rupturing of the skin. This is supported by the positive correlation between input variable V21 and output variables V27 and V30, seen in *Table 8.1*. The positive correlation implies that as the total rainfall for the two weeks prior to harvest increases, so would the number of split berries likely to be found in the packhouse post-harvest.

There did not seem to be a meaningful correlation between any of the environmental variables (V21 – V26) and decayed berries found in the packhouse reports (V32). This is most likely due to fact that a *Botrytis* is a post-harvest pathological disorder, which only becomes visible two to three weeks after infection, once the fruit has already been packed (García, 2022). Split berries are, therefore, a precursor to potential berry decay further down the supply chain, which would result in a reduced arrival quality score. This is supported by the Spearman's correlation coefficient result of 0.2 (a weak correlation), with a p-value of 0.000007, implying statistical significance.

Table 8.1: Correlation matrix - climatic conditions and shed QC predictors

		V27	V28	V29	V30	V32	V33
	Y / X	IN_Split berries	IN_Soft berries	IN_Slip skin	IN_Neck splits	IN_Decayed berries	IN_Berry drop
V21	VIN_sum.Rain_LF (mm)	0.36	-0.07	0.18	0.20	0.06	-0.08
V22	VIN_ET0_LF (mm)	-0.08	0.11	-0.13	-0.02	0.04	0.07
V23	VIN_av.RH_LF (%)	0.31	-0.08	0.21	0.19	0.10	-0.05
V24	VIN_av. Tn_LF (°C)	0.05	-0.01	-0.02	0.09	-0.02	0.01
V25	VIN_av. T_LF (°C)	-0.22	0.06	-0.12	-0.10	-0.04	0.09
V26	VIN_av. T (°C)	-0.06	0.16	-0.14	-0.06	0.04	0.04

Source: Author's own

The implications of the results are somewhat confounding. The correlation results show that heavy rain leads to more splits, which are infection points for spores. It could, therefore, be expected that there would be a strong negative correlation between intake split berries and the arrival quality, but according to *Table 7.4*, there is no real relationship between the two variables. This could imply a 'third-variable effect' impacting the relationship between heavy rainfall preharvest and poor quality on arrival.

8.2.2 Grape maturity and decay susceptibility

In *Section 2.11*, it was noted that mature berries tend to be more susceptible to fungal infection as naturally occurring fungistatic agents within the berry skin diminish with maturation. Secondly, as stated in *Section 8.2.1,* mature grape berries suffer from berry skin cracking or splitting due to increased turgor pressure caused by the sudden increased moisture availability, which is rapidly absorbed osmotically and through the root system. The sudden absorption of moisture results in the skin bursting leaving open wounds where fungal infection can occur. Immature grape clusters on the other hand, are less susceptible to infection. According to Vercesi, Locci and Prosser (1997:139), fungi require a carbon source to grow and colonise. The only carbon source available within immature grape berries are organic acid such as tartaric and malic acid, a poor energy source. These acidic compounds tend to decrease during the period of berry set (young berries enlarge – ELP 27) through to the onset of ripening (ELP 34) and are replaced by natural sugars such as glucose, fructose and sucrose, a more favourable carbon source. These sugars allow fungi to achieve greater growth rates and allow for an improved colonisation potential. The phenological stages of table grape maturation are summarised in *Figure 8.3*.

Phenological scale (ELP)	Phenological stage
1	Winter bud
4	Green tip, first leaf tissue visible
7	First leaf separated from shoot tip
12	5 leaves separated; shoots about 10 cm long; inflorescence clear
17	12 leaves separated; inflorescence well developed; single flowers separated
23	17-20 leaves separated; 50% caps off, full bloom
26	Cap-fall complete
27	Setting; young berries enlarging (> 2 mm diam.), bunches at right angles to stem
31	Berries pea-size (7 mm diam.)
32	Beginning of bunch closure, berries touching (if bunches are tight)
33	Berries still hard and green
35	Berries begin to colour and enlarge: veraison
38	Berries harvest-ripe (22°Brix)
39	Berries over-ripe
41	After harvest: cane maturation complete
43	Beginning of leaf fall
47	End of leaf fall

Figure 8.3: Table grape phenological scale (ELP)

Source: Coombe, 1995; Verdugo-vásquez et al., 2017

The reason why the heavy rainfall only affected fully mature table grapes was, therefore, due to the effect of turgor pressure causing open wounds, ideal for fungal infection and colonisation fuelled by the sugary food source. In contrast, unripe grapes, which are hard and small, are unaffected by rainfall and do not yet contain the sufficient nutrients for fungal growth. For these reasons the entire seasons crop was not lost.

8.3 Varieties

From *Figure 7.10* and *Figure 7.11*, and the Cramér's *V* score, it is evident that different varieties have a varying tolerance to rain prior to harvest. The identification and reasons are beyond the scope of this study, however, the implications for exporters and producers alike are evident.

The stakeholders involved in variety selection need to identify and plant varieties which are less susceptible to unsuitable preharvest weather conditions. The stakeholders also need to identify currently planted varieties, which require additional precautions to reduce potential losses. If these precautions are not implemented across the exporters total supply base, negative importer and consumer sentiment could become associated with those varieties, ultimately limiting market access. This negative sentiment can result in the exclusion of the low tolerance varieties to specific markets, which would limit exporters and producers market access options leading to additional exposure, loss in revenue, and big overhead expenses required to replace the unwanted variety.

According to planting trends for the Hex River region, illustrated in *Appendix L*, varieties such as IF6 (Sweet Sapphire®), the worst performing variety in this dataset (refer to *Table 7.3*), saw no additional vines being planted for the subsequent 2021 and 2022 seasons. Instead, there was a 32% reduction in IF6 vines planted year-on-year between 2021 and 2022. The implications of the mass plantings of IF6, 3-9 years ago and the subsequent removal of vines, results in a loss of income and future cashflow constraints for up to four years, two rounds of planting costs and two periods of vine establishment with no income. With the recommendations above, costly examples such as this could hopefully be avoided.

8.4 Supply chain impact

The correlation analysis conducted on the transit temperature input variables resulted in no meaningful association with any of the output variables. These findings contradict those of van der Klein (2018:86), who found a high correlation (0.6) between the number of temperature breaks and the negative status of the arrival quality report. It must be noted though, that van der Klein's study was constructed differently, consisting of 18 temperature recorders placed in 12 containers. As noted by van der Klein (2018:86) "…because of the small sample size more research needs to be done", no generalisable conclusions can be drawn from those findings as the sample size of 12 shipping containers does not meet the minimum requirement of 100 units (Zikmund *et al.*, 2013: 436).

The fruit age findings, illustrated in *Figure 7.5*, contradicted the literature and instead displayed a consistent distribution per fruit age buckets (roughly, 30% green; 40% amber; and 30% red). These findings indicated that voyage length did not contribute to poor arrival quality.

8.4.1 Temperature fluctuations

The greatest potential negative impact of temperature fluctuations, in combination with Relative Humidity, would be additional moisture created through condensation within the packaging liner due to a rise in temperature caused by a break in the cold chain.

The additional moisture coupled with the rise in relative humidity would activate the gas sheets (Tessara, 2022) within the packaging (used as a postharvest fungicide), resulting in additional production of SO_2. This SO_2 then mixes with the free moisture resulting in the formation of sulphite precipitate. Table grapes exposed to sulphate precipitate develop a phenomenon known as SO_2 burn, characterised by areas of discoloration (bleaching) and pitting on the berry surface, seen in *Figure 8.4*, (Langenhoven, 2022).

Figure 8.4: SO2 burn on 'Red Globe; table grapes. a) SO2 damage on undamaged grapes, b) SO2 damage on split berries.

Source: Langenhoven, 2022

Post-harvest storage temperature fluctuations and fruit age do not reduce decay but merely act as a retardant to the speed of decay growth and spread. As stated by Harvey (1955), the length of storage has less of an effect on post-harvest decay than the environment to which the fruit is exposed to preharvest, implying that preharvest infection prevention measures are more important than supply chain conditions. The only potential difference between longer and shorter transit times is the level of decay spread between bunches, also known as nested decay, whereby earlier intervention could reduce the spread of fungal growth between berries and adjacent bunches.

8.5 Model

When referring to the total error rate for each class-label prediction, the model did not perform well for amber (31% errors) and green (25% errors) class-labels. There are two potential solutions to this problem. Firstly, the model could be turned into a binary classifier by changing the response variable. Or secondly, the model could be modified to incorporate cost-sensitivity through threshold tuning.

8.5.1 Sensitivity vs. specificity: cost-sensitive models

As discussed in *section 3.6.4*, There are two types of error that can occur for classification prediction tasks. Type I or false positives (FP) error, predicting the event when there was no event or type II false negatives (FN) error, predicting no event when there was in fact an event. The type of error's importance to classification tasks are domain specific, implying that different tasks associate more importance to one type of error over the other.

For this business context, there are different costs associated with class-label misclassifications. If a red class sample is misclassified as one of the other two classes (known as a false negative), the fruit exported will arrive in a sub-optimal condition, resulting in a loss of income and possible additional costs. The optimal performing model for this task would, therefore, need to reduce the FN error to ensure that all red class fruit is correctly identified so that allocation decisions can be made that maximise financial returns. This can be achieved by applying threshold-tuning to the model during the validation process to include cost-sensitive measures.

This would, however, affect the trade-off between sensitivity and specificity, as the optimal model would need to increase the recall / sensitivity level to reduce the FN error. This would negatively impact the 1- specificity score, as well as, the ROC AUC score, which is a trade-off between two performance metrics.

8.5.2 Binary classifier

As stated in *section 7.3.1.1*, the precision score for the amber class was 0.57, which implies that the model struggled to correctly predict amber class observations, with many false positive (FP) predictions. This misclassification is detrimental to the accuracy of other class-label predictions, which are of greater importance from a financial context. A potential solution would be to select a binary response variable to be predicted, which removes the ambiguous amber class-label, which is of no relevance for to decision makers within the particular business context.

The model could be altered to rather predict variable V69 (AR_claim_type - data base) instead of the arrival QC score. V69 is a binary variable, which indicates the split between the proportion of observations where the arrival quality did and did not impact the final sales price. The split is exhibited in *Figure 7.2*, where 291 (62%) observations had a sound arrival, and 174 (38%) observations had a financial claim due to poor arrival quality.

8.6 Conclusion

The chapter illustrated that the cause of poor arrival quality consisted of a combination of preharvest, harvest, and postharvest factors all interacting. The net result is a deterioration of the product.

To summarise the findings, the preharvest climactic variables are of vital importance to the arrival quality of table grapes. The ideal conditions would see no rain, low relative humidity and temperatures that are warm but not too hot. The effect of wind was not directly measured and analysed in this study.

When to harvest, based on the level of the fruit's ripeness (measured by the ˚Brix), has a quadratic relationship to the arrival quality scores. This means that scores that are too low or too high result in an increased probability of poor arrival quality. Currently, the average ˚Brix at harvest is lower than what the analysis found to be optimal.

Chapter 9: Conclusion and Recommendations

The current table grape export supply chain sees a large proportion of consignments arriving in Europe in a suboptimal condition. Fruit arriving in poor quality incurs additional costs as the product needs to be reworked to appear presentable for the final consumer. Therefore, the aim of this study was to determine the factors attributing to the negative arrival quality of table grapes, and to what extent the factors play a mutual role in causing quality claims. In addition, machine learning modelling techniques were used to predict the arrival quality prior to export, which can aid exporters' allocation decisions, mitigating losses and reducing food waste. The effect of the supply chain was also considered and evaluated.

The objectives for this study were developed with the aim of providing an understanding of what the main factors were that lead to poor arrival quality, and if machine learning could in fact be applied to the classification task.

The literature review was conducted as secondary research and was split into two chapters: *Chapter 2* examined the essential theory related to table grapes and the cold chain whereas *Chapter 3* focused on the machine learning aspects of the study. *Chapter 2* constructs an impression of the South African table grape industry and the key stakeholders involved at present. Further understanding was gained regarding the preharvest and postharvest quality factors that affect table grapes as well as the postharvest pathological and physiological disorders that can impact the product. Insights were also gained into the role that the cold chain plays as well as how cosmetic standards impact food waste. The chapter concluded by offering the current global best practises on how post-harvest decay can be reduced. *Chapter 3* described the general machine learning process with a specific focus on supervised learning for classification tasks.

No primary research was conducted in this study. Instead, data from five various secondary sources were incorporated to generate a robust dataset. The data preparation process is discussed in *Chapter 5*, where data was extracted, wrangled, and merged, resulting in a dataset of 467 observations. Introductory descriptive statistics were used to describe the data, thereafter inferential statistics were applied to determine the relationships between variables, later used for feature selection in the machine learning modelling process. Feature selection techniques made use of inferential statistics to identify variables that had a degree of association toward the response variable, the arrival quality scores for table grapes. The features selected were preharvest climatic variables measured over a period two weeks prior to harvest, the variety and ˚Brix at harvest, the incidence of decay found in the packhouse, and packaging decisions.

The machine learning process is described in *Chapter 6,* which saw four types of models being trained. The Random Forest classifier was the best performing model architecture but displayed varying prediction accuracy for the different classes. The model performed best when predicting red class-labels but struggled on green and performed worst for amber class-labels. The inherent noise in the amber class led to a reduction in class-label prediction performance. The analysis performed, is interpreted in *Chapter 8* and discussed how the features individually and clustered, affected the table grape arrival scores. The relationships between the variables found in the analysis were linked to the relevant literature and the implications thereof were identified. Potential model improvements that would increase the prediction accuracy were discussed with specific focus on industry's requirements.

The outcomes of the study are aimed to benefit both industry and academia. For industry, the study illustrated that machine learning techniques can be applied to perishable products to improve allocation decisions. The study also identified what constitutes a poor arrival and what the main causes are. Since the study was exploratory in nature, the outcome for academia is that a new avenue of research has been established where areas of further investigation have been suggested, opening a new knowledge base.

9.1 Did the study answer the research questions?

The overarching research question that the study intended to answer was:

What are the causes of poor table grape quality on arrival, and can these variables be modelled to predict the arrival quality prior to export?

From the correlation matrix and Cramér's *V* correlations, the variables most associated to arrival quality were identified and were selected for modelling. These variables can be found in *Table 6.1*, and consisted of the type of variety, the packaging, climactic variables measured two weeks prior to harvest, the ˚Brix at harvest and packhouse QC data consisting of the number of decayed berries found as well as the intake QC score.

Machine learning was applied to the variables with four types of models being trained and tested, with the best performing model being the Random Forest classifier, which was retained for further evaluations. The feature importance scores from the Random Forest classifier indicate which variables were weighted highest by the model when making predictions and can be found in *Figure 7.17*. The three most important variables for the model, accounting for over fifty percent of the models' predictions, consisted of the average ˚Brix at harvest (V20) 18.3%, the average temperature two weeks prior to harvest (V25) 17.3%, and the average Relative Humidity two weeks prior to harvest (V23) 16.1%. The variety variable also had a large effect on the model.

The sub-research questions were answered as follows:

9.1.1 Question one: what are the preharvest factors that influence table grape quality?

The preharvest variables that were most associated with the response were the total sum of rainfall two weeks prior to harvest (V21), the average relative humidity two weeks prior to harvest (V23), and the average daily temperature two weeks prior to harvest (V25).

A Cramér's *V* test of variable association was conducted on the variety variable (V6) and found a 0.435 level of association. This indicates that the variety harvested has a relatively strong level of association to the arrival quality.

9.1.2 Question two: What are point of harvest factors that influence table grape quality?

The only point of harvest variable collected was the average ˚Brix at harvest (V20), which is a measure of Total Soluble Solids (TSS), used to determine the fruits' ripeness. As stated in the literature, table grapes are a non-climacteric fruit meaning the product must be harvested at the optimal ripeness as it does not ripen further postharvest. The variable V20 had a negative correlation of 0.21 towards the arrival quality scores, however, as noted in the data analysis, the relationship is most likely quadratic, implying that ˚Brix readings that are either too low or high will result in a reduction of arrival quality.

9.1.3 Question three: What are the postharvest factors that influence table grape quality?

The most relevant postharvest variables were related to packaging decisions and the results from the packhouse QC reports. Packaging variables consisted of NOM_pack (V9) and NOM_inv_code (V10), which were intercorrelated with the effect relating to the type of external packaging (V9) either being punnet or loose pack.

The most important packhouse QC variables were the IN_Decayed berries* (V32) and IN_QC Score (V45). With correlation scores of 0.24 and 0.2 respectively.

9.1.4 Question four: What are the characteristics of a product that arrives in poor quality?

Content analysis was used in *section 7.1.2*, to determine the characteristics of table grapes that arrive in poor quality. The results can be seen in *Appendix I*, which overwhelmingly found postharvest decay to be the cause of amber and red arrival reports. Other lesser issues were the occurrence of berry shrivel caused by dehydration, berry drop, cracking, and stem browning.

9.1.5 Question five: Does the supply chain (from the point of loading onwards) contribute to poor arrival quality of table grapes?

Variables that measure the cold chain performance consisted of input variables V46 to V48 (denoted with the prefix 'VOY'). As stated in *Section 5.3.3,* data consisted of the frequency, or count of occurrence, of temperature spikes and breaks recorded for each container (V46 &

V47), as well as the total fruit age from time of harvest to arrival quality surveyance (V48), as seen in *Table 7.4*. The correlation matrix indicates no correlation between the supply chain variables and the arrival quality scores. Therefore, the supply chain had little impact on the arrival quality of table grapes exported to Europe for the 2020 season.

It must also be noted that if there are severe supply chain failures, exporters can claim the loss on income from their insurers, as noted in *section 7.1.1*. The implications are that severe supply chain breaks do not directly impact the shipment's financial return, but there are however, negative externalities in the form of food waste.

9.1.6 Question six: Can machine learning be employed to predict the downstream quality of table grapes?

Four machine learning models were trained to predict the arrival quality scores for table grapes, which consisted of Multinominal Logistic Regression, k-Nearest Neighbours, Decision Tree, and Random Forest Classifier. The cross-validation scores, which estimate how the model is expected to perform at predicting unseen data, is summarised in *Table 7.5*. The best performing model of the four tested was the Random Forest Classifier.

9.1.7 Question seven: Does the model perform equally well at predicting the three arrival score class-labels?

The Random Forest Classifier performance is summarised in *Table 7.6*, and performed relatively well with an overall accuracy of 63%. The model performed best when predicting red class-labels with a recall score of 75%, which indicates the ratio of correct red class predictions to the actual number of red class observations. The performance was hindered by amber and green class predictions, where the model seemed to misclassify the one class for the other. Techniques to improve the RF classifier's precision were discussed in *section 8.5*.

9.2 Recommendations

This section presents recommended mitigation practices to reduce the incidence of poor arrival quality of table grapes, which would reduce financial losses and food waste.

9.2.1 Site and variety selection decisions

When establishing a new vineyard, the site selected and variety planted has a long-term impact on the producers cashflow as it takes roughly three to four years before a vine is in full production. It is, therefore, of vital importance that the correct decisions are made prior to any money being spent.

Evaluating the *terroir*, a French term used to describe the environmental factors that affect a crop's phenotype, including unique environment contexts, farming practices and a crop's specific growth habitat, of any potential area is crucial. The environmental conditions such as rain, temperature, and relative humidity are of vital importance when evaluating different regional microclimates. Semi-desert conditions are ideal for growing table grapes, as long as there is sufficient water for irrigation and acceptable levels of wind.

Producers need to switch to varieties less susceptible to decay, the main cause of poor arrival report scores, to minimise postharvest losses. This is not as easy as it seems since retailers have preferred varieties and will either ask a price premium for those, or red list everything else. Producers, therefore, have somewhat limited choice in the matter. The best option is to trial new varieties on the specific farm's *terroir* and critically evaluate each prior to full-scale planting. It must be noted that there are other constraints such as labour hours, yield per hectare and sales price that also impact table grape variety selection (Helberg, 2021: 87).

9.2.2 ˚Brix at harvest

From the feature importance (*Figure 7.17*) scores for the Random Forest classifier, the most important variable was the average ˚Brix at harvest (18.3% importance score). As stated in the literature review, table grapes are a non-climacteric fruit and need to be harvested at the optimal TSS or ˚Brix as the fruit does not ripen further postharvest. It is, therefore, of vital importance that producers harvest the fruit around 18˚ to 22˚Brix to realise the best arrival quality scores.

9.2.3 Reducing pathogen infection and postharvest decay

From the content analysis results (*§7.1.2*) postharvest fungal decay can be attributed to poor arrival quality scores, additional costs, and food waste. It is, therefore, of vital importance to mitigate infection avenues prior to packaging.

The best practises to reduce infection are discussed in the literature review (*§2.12*), and consist of retractable canopy covers which can reduce the effects of the weather, fungicide application practises, sanitation practises from the vineyard to the point of packaging, and the use of fumigation chambers to kill any potential inoculum that could germinate postharvest.

The levels of fungal infection within the vineyard should be determined periodically at the main infection stages in the physiological growth cycle to effectively apply the mitigation strategies mentioned.

9.2.4 Producer packaging specification perceptions

As noted in *section 8.1*, the arrival quality scores varied depending on the type of packaging, possibly due to packhouse personnel's perception of the pack and product specification standards required for each packaging type. A potential solution would be to create a metric that measures the standard of the packhouse throughput. This could be achieved by sampling fruit prior to sorting and packing when the fruit arrives at the packhouse in field trays (lugs). This would develop a baseline for the general condition of the grapes prior to packing, which could then be compared to the final product at the end of the packing process. This metric would be beneficial for two reasons. Firstly, it would become a key performance indicator (KPI) for packhouse managers, and secondly, it would become an additional feature, which could aid in model performance if the machine learning model were to be deployed in industry.

9.2.5 Deployment of prescriptive analytics for industry

It is recommended that exporters embrace the benefits of machine learning algorithms to create usable models which can provide insights into allocation decisions. Improved allocation decisions would optimise financial returns through reducing the proportion of fruit arriving in poor quality.

Once predictive models have been built, the final phase of business is prescriptive analytics, visible in *Figure 9.1*.

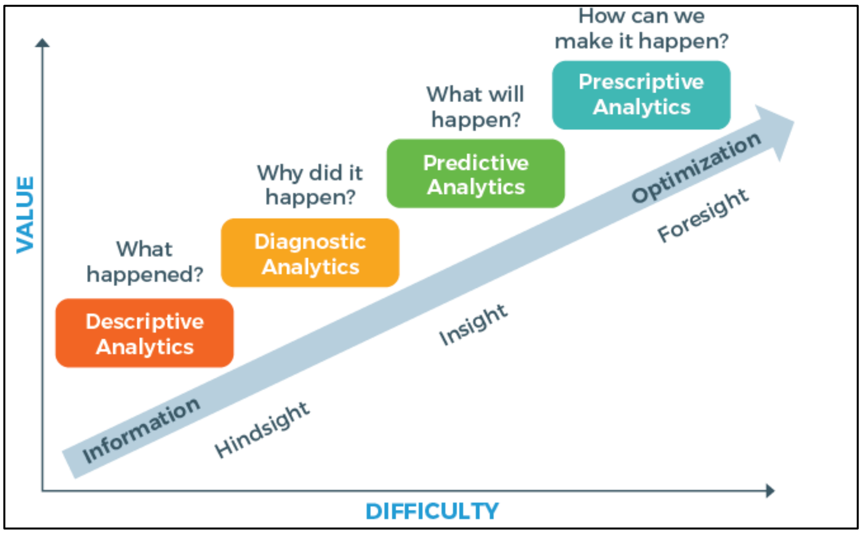


Figure 9.1: A framework for understanding the value of differing types of analytics

Source: Cromer, Goldstein & Koenemann, 2018

Prescriptive analytics builds on findings from both descriptive and predictive analytics and assesses the consequences of the potential decisions to determine the best course of action to mitigate a potential future risk (Greasley, 2019: 5). This type of analytics attempts to quantify the effects of future decisions to advise on the possible outcomes before the decisions are made. It, therefore, goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the implications of each decision option.

The following section discusses the prerequisites required to create a predictive model to be deployed in a real-world setting.

9.2.5.1 Prerequisites for a real-world predictive model

As stated in *section 4.2*, the data for this study was collected from secondary sources, meaning that the information was initially used for other primary means. Due to this, the data pre-processing applied was time-consuming owing to the file formats the data was received in.

To scale the methodology into a real-world table grape quality risk management tool, the data collection and storage methods would need to be improved. A potential solution to this could be the use of crowd-sourced data collection. This is a participatory method of building a dataset through the help of a large group of people. This method allows researchers to cheaply outsource data collection tasks allowing for scalability in the data extraction process. Amazon SageMaker Ground Truth, offers such services allowing analysts to build accurate training datasets in a short timeframe (Ahmad & Moos, 2019).

The optimal process would require a redesign of the packhouse QC system based on the optimal features identified using both the filter and wrapper method, discussed in *section 3.5.* Types of data gathered on those features should ideally consist of quantitative data, as it is more reliable and is preferred for analysis. Collected data should then be stored on a relational database, formatted in columns and rows, allowing for ease of access for analysis.

There is currently a plethora of new agricultural internet of things (IoT) enabled hardware solutions designed to gather infield data throughout the growing season. A couple of examples are irrigation management tools, fruit diameter sensors, dendrometer for vine trunk radius, leaf temperature sensors, soil sensors, and weather stations (Advanced Ai-Powered Precision Farming Platform, 2021). Gathering quality infield data throughout the growing season for each vineyard would greatly improve the prediction precision of the machine learning model due to the constant precise record keeping of site-specific environmental fluctuations and management practices.

The current practice of how cold chain data is captured by temperature monitors placed in the shipping containers would also need to be improved if scalability is intended. Data captured by monitors currently used in industry are not always uploaded onto the service providers database. This is due to the monitor hardware constraints, which require manual downloading and then subsequent uploading by receivers at the end of the supply chain. This tedious procedure often results in missing data. There is a growing demand for modern monitors, equipped with cellular transmitting functionality to provide real-time supply chain visibility. A biproduct of the enhanced functionality is improved data integrity since all data is retrieved, captured, and stored in a centralised database.

What would need improving is the method in which the model retrieves the data. Currently, to access the cellularly transmitted data, a user must manually download each individual shipment from a web-based platform. This would pose a challenge to the deployment of a scaled model. A possible solution would be an Electronic Data Interchange which automates the process of electronically sharing data between two or more systems.

9.3 Future research

Future research includes aspects of additional research that were not covered in this study. Topics considered for future research consist of the effect of the cold chain on shelf life, the effect of netting of table grape quality, selective harvesting depending on market specifications, omitted variables, extending the length of the climate variables, and testing more machine learning models to improve prediction accuracy.

9.3.1 The effect of cold chain on fruit quality

Future studies relating to the cold chain are discussed in the following sections and relate to various aspects in which the cold chain could possibly impact fruit quality but was not addressed in this research.

9.3.1.1 Early stages of the cold chain

Future research should be conducted into the early stages of the supply chain – from the point of harvest to the point of cooling of the packaged product. Haasbroek (2013: 122), made recommendations that fruit should be harvested early in the morning, around first light, when the pulp temperature is low, to mitigate moisture loss through transpiration and respiration. It has also been noted that the photosynthesis process during the day removes moisture from the grape berries and is replaced at night, therefore, fruit harvested midday has a moisture deficit (Fresh Produce Exporters' Forum, 2016).

It must be noted however, that fruit harvested in the morning will be stored in the packhouse at ambient temperatures throughout the day, up to 12 hours from the time of harvest. The trade-off between building a buffer stock of lower pulp temperature fruit compared to picking throughout the day and the resultant effect on table grape quality should be investigated further.

9.3.1.2 Measure of supply chain performance

It is also possible that the measure of supply chain effectiveness, the sum of temperature spikes and breaks (§*5.3.3*), does not truly represent how well the supply chain is performing. Other variables effect on the supply chain that should be studied are the mean and standard deviation of ambient temperature and the mean kinetic temperature (a heat transfer variable that expresses the overall effect of temperature fluctuations during storage or transit of perishable goods).

9.3.1.3 Pulp and air temperature monitoring

As stated in the literature review, due to the recommended table grape storage conditions of -0.5°C at a relative humidity (RH) of 95% (Henning & Chetty, 2019), a small temperature fluctuation from these conditions can cause the dew point to be reached, resulting in condensation (Linke & Geyer, 2013). To better monitor this, exporters could make use of a combination of temperature probes to determine the dew point throughout the voyage. Monitors that measure the air temperature, RH, and the fruit pulp temperature would need to be used in combination to triangulate the dew point and resultant condensation (Ainsworth, 2018).

9.3.1.4 Effect of supply chain shocks

The 2021 and 2022 seasons that followed this study, saw supply chain shocks in the form of delays and temperature breaks. Future research could determine how a faulty supply chain effects the arrival quality of table grapes and whether the supply chain variables should be added to the model features for arrival quality predictions.

9.3.1.5 The effect of the supply chain on table grape shelf life

It must be noted that the supply chain results for this study contradict previous work on how the export cold chain impacts fresh fruit, see (Conradie, 2019; Fedeli, 2019; Freiboth *et al.*, 2013; Goedhals-Gerber *et al.*, 2015; Joubert, 2019; Khumalo, 2018; Kisten, 2020; van der Klein, 2018; Valentine & Goedhals-gerber, 2017). Reasons for lack of a correlation between the supply chain variables and fruit quality scores might be a result of the point in the cold chain that the fruit quality is measured. The arrival quality control is conducted roughly 30 days postharvest, prior to the fruit entering the retailer's distribution centres, the retail stores, and the final consumer's fridge. Therefore, it is possible that the negative effects of temperature breaks during the shipping leg might only manifest further downstream, reducing the total shelf life of the product.

Future research would need to either study supply chains with longer transit times, for example table grapes shipped from South Africa to the Far east (e.g., China and Taiwan), where total fruit age on arrival can be closer to 45 days on average. Or, additional fruit quality inspections would need to be conducted further downstream. However, this might be impractical.

If it is the case that the underperforming cold chain results in a loss of shelf life, the negative connotations would be two-fold. From a sustainability perspective the total food waste for the table grape supply chain may be worse than the data for this study suggests due to downstream waste. The second negative effect would be for Company X and the South African table grape industry as a whole. If consumers develop a negative perception of South African table grapes, consumers might substitute South African grapes for grapes produced in other parts of the globe, or possibly substituting grapes for another fruit or product entirely. It is, therefore, important for in-depth end-of-chain studies to be conducted in collaboration with upstream role players.

9.3.2 The effect of netting on table grape quality

Some orchards observed in the dataset make use of permanent protective nets to cover the vineyard. These flat-roof shade-nets are used to protect the grapes against strong winds, which can negatively impact the grapes throughout the growing process. A drawback to the shade-nets is that they create a greenhouse-like environment, which is warm and humid. If unseasonal rainfall occurs, wind cannot effectively enter a vineyard covered in shade-nets, limiting the evaporation of freestanding moisture, further increasing the relative humidity within the covered environment. As stated in *section 2.10.3*, these are the ideal conditions for pathogen growth. Unfortunately, the dataset provided did not distinguish between orchards that had netting and those that did not.

A future study could conduct a cost-benefit analysis to compare farming table grapes under nets compared to alternative methods.

9.3.3 Matching product specifications to harvesting practices

All producers have some good quality, mediocre, and poor-quality grapes. The challenge is to allocate the correct product to the correct market. A strategy that has been suggested by a producer (Rossouw, 2022), is for exporters to better indicate on the packing programs the product specification requirements for each market so that an appropriate selective harvesting strategy can be applied rather than a strip picking approach.

Selective harvesting is when the picking teams only harvest grapes which adhere to a specific set of standards enforced by the producer. In contrast, strip picking is when grapes are harvested sequentially, vine by vine, row by row, without any priority placed on specific characteristics of the product.

As stated, there is a range of quality in every vineyard, therefore, future research could try to identify the best strategy when giving instruction to the picking teams, to best match market requirements to the product, maximising returns and reducing food waste downstream in the supply chain.

9.3.4 Omitted variables

A variable not tested in this research was the effect of wind prior to harvest, due to a lack of sample data. Wind aids in removing humidity and drying out the vineyard canopy post rainfall, possibly reducing the likelihood of physiological disorders such as berry splitting and cracking. Evapotranspiration is another measure, which could be tested in future research.

9.3.5 How long the effect of rain and suboptimal climatic conditions effect unharvested table grapes

As stated in *section 5.2*, climate variables for this study measured the effects of the climate for a two-week period. The length of time chosen was linked to the literature from previous studies. In general, the variables seemed to be relevant when predicting the arrival quality of table grapes, based on the results from the feature importance matrix, *Figure 7.17*, however, when referring to *Figure 7.9*, the rainfall measure did not align appropriately to the red arrival scores. It is, therefore, possible that the two-week period suggested is not a long enough period of time to adequately capture the effect of the climate. Future research should, therefore, try and determine the optimal length of time for each climactic variable to better represent the effect on the arrival quality score.

9.3.6 Investigating the use of machine learning further

There are more machine learning models available that could be applied to the dataset, which might yield more accurate results. Future research could apply Kernel methods such as support-vector machine (SVM), Artificial neural networks, or graphical models.

List of References

Advanced Ai-Powered Precision Farming Platform. 2021. [Online], Available: https://supplant.me/technology/ [2022, August 23].

Agricultural Research Council (ARC). 2019. *Annual Report*. [Online], Available: https://www.arc.agric.za/Pages/ARC-Annual-Report-.aspx [2021, October 17].

Agricultural Standards Unit - United Nations Economic Commission for Europe. 2017. *STANDARD FFV-19 Table Grapes*. New York and Geneva: United Nations Economic Commission for Europe (UNECE). [Online], Available: http://www.unece.org/.

Ahmad, S. & Moos, L. 2019. *Use the wisdom of crowds with Amazon SageMaker Ground Truth to annotate data more accurately*. [Online], Available: https://aws.amazon.com/blogs/machine-learning/use-the-wisdom-of-crowds-with-amazon-sagemaker-ground-truth-to-annotate-data-more-accurately/ [2022, October 03].

Ainsworth, N. 2018. *Should I monitor air or product (pulp) temperature?* [Online], Available: https://afccc.org.au/images/resources/hort projects/cold-supply-air-pulp.pdf.

Alghamdi, A.H. & Li, L. 2013. Adapting design-based research as a research methodology in educational settings. *International Journal of Education and Research*. 1(10):1–12.

Alle scholen, cafés en restaurants tot en met 6 april dicht om coronavirus. 2020. [Online], Available: https://nos.nl/artikel/2327194-alle-scholen-cafes-en-restaurants-tot-en-met-6-april-dicht-om-coronavirus [2022, September 13].

Bagaria, J. 2019. *Basic Set Theory*. [Online], Available: https://plato.stanford.edu/entries/set-theory/basic-set-theory.html [2021, October 07].

Barkai-Golan, R. 2001. *Postharvest Diseases of Fruit and Vegetables: Development and Control*. Amsterdam: Elsevier.

Becker, B.R. & Fricke, B.A. 2014. Transpiration and Respiration of Fruits and Vegetables. *Universidad of Missouri*. 18. [Online], Available: http://b.web.umkc.edu/beckerb/publications/chapters/trans_resp.pdf.

Becker, B., Misra, A. & Fricke, B. 1996. Bulk Refrigeration of Fruits and Vegetables Part I: Theoretical Considerations of Heat and Mass Transfer. *HVAC&R Research*. 2(2):122–134.

Beede, R.H. n.d. *Berry cracking in table grapes*. Kings County, California USA. [Online], Available: https://cekings.ucanr.edu/files/18991.pdf.

Bentahar, O. & Cameron, R. 2015. Design and implementation of a mixed method research study in project management. *Electronic Journal Of Business Research Methods*. 13(1):3–15. [Online], Available: http://web.a.ebscohost.com.ezp.waldenulibrary.org/ehost/search/advanced?sid=2d567c58-d245-4b74-bf4c-285a67f1d8fa@sessionmgr4003&vid=19&hid=4214.

Beraha, M., Metelli, A.M., Papini, M., Tirinzoni, A. & Restelli, M. 2019. Feature Selection via

Mutual Information: New Theoretical Insights. in *2019 International Joint Conference on Neural Networks (IJCNN)* Vols 2019-July. IEEE. 1–9.

Berk, R.A. & Bleich, J. 2013. Statistical Procedures for Forecasting Criminal Behavior. *Criminology & Public Policy*. 12(3):513–544.

Børve, J., Skaar, E., Sekse, L., Meland, M. & Vangdal, E. 2003. Rain Protective Covering of Sweet Cherry Trees—Effects of Different Covering Methods on Fruit Quality and Microclimate. *HortTechnology*. 13(1):143–148.

Bottini, N., Ernst, C. & Luelbker, M. 2007. *Offshoring and the labour market: What are the issues?*

Bowcher, A., AnDi Communications, Clarke, S., Rogiers, S., Hardie, J. & NWGIC. 2012. *Fruit splitting of wine grape berries*. Wagga Wagga NSW, Australia. [Online], Available: https://cdn.csu.edu.au/__data/assets/pdf_file/0010/455194/NWGIC-fs1-fruit-splitting.pdf.

Brink, J.C., Holz, G. & Fourie, P.H. 2006. Effect of Fungicide Spray Cover on Botrytis Cinerea Infection in Grape Bunches. *South African Journal of Enology & Viticulture*. 27(1):51–56.

Brink, J.C., Calitz, F.J. & Fourie, P.H. 2016. Spray deposition and control of Botrytis cinerea on grape leaves and bunches: Part 2 (wine grapes). *South African Journal of Enology and Viticulture*. 37(2):157–168.

Brownlee, J. 2019. *How to Choose a Feature Selection Method For Machine Learning*. [Online], Available: https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/ [2022, August 11].

Brownlee, J. 2021. *A Gentle Introduction to Threshold-Moving for Imbalanced Classification*. [Online], Available: https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/ [2022, September 11].

Bryman, A. & Bell, E. 2015. *Business Research Methods. Methods*.

Bunch Rot Part 1: Botrytis cinerea. 2021. [Online], Available: https://www.lodigrowers.com/botrytis-cinerea/ [2021, May 19].

Bunyard, M. 2020. Improved Crop Protection with SOLARIG® Covers. *IsraelAgri*. (August). [Online], Available: https://www.israelagri.com/?CategoryID=402&ArticleID=1880.

Burger, D.A. 2000. Postharvest berry split and abscission in “Thompson Seedless” and “Waltham Cross” table grapes. Stellenbosch University, Stellenbosch. [Online], Available: http://scholar.sun.ac.za/handle/10019.1/51877.

Burkov, A. 2019. *The Hundred-Page Machine Learning Book*. Andriy Burkov. [Online], Available: https://books.google.co.za/books?id=0jbxwQEACAAJ.

Cappellini, R.A., Ceponis, M.J. & Lightner, G.W. 1986. Disorders in table grape shipments to the New York market: 1972 - 1984. *Plant Dis.* 70:1075–1079.

Chabada, L., Damgaard, C.M., Dreyer, H.C., Hvolby, H.H. & Dukovska-Popovska, I. 2014. Logistical Causes of Food Waste: A Case Study of a Norwegian Distribution Chain of

Chilled Food Products. in *IFIP Advances in Information and Communication Technology* Vol. 438. 273–280.

Choueiry, G. 2022. *Identify Variable Types in Statistics*. [Online], Available: https://quantifyinghealth.com/variable-types/ [2022, August 26].

Ciliberti, N., Fermaud, M., Roudet, J. & Rossi, V. 2015. Environmental conditions affect Botrytis cinerea infection of mature grape berries more than the strain or transposon genotype. *Phytopathology*. 105(8):1090–1096.

Civilian Secretariat for Police Service. 2016. *2016 White Paper on Safety and Security - Annexure B*. Pretoria, South Africa. [Online], Available: https://www.saferspaces.org.za/resources/entry/2016-white-paper-on-safety-and-security.

Conradie, C. 2019. Identifying temperature breaks in the export cold chain of navel oranges: A Western Cape case. Stellenbosch University, Stellenbosch. [Online], Available: http://hdl.handle.net/10019.1/107280.

Convection Heat Transfer – Natural and Forced Convection. 2019. [Online], Available: https://www.smlease.com/entries/thermal-design/convection-convective-heat-transfer/ [2020, May 04].

Coombe, B.G. 1995. Growth Stages of the Grapevine: Adoption of a system for identifying grapevine growth stages. *Australian Journal of Grape and Wine Research*. 1(2):104–110.

Council of Supply Chain Management Professionals (CSCMP). 2017. *Supply chain definitions*.

Crisosto, C.H., Mitcham, E.J. & Kadar, A.A. 1998. *Table Grape Fact Sheet*. [Online], Available: http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Fruit_English/?uid=24&ds=798 [2020, May 13].

Cromer, B., Goldstein, N. & Koenemann, K. 2018. *How Manufacturers Are Using Prescriptive Analytics to Optimize Profits*. [Online], Available: https://www.tbmcg.com/resources/blog/technology-prescriptive-analytics/ [2022, September 28].

DAFF. 2017. *A profile of the South African table grape market value chain*.

Data Flair. 2021. *Joins in Tableau – Learn the rules to join tables in Tableau*. [Online], Available: https://data-flair.training/blogs/tableau-joins/ [2021, October 08].

Defilippi, B.G., Rivera, S.A., Preez-Donoso, A., Gonzalez-Aguero, M. & Campos-Vargas, R. 2019. Table Grapes. in *Postharvest Physiological Disorders in Fruits and Vegetables* S. Tonetto de Freitas & S. Pareek (eds.). Boca Ratan : Taylor & Francis, 2018.: CRC Press S. Tonetto de Freitas & S. Pareek (eds.). 237–260. [Online], Available: https://www.taylorfrancis.com/books/9781351973175.

Deforge, B.R. 2010. Research Design Principles. in *Encyclopedia of Research Design* N. J. Salk ed. Thousand Oaks: SAGE Publications, Inc. 1253–1260.

Department of Agriculture Forestry and Fisheries (DAFF). 2015. *Standards and Requirements Regarding Control of the Export of table grapes*. Pretoria, South Africa. [Online], Available: http://www.nda.agric.za/doaDev/sideMenu/foodSafety/doc/grapes/TableGrapesreg(2012).doc.

Dickson, A., Adu-Agyem, J. & Emad Kamil, H. 2018. Theoretical and conceptual framework: mandatory ingredients of a quality research. *International Journal of Scientific Research*. 7(1):438–441.

Dix, P. 2021. *Why Time Series Matters for Metrics , Real-Time Analytics and Sensor Data*. [Online], Available: https://www.influxdata.com/what-is-time-series-data/.

Elad, Y., Williamson, B., Tudzynski, P. & Delen, N. 2007. *Botrytis: Biology, pathology and control*.

European Union. 2020. *Brief on food waste in the European Union*. [Online], Available: https://www.eu-fusions.org/.

EXSA webpage. 2021. *Production Regions: South Africa and Namibia*. [Online], Available: https://exsa.com/our-production-regions/ [2021, October 18].

FAO. 2016. *Table and Dried Grapes: FAO -OIV FOCUS 2016*.

Fedeli, S. 2019. Identifying temperature breaks in pome fruit and table grape export cold chains from South Africa to the United Kingdom and the Netherlands : A Western Cape case Savia Fedeli. Stellenbosch University, Stellenbosch. [Online], Available: https://scholar.sun.ac.za/handle/10019.1/107234.

Fidelibus, M.W., Vasquez, S.J. & Kaan Kurtural, S. 2016. Late-season plastic canopy covers affect canopy microclimate and fruit quality of ‘autumn king’ and ‘redglobe’ table grapes. *HortTechnology*. 26(2):141–147.

Fischer, D., Craig, W.L., Watada, A.E., Douglas, W. & Ashby, B.H. 1992. Simulated in-transit vibration damage to packaged fresh market grapes and strawberries. *Applied Engineering in Agriculture*. 8(3):363–366.

Fortmann-Roe, S. 2012. *Understanding the Bias-Variance Tradeoff*. [Online], Available: http://scott.fortmann-roe.com/docs/BiasVariance.html [2022, September 03].

Fourie, J.F. 2008. Harvesting, Handling and Storage of Table Grapes (With Focus on Pre- And Post-Harvest Pathological Aspects). *Acta Horticulturae*. 785(785):421–424.

FPEF. 2018. *Fresh fruit directory 2018*. Cape Town, South Africa. [Online], Available: https://www.fpef.co.za/wp-content/uploads/2018/04/FPEF-ED-2018-rev3.pdf.

Franck, J., Latorre, B.A., Torres, R. & Zoffoli, J.P. 2005. The effect of preharvest fungicide and postharvest sulfur dioxide use on postharvest decay of table grapes caused by Penicillium expansum. *Postharvest Biology and Technology*. 37(1):20–30.

Frees, E.W. 2004. *Longitudinal and Panel Data*. Cambridge University press.

Freiboth, H.W., Goedhals-Gerber, L., Van Dyk, F.E. & Dodd, M.C. 2013. Investigating temperature breaks in the summer fruit export cold chain: A case study. *Journal of Transport and Supply Chain Management*. 7(1):1–7.

Fresh Food Trade SA. 2021. Pretoria, South Africa. [Online], Available: https://www.dalrrd.gov.za/doaDev/sideMenu/internationalTrade/docs/tradeFacilitation/Fresh-Food-Trade-SA-2021-eBook.pdf.

Fresh Produce Exporters' Forum. 2016. *Harvest to Home: The fresh Fruit Trade Chain*. Cape Town, South Africa.

García, I. 2022. *Botrytis Cinerea: a highly infectious crop killer*. [Online], Available: https://www.cannagardening.com/botrytis_cinerea_in_detail#:~:text=The mold grows on dying,already have penetrated the plant. [2022, May 10].

Goedhals-Gerber, L.L., Haasbroek, L., Freiboth, H. & Van Dyk, F.E. 2015. An analysis of the influence of logistics activities on the export cold chain of temperature sensitive fruit through the Port of Cape Town. *Journal of Transport and Supply Chain Management*. 9(1):1–9.

Goedhals-Gerber, L.L., Stander, C. & Van Dyk, F.E. 2017. Maintaining cold chain integrity: Temperature breaks within fruit reefer containers in the Cape Town Container Terminal. *Southern African Business Review*. 21(1):362–384.

Gogh, B. Van, Aramyan, L., Sluis, A. Van Der, Soethoudt, H. & Scheer, F. 2013. *Feasibility of a network of excellence postharvest food losses*. Wageningen, Netherlands.

Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*. 24(1):44–65.

González-Domínguez, E., Caffi, T., Ciliberti, N. & Rossi, V. 2015. A mechanistic model of botrytis cinerea on grapevines that includes weather, vine growth stage, and the main infection pathways. *PLoS ONE*. 10(10):1–23.

Gopala Rao, C. 2015. WITHDRAWN: Postharvest Physiology of Fruits and Vegetables. in *Engineering for Storage of Fruits and Vegetables* Elsevier Inc. 13–38.

Government Gazette. 2020. *Alert level 4*.

Grant, C. & Osanloo, A. 2014. Understanding, Selecting, and Integrating a Theoretical Framework in Dissertation Research: Creating The Blueprint for Your "House". *Administrative Issues Journal Education Practice and Research*. 12–26.

Greasley, A. 2019. *Simulating Business Processes for Descriptive, Predictive, and Prescriptive Analytics*. Berlin, Germany: De Gruyter.

Gustavsson, J., Cederberg, C. & Sonesson, U. 2011. Global Food Losses and Food Waste: Extent, Causes and Prevention. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*.

Haasbroek, L.M. 2013. An analysis of temperature breaks in the summer fruit export cold chain from pack house to vessel. Stellenbosch University, Stellenbosch.

Harvey, J.M. 1955. A method of forecasting decay in California storage grapes. *Phytopathology*. 45:229–32.

Hatch, W. 1989. Mass Transfer Characteristics of Fresh Fruits Stored in Regular and Controlled Atmosphere Conditions. Oregon State University. [Online], Available: https://ir.library.oregonstate.edu/downloads/jh343w66f.

Helberg, C.J. 2021. A systematic approach to select new table grape varieties for cultivation by. Stellenbosch University.

Henning, B. & Chetty, V. 2019. [Online], Available: https://ppecb.com/wp-content/uploads/2019/04/Intransit-Handling-Protocol-HP34G-China-Table-Grapes.pdf.

Herholdt, R. 2021.

Hill, G.N., Henshall, W.R. & Beresford, R.M. 2017. Manipulating rainfall to study symptom expression of *Botrytis cinerea* infection in wine grapes. *New Zealand Plant Protection*. 70:301–309.

Hill, G.N., Beresford, R.M. & Evans, K.J. 2019. Automated analysis of aggregated datasets to identify climatic predictors of Botrytis bunch rot in wine grapes. *Phytopathology*. 109(1):84–95.

Holcroft, D. 2015. *Water Relations in Harvested Fresh Produce Water Relations in Harvested Fresh Produce*. (15–01). La Pine, Oregon, USA. [Online], Available: https://apps.who.int/iris/handle/10665/41227.

de Hooge, I.E., van Dulm, E. & van Trijp, H.C.M. 2018. Cosmetic specifications in the food waste issue: Supply chain considerations and practices concerning suboptimal food products. *Journal of Cleaner Production*. 183(April):698–709.

Inferential Statistics. 2022. [Online], Available: https://www.cuemath.com/data/inferential-statistics/ [2022, August 22].

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. *An introduction to statistical learning : with applications in R*. New York: New York : Springer. [Online], Available: https://search.library.wisc.edu/catalog/9910207152902121.

Jansen, C. 2021. Grape postharvest disease risk evaluation service grows by leaps and bounds. *Fresh Plaza*. 17 December. [Online], Available: https://www.freshplaza.com/article/9384278/grape-postharvest-disease-risk-evaluation-service-grows-by-leaps-and-bounds/.

Johnson, R.B., Onwuegbuzie, A.J. & Turner, L.A. 2007. Toward a Definition of Mixed Methods research. *Journal of Mixed Methods Research*. 1(2):112–133.

Join Your Data. 2021. [Online], Available: https://help.tableau.com/current/pro/desktop/en-us/joining_tables.htm [2021, October 07].

Jooste, A. 2020. Item descriptions, E-mail to C.Rossouw. 11 Septemb:1–9. [Online], Available:

e-mail: 18363180@sun.ac.za.

Joubert, L. 2019. An investigation to evaluate the impact of quality-controlled logistics on food waste and food quality with the assistance of Internet of Things. Stellenbosch University, Stellenbosch.

Kakalíková, Ľ., Jankura, E. & Šrobárov́, A. 2009. First report of Alternaria bunch rot of grapevines in Slovakia. *Australasian Plant Disease Notes*. 4(1):68–69.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N.H., Weaver, C., Lee, B., et al. 2011. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*. 10(4):271–288.

Kaur, S., Arora, N.K., Gill, K.B.S., Sharma, S. & Gill, M.I.S. 2019. Hexanal formulation reduces rachis browning and postharvest losses in table grapes cv. 'Flame Seedless'. *Scientia Horticulturae*. 248(January):265–273.

Khumalo, G. 2018. Identifying Temperature Breaks in the Export Cold Chain of Navel Oranges: A Western Cape Case. Stellenbosch University, Stellenbosch. [Online], Available: http://scholar.sun.ac.za/handle/10019.1/103667.

Killer, A. 2019. *VBA to copy text from multiple txt files into successive columns in EXCEL*. [Online], Available: https://answers.microsoft.com/en-us/msoffice/forum/all/vba-to-copy-text-from-multiple-txt-files-into/40d38e3e-42a3-4e11-8e9e-b066e4d6761e?page=2 [2019, September 14].

Kisten, C. 2020. Investigating quality issues after the introduction of humidifiers into table grape pack houses : A Northern Cape case. Stellenbosch University, Stellenbosch.

van der Klein, M. 2018. An analysis of temperature breaks along the table grape export cold chain from exporter's pack-house to the importer's coldstore and beyond. Netherlands Maritime University, Rotterdam.

Kuhn, M. & Johnson, K. 2013. *Applied Predictive Modeling*. New York, NY: Springer.

lal Basediya, A., Samuel, D.V.K. & Beera, V. 2013. Evaporative cooling system for storage of fruits and vegetables - a review. *Journal of Food Science and Technology*. 50(3):429–442.

Langenhoven, S. 2022. *Uvasys Vivo and Uvasys Blue - trial report*. Stellenbosch, Republic of South Africa.

Lavrakas, P. 2008. Panel data analysis. in *Encyclopedia of Survey Research Methods* Thousand Oaks, California: SAGE Publications, Inc.

Li, Z. & Thomas, C. 2014. Quantitative evaluation of mechanical damage to fresh fruits. *Trends in Food Science & Technology*. 35(2):138–150.

Li, L., Kaplunov, T., Zutahy, Y., Daus, A., Porat, R. & Lichter, A. 2015. The effects of 1-methylcyclopropane and ethylene on postharvest rachis browning in table grapes. *Postharvest Biology and Technology*. 107:16–22.

Liberty, J.T., Okonkwo, W.I. & Echiegu, E. a. 2013. Evaporative Cooling: A Postharvest Technology for Fruits and Vegetables Preservation. *International Journal of Scientific & Engineering Research*. 4(8):2257–2266.

Linke, M. & Geyer, M. 2013. Condensation dynamics in plastic film packaging of fruit and vegetables. *Journal of Food Engineering*. 116(1):144–154.

Lopez Camelo, A.F. 2004. *Manual for the Preparation and Sale of Fruits and Vegetables: From Field to Market*. (FAO Agricultural services bulletin 151). Rome, Italy: Food and Agriculture Organization of the United Nations. [Online], Available: https://books.google.co.za/books?id=DwUdO9hPZ7sC.

Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U. & Weber, E. 1995. Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L. ssp. vinifera)—Codes and descriptions according to the extended BBCH scale. *Australian Journal of Grape and Wine Research*. 1(2):100–103.

Lorenzini, M. & Zapparoli, G. 2014. Characterization and pathogenicity of Alternaria spp. strains associated with grape bunch rot during post-harvest withering. *International Journal of Food Microbiology*. 186:1–5.

Louw, L. 2017. Economic Aspects of Losses and Waste: Case Study of the South African Table Grape Supply Chain. University of Pretoria, Pretoria. [Online], Available: https://repository.up.ac.za/handle/2263/1682.

Mashabela, T.E. 2007. Measuring the Relative Competitiveness of Global Deciduous Fruit Supply Chains: South Africa versus Chile. University of Stellenbosch. [Online], Available: https://scholar.sun.ac.za/handle/10019.1/18221.

Menzel, K. 2009. *Introduction to Statistical Methods in Economics*. Cambridge, MA USA. [Online], Available: https://ocw.mit.edu/courses/economics/14-30-introduction-to-statistical-methods-in-economics-spring-2009/assignments/MIT14_30s09_sol_pset01.pdf.

Ministry of Electronics and Information Technology: Government of India. 2019. *Grapes: Diseases and Symptoms*. [Online], Available: http://vikaspedia.in/agriculture/crop-production/integrated-pest-managment/ipm-for-fruit-crops/ipm-strategies-for-grapes/grapes-diseases-and-symptoms#section-11 [2019, July 27].

Molitor, D., Baus, O., Hoffmann, L. & Beyer, M. 2016. Meteorological conditions determine the thermal-temporal position of the annual Botrytis bunch rot epidemic on Vitis vinifera L. cv. Riesling grapes. *Oeno One*. 50(4):231–244.

Montero, C.R.S., Schwarz, L.L., Santos, L.C. dos, Andreazza, C.S., Kechinski, C.P. & Bender, R.J. 2009. Postharvest mechanical damage affects fruit quality of “Montenegrina” and “Rainha” tangerines. *Pesquisa Agropecuária Brasileira*. 44(12):1636–1640.

Morgan, D.P. & Michailides, T.J. 2004. First Report of Melting Decay of ‘Red Globe’ Grapes in

California. *Plant Disease*. 88(9):1047–1047.

Murugan, K. 2022. Classification of Selected Cardiac Abnormalities through Machine Learning. Stellenbosch University, Stellenbosch. [Online], Available: https://scholar.sun.ac.za/handle/10019.1/124803.

Müssmann, C. 2015. Supply Chain Finance : Improving the efficiency of the table grape industry – A case study. Stellenbosch University, Stellenbosch. [Online], Available: http://hdl.handle.net/10019.1/97061.

Mycotoxins. 2018. [Online], Available: https://www.who.int/news-room/fact-sheets/detail/mycotoxins [2020, May 20].

Nel, I. 2020. [Online], Available: http://www.arc.agric.za/Pages/Home.aspx.

Nelson, K.. 1978. Pre-cooling - its significance to the market quality of table grapes. *International Journal of Refrigeration*. 1(4):207–215.

Ngcobo, M.E.K. 2013. Resistance to Airflow and Moisture Loss of Table Grapes Inside Multi-scale Packaging. Stellenbosch University, Stellenbosch. [Online], Available: http://hdl.handle.net/10019.1/80192.

Nunan, D., Bankhead, C. & Aronson, J. 2017. *Catalogue of Bias*. [Online], Available: http://www.catalogofbias.org/biases/selection-bias/ [2021, August 05].

OECD. 2020. Preliminary Report: Evaluation of the Impact of the Coronavirus (COVID-19) on Fruit and Vegetables Trade. 1–14. [Online], Available: https://www.oecd.org/agriculture/fruit-vegetables/oecd-covid-19-impact-on-fruit-and-vegetables-trade.pdf.

Papargyropoulou, E., Wright, N., Lozano, R., Steinberger, J., Padfield, R. & Ujang, Z. 2016. Conceptual framework for the study of food waste generation and prevention in the hospitality sector. *Waste Management*. 49:326–336.

Peacock, B. & Smilanick, J. 1996. Postharvest decay of late season table grape. (510):5.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., et al. 2011. *Cross-validation: evaluating estimator performance*. Vol. 12. [Online], Available: https://scikit-learn.org/stable/modules/cross_validation.html [2022, September 12].

Phaleng, L. & Tshitiza, O. 2019. *South African Fruit Trade Flow*. [Online], Available: https://www.namc.co.za/wp-content/uploads/2019/04/South-African-Fruit-flow-report-March-2019-Issue-33-Final.pdf.

Pieterse, J.L. 2022. A predictive model for precision tree measurements using applied machine learning. Stellenbosch University, Stellenbosch. [Online], Available: https://scholar.sun.ac.za/handle/10019.1/124512.

Du Plessis, S.J. 1935. Studies of the Wastage of Export Grapes: With Special Reference to that caused by Botrytis Cinerea, Pers. Stellenbosch University, Stellenbosch.

PPECB. 2003. *160 Years of Export: The History of the Perishable Products Export Control Board*. Paarl, South Africa.

PPECB. 2013. *PPECB Blue Book*. [Online], Available: https://ppecb.com/.

PPECB. 2018. [Online], Available: https://ppecb.com/wp-content/uploads/2015/03/HP22-M-files-Carrying-Temperature-Regime-Codes-of-Perishable-Produce-16032018.pdf.

Priefer, C., Jörissen, J. & Bräutigam, K.-R. 2016. Food waste prevention in Europe – A cause-driven approach to identify the most relevant leverage points for action. *Resources, Conservation and Recycling*. 109:155–165.

Primary vs. Secondary Sources. 2018. [Online], Available: https://ncu.libguides.com/researchprocess/primaryandsecondary [2018, December 13].

Romanazzi, G. & Feliziani, E. 2014. Botrytis cinerea (Gray Mold). in *Postharvest Decay: Control Strategies* S. Bautista-Baños (ed.). Aconda, Italy: Elsevier S. Bautista-Baños (ed.). 131–146.

Rossouw, F. 2022. Semi-structured interview: Harvest practices. (Worcester, South Africa).

Russell, S. & Norvig, P. 2021. *Artificial Intelligence A Modern Approach (4th Edition)*. [Online], Available: https://books.google.com.br/books?id=koFptAEACAAJ.

Saghir, M. 2004. the Concept of Packaging Logistics. *ResearchGate*. (May 2002):1–31.

SATI. 2018. *SATI Statistical Booklet 2018*. Paarl, South Africa. [Online], Available: https://www.satgi.co.za/industry-information/statistics-booklet/.

SATI. 2019. *SATI Statistical Booklet 2019*. Paarl, South Africa. [Online], Available: https://www.satgi.co.za/industry-information/statistics-booklet/.

SATI. 2020. *SATI Statistical Booklet 2020*. Paarl, South Africa. [Online], Available: https://www.satgi.co.za/industry-information/statistics-booklet/.

SATI. 2021. *SATI Statistical Booklet 2021*. Paarl, South Africa. [Online], Available: https://www.satgi.co.za/industry-information/statistics-booklet/.

SATI. 2022. *SATI Statistical Booklet 2022*. Paarl, South Africa. [Online], Available: https://www.satgi.co.za/industry-information/statistics-booklet/.

Saunders, M., Lewis, P. & Thornhill, A. 2019. *Research Methods for Business Students*. Pearson. [Online], Available: https://books.google.co.za/books?id=LtiQvwEACAAJ.

SAWS. 2022. *Organisational overview*. [Online], Available: https://www.weathersa.co.za/home/overview [2022, May 24].

Schouten, P., Lemckert, C., Parisi, A., Downs, N., Underhill, I. & Turner, G. 2011. Variable Wind Speed and Evaporation Rates : A Practical and Modelling Exercise for High School Physics and Multi-Strand Science Classes. *Teachingscience*. 57(2):47–51. [Online], Available: http://hdl.handle.net/10072/41089.

Sengupta, D. 2012. *What is Price Pooling?* [Online], Available: https://economictimes.indiatimes.com/news/economy/policy/what-is-price-pooling/articleshow/15509206.cms [2021, November 07].

Shankar, E.G., Sharma, O.P., Raj Boina, D. & Narsl Reddy, M. 2014. *Agro-Eco System Analysis: Intergrated Pest Mangement package*. Hyderbad.

De Simone, N., Pace, B., Grieco, F., Chimienti, M., Tyibilika, V., Santoro, V., Capozzi, V., Colelli, G., et al. 2020. Botrytis cinerea and Table Grapes: A Review of the Main Physical, Chemical, and Bio-Based Control Treatments in Post-Harvest. *Foods*. 9(9):1138.

Singh, P. & Heldman, D. 2014. *Introduction to Food Engineering*. Fifth Edit ed. S. Taylor (ed.). San Diego, CA, USA: Elsevier. [Online], Available: https://www.elsevier.com/books/introduction-to-food-engineering/singh/978-0-12-398530-9.

Srivastava, T. 2014. *Introduction to k-Nearest Neighbors: A powerful Machine Learning Algorithm*. [Online], Available: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/ [2022, August 05].

Statistical Report on World Vitiviniculture. 2019. [Online], Available: http://www.oiv.int/public/medias/5029/world-vitiviniculture-situation-2016.pdf.

Stenmarck, Å., Jensen, C., Quested, T., Moates, G., Buksti, M., Cseh, B., Juul, S., Parry, A., et al. 2016. *Estimates of European food waste levels. Reducing food waste through social innovation*. [Online], Available: https://www.eu-fusions.org/phocadownload/Publications/Estimates of European food waste levels.pdf%5Cnhttps://phys.org/news/2016-12-quarter-million-tonnes-food-logistics.html#nRlv.

Stevens, K. & Fuller, M. 2015. Thermoregulation and clothing comfort. *Textile-Led Design for the Active Ageing Population*. (January, 1):117–138.

Stroecken, R. 2017. Food Waste in the Fresh Produce Supply Chain. Wageningen University & Research, Wagenigen.

Strydom, J. 2019. *Force Air-cooling: A Summary of Basic Principles*. Paarl, South Africa. [Online], Available: https://www.hortgro.co.za/wp-content/uploads/docs/2019/10/fa-cooling-guidelines-2020.pdf.

Swart, A.E. & Lennox, C.L. 1995. Infection of Table Grape Bunches by Alternaria alternata. *South African Journal of Enology & Viticulture*. 16(1):3–6.

Swift, J.G., May, P. & Lawton, E.A. 1974. Concentric cracking of grape berries. *Vitis*. 13:30–35.

Symington, S. 2008. Staying ahead of the global pack : [creating sustainable competitive advantage in the marketing of South African table grapes to the United Kingdom in the deregulated era]. University of Cape Town, Cape Town. [Online], Available: http://hdl.handle.net/11427/5685.

Tan, P.-N., Steinbach, M., Kumar, V. & Karpatne, A. 2018. Classification: Basic Concepts, Decision Trees, and Model Evaluation. in *Introduction to Data Mining* Second ed. Pearson. 137–192.

TempTale® 4: USB Interface Plus Reader. n.d. [Online], Available: https://www.sensitech.com/en/products/monitors/accessories/.

Tessara. 2022. *Uvasys Green - How does it work?* [Online], Available: https://www.tessara.co.za/products/uvasys/uvasys-blue/ [2022, May 16].

Tharwat, A. 2018. *Classification assessment methods : a detailed tutorial*. [Online], Available: https://www.researchgate.net/publication/327403649_Classification_assessment_methods_a_detailed_tutorial.

Thompson, J.F., Mitchell, G.F., Rumsey, T.R., Kasmire, R.F. & Crisosto, C.H. 2008. *Commercial Cooling of Fruits, Vegetables and Flowers*. Revised Ed ed. Davis, CA: Regents of the University of California. USA.

Tonetto de Freitas, S. & Pareek, S. 2019. *Postharvest Physiological Disorders in Fruits and Vegetables*. S. Tonetto de Freitas & S. Pareek (eds.). Boca Raton: Taylor & Francis, 2018.: CRC Press.

Trading Economics. 2019. *South Africa Unemployment Rate*. [Online], Available: https://tradingeconomics.com/south-africa/unemploymet-rate [2019, June 11].

Trienekens, J. & Willems, S. 2007. Innovation and Governance in International Food Supply Chains: The Cases of Ghanaian Pineapples and South African Grapes. *The International Food and Agribusiness Management Review*. 10(4):42–63.

Trung Thanh, L., Thi Anh Dao, N., Linh-Trung, N. & Vu Le, H. 2017. On the overall ROC of multistage systems. in *2017 International Conference on Advanced Technologies for Communications (ATC)* Vols 2017-Octob. IEEE. 229–234.

Tuychiev, B. 2021. *Comprehensive Guide to Multiclass Classification Metrics*. [Online], Available: https://towardsdatascience.com/comprehensive-guide-on-multiclass-classification-metrics-af94cfb83fbd [2022, August 15].

Valentine, A.G.D.T. & Goedhals-gerber, L. 2017. The temperature profile of an apple supply chain : A case study of the Ceres district. *Journal of Transport and Supply Chain Management*. 11:1–8.

Vanhamäki, S., Heinonen, A., Manskinen, K. & Kälviäinen, M. 2017. Information design as a tool for promoting renewable energy. *The Design Journal*. 20(sup1):1827–1835.

Vercesi, A., Locci, R. & Prosser, J.I. 1997. Growth kinetics of Botrytis cinerea on organic acids and sugars in relation to colonization of grape berries. *Mycological Research*. 101(2):139–142.

Verdugo-vásquez, N., Fuente, C.P., Ortega-farías, S., Talca, U. De & Agrarias, F.D.C. 2017. Model development to predict phenological scales of table grapes (cvs . Thompson , Crimson and Superior Seedless and Red Globe) using growing degree days. *Oeno One*. 51(3):277–288.

Vishwakarma, R.K., Bashir, A.A., Kumar, Y., Yadav, D.S., Sharma, A.K. & Lohakare, N.C. 2022. Development of automated fumigation chamber for treatment of grapes with SO2 and CO2. *Journal of Food Process Engineering*. 45(4).

De Visser, P., Nannes, L., Van Bokhoven, H. & Buwalda, F. 2015. *Decision Support System (DSS) for prevention of Botrytis in tomato in greenhouses*. [Online], Available: https://edepot.wur.nl/339617.

Weaver, K. 2018. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. in B.B. Frey (ed.). 2455 Teller Road, Thousand Oaks, California 91320: SAGE Publications, Inc. B.B. Frey (ed.). 1287–1288.

WFP. 2018. *Nutirition Statistics*. [Online], Available: https://www1.wfp.org/nutrition [2019, June 10].

Wheelan, C. 2014. *Naked Statistics: Stripping the Dread from Data*. 1st ed. W. W. Norton & Company.

Witbooi, W.R., Fourie, J.F. & Taylor, M.A. 2010. Information on the Little Understood Problem of Soft Tissue Breakdown in Cold Stored Table Grapes. *SA Fruit Journal*. 9(2010):45–49.

World Wide Travel Organisation. 2021. *Average Monthly Snow And Rainfall In Worcester (Western Cape) In Millimeter*. [Online], Available: https://weather-and-climate.com/average-monthly-precipitation-Rainfall,worcester-western-cape-za,South-Africa [2021, November 11].

Yarwood, M. 2021. Temperature Controlled Cargo: Managing Risk through the Supply Chain. in *Cool Logistics Global* TT Club. [Online], Available: https://coollogisticsresources.com/.

Ye, J. 2020. *Everything You Need To Know About Correlation*. [Online], Available: https://towardsdatascience.com/everything-you-need-to-know-about-correlation-3ef78f22fcad [2022, May 07].

Zheng, A. 2015. *Chapter 4. Hyperparameter Tuning*. O'Reilly Media, Inc. [Online], Available: https://www.oreilly.com/library/view/evaluating-machine-learning/9781492048756/ch04.html.

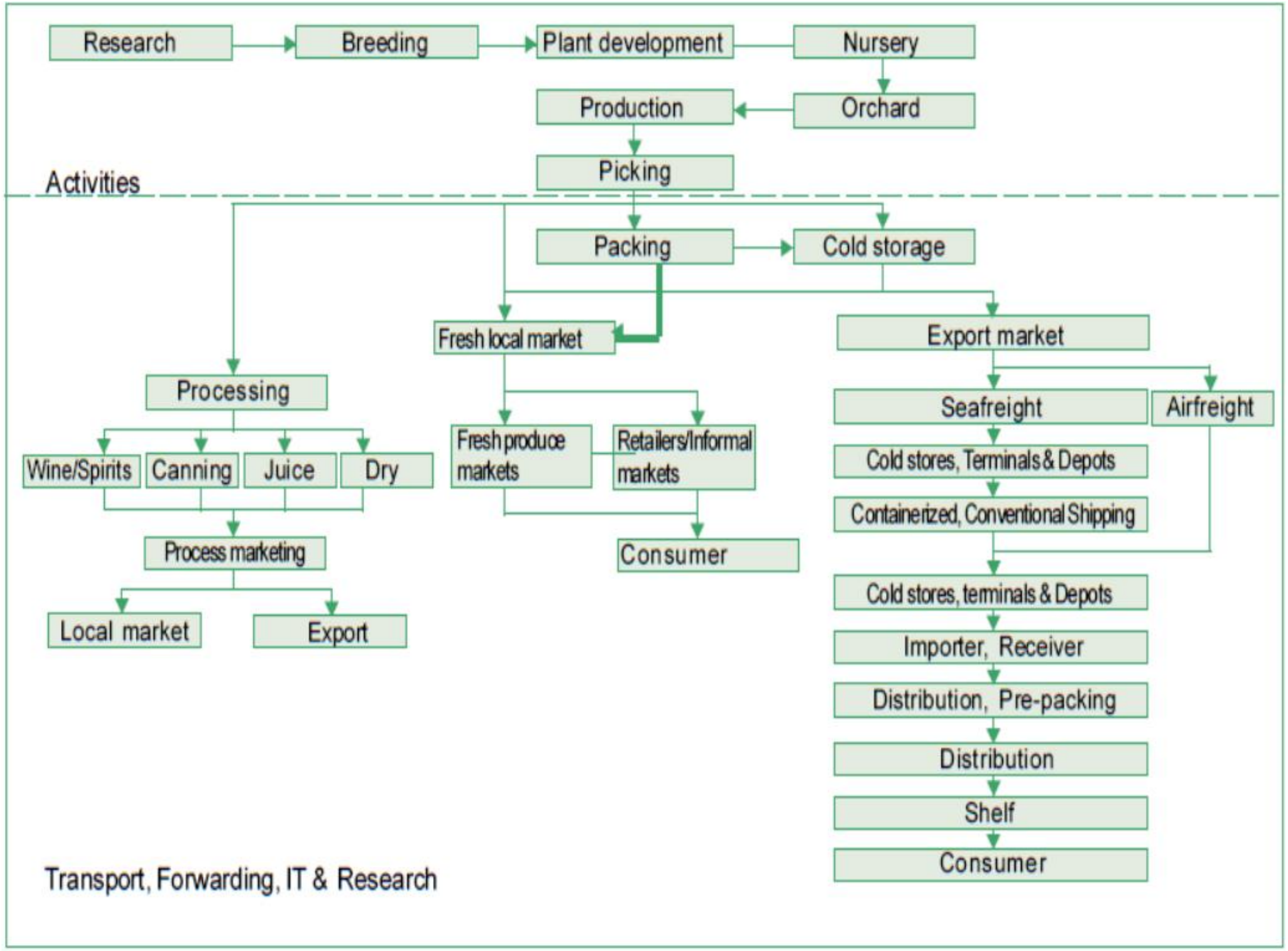
Zikmund, W.G., Babin, B.J., Carr, J.C. & Griffin, M. 2013. *Business Research Methods*. 9th editio ed. South-Western, Cengage Learning.

Zoffoli, J.P. 2008. Postharvest Handling of Table grape. *Acta Horticulturae*. 785(785):415–420.

Zoffoli, J.P. & Latorre, B.A. 2011. Table grape (Vitis vinifera L.). in *Postharvest Biology and Technology of Tropical and Subtropical Fruits* Chile: Elsevier. 179–214.

Žukauskas, P., Vveinhardt, J. & Andriukaitienė, R. 2018. Philosophy and Paradigm of Scientific Research. in *Management Culture and Corporate Social Responsibility* BoD – Books on Demand. 121–135.

Appendix A: Table grape supply chain



Source: DAFF, 2017

Appendix B: Insurance claim account sale

Goods: South African Grapes

Received: 24.02.2021

Size	Cal	Variety	Qty	Price	Value
10x500g	CAT I	South African Tawny	360	€8.00	€2 880.00
		Grape Punnets			
4.5kg	CAT I	South African Tawny	3 060	€8.00	€24 480.00
		Grapes			
		QC 4 - Issues			
		Delayed vessel			
		Container No:-			
		MEDU 9098020			

Total: 3 420 €27 360.00

5% Commission:	€1 368.00
Recycling Costs:	€234.00
Terminal/ Discharge:	€355.00
Clearance:	€159.00
Handling:	€430.00
Transport to Depot:	€175.00
Quality Control:	€239.40
Decay:	€310.36
Repacking Costs:	€720.00
RHD/ Storage:	€184.75
TOTAL COSTS	**€4 175.51**

4.5 = R93

Pu = R71

C&F Hamburg: €23 184.49

Euro Advance Payment: **€14 040.00**

TOTAL FINAL EURO RETURN **€9 144.49**

Sales difference:

(9.40 - 8) x 360 = €504

(8.5 - 8) x 3060 = € 1530

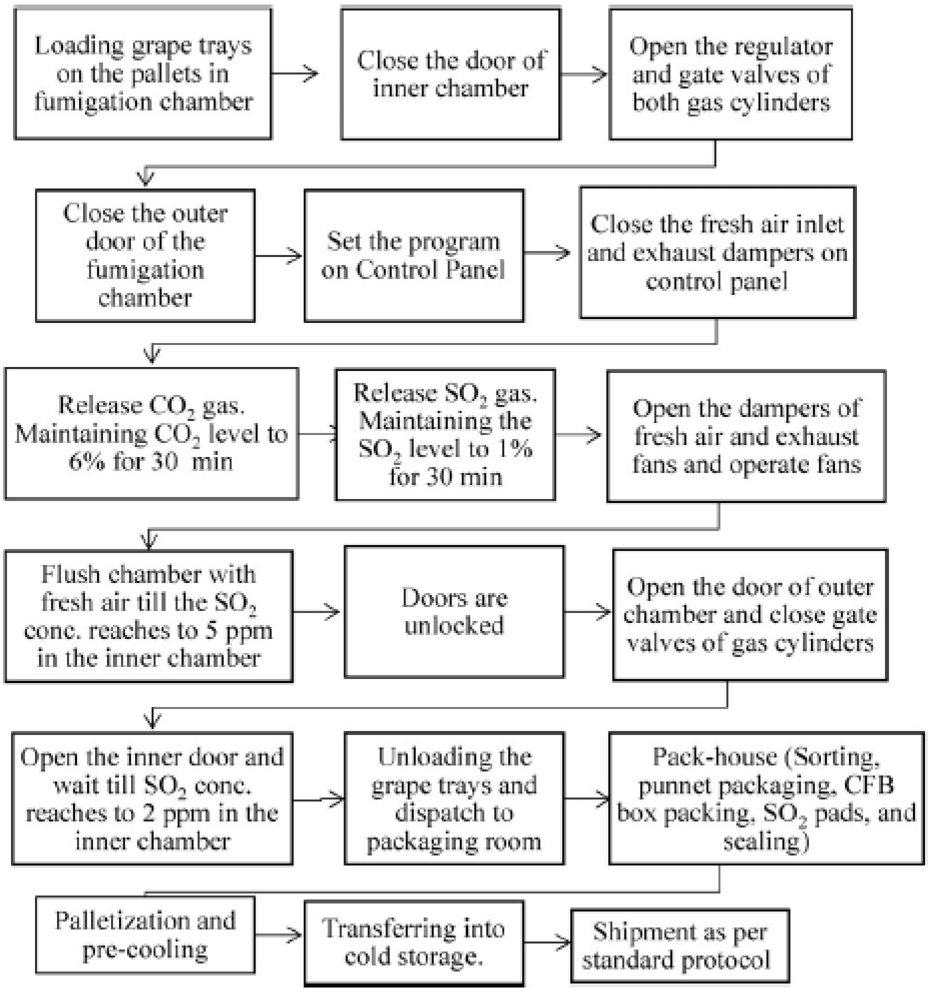
Decay + Repack:

310.36 + 720 = €1030.36

Total € 3064.36

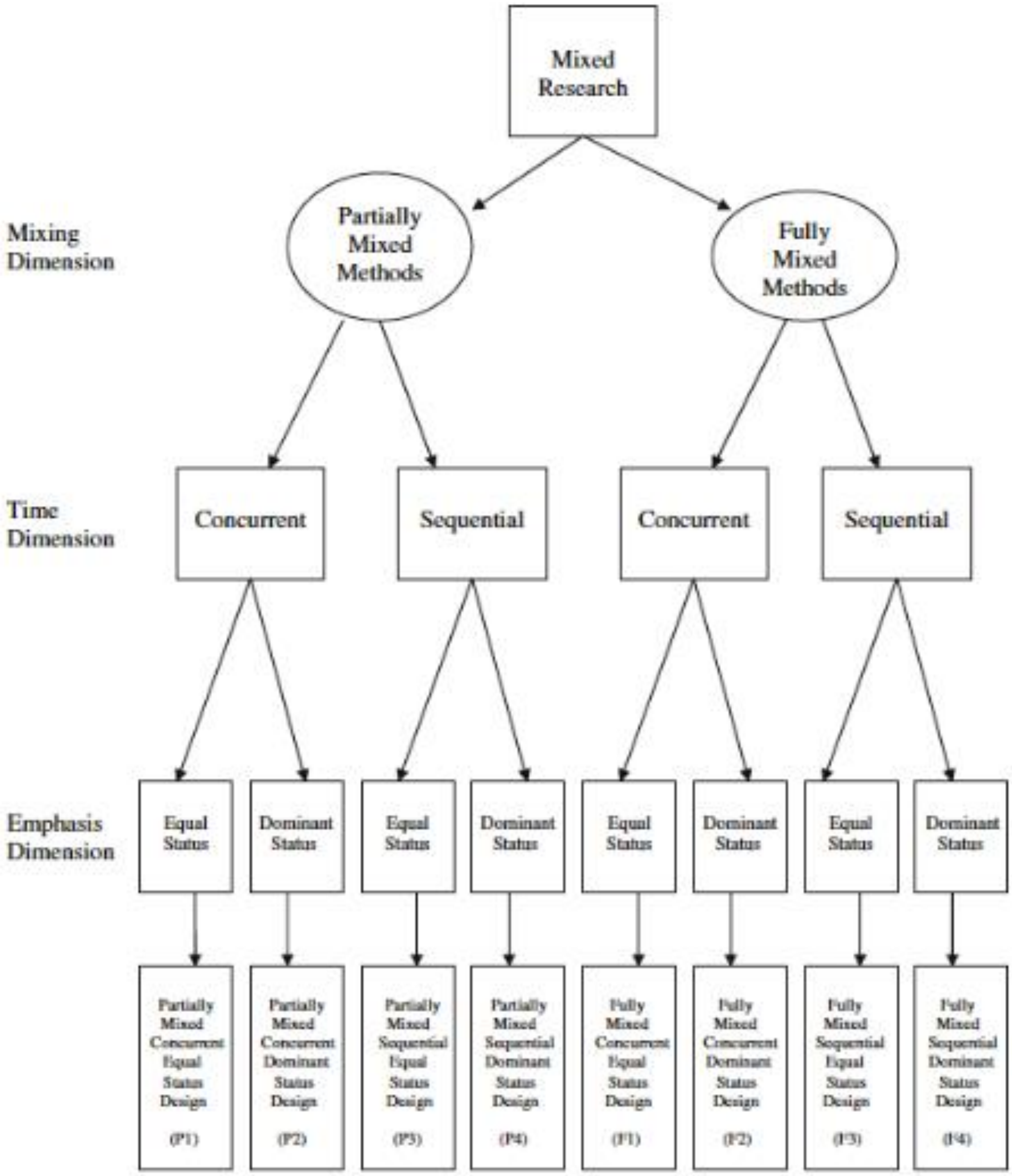
Source: Author's own

Appendix C: Standard operating procedure (SOP) for fumigation treatment of grapes



Source: Vishwakarma, Bashir, Kumar, *et al.*, 2022

Appendix D :Typology of mixed research



Source: Bentahar & Cameron, 2015

Appendix E: Description of variables collected from the five data sources (colour-coded)

REF#	FIELD NAME	DATA TYPES	VARIABLE TYPE	DESCRIPTION	EXAMPLE
V0	NOM_#	QUALITATIVE	NOMINAL	Observation number	657
V1	NOM_Container number	QUALITATIVE	NOMINAL	Container number	TEMU9185430
V2	NOM_Ship name + Voyage	QUALITATIVE	NOMINAL	Ship name and voyage number	SANTA ISABEL 201B
V3	NOM_region	QUALITATIVE	NOMINAL	Geographical area the farm is located	[redacted]
V4	NOM_Farm Name	QUALITATIVE	NOMINAL	Name of farm	[redacted]
V5	NOM_farm	QUALITATIVE	NOMINAL	Farm code	[redacted]
V6	NOM_variety	QUALITATIVE	NOMINAL	Grape variety	CSS
V7	NOM_var_grp	QUALITATIVE	NOMINAL	Colour of grape group (red, green, black)	RS
V8	NOM_size_count	QUALITATIVE	NOMINAL	Berry size	L
V9	NOM_pack	QUALITATIVE	NOMINAL	Type of carton code (Secondary Packaging)	B05P
V10	NOM_inv_code	QUALITATIVE	NOMINAL	Code for type of tertiary packaging	UL
V11	NOM_net mass	QUALITATIVE	NOMINAL	Carton weight	5 kg
V12	NOM_pack_date	QUALITATIVE	NOMINAL	Harvest/packaging date	23/03/2020
V13	NOM_load_date	QUALITATIVE	NOMINAL	Date container was loaded	30/03/2020
V14	NOM_etd	QUALITATIVE	NOMINAL	Estimated Time of Departure	31/03/2020
V15	NOM_atd	QUALITATIVE	NOMINAL	Actual Time of Departure	07/04/2020
V16	NOM_eta	QUALITATIVE	NOMINAL	Estimated Time of Arrival	22/04/2020
V17	NOM_Arriva QC Date	QUALITATIVE	NOMINAL	Date the fruit was QC'd on arrival	24/04/2020
V18	VIN_Brix readings - MIN	QUANTITATIVE	INTERVAL	Minimum fruit sugar reading	18.5
V19	VIN_Brix readings - MAX	QUANTITATIVE	INTERVAL	Maximum fruit sugar reading	21.2
V20	VIN_Brix readings - AVERAGE	QUANTITATIVE	INTERVAL	Average fruit sugar reading	19.8
V21	VIN_sum.Rain_LF (mm)	QUANTITATIVE	RATIO	Sum of rainfall in the last fortnight (two-weeks / 14 days)	0.51
V22	VIN_ET0_LF (mm)	QUANTITATIVE	RATIO	Total evapotranspiration for the past 14 days	60.98
V23	VIN_av.RH_LF (%)	QUANTITATIVE	RATIO	Mean daily relative humidity for the past 14 days	0.5998
V24	VIN_av. Tn_LF (˚C)	QUANTITATIVE	INTERVAL	Mean daily min temperature for the past 14 days	12.75

V25	VIN_av. T_LF (˚C)	QUANTITATIVE	INTERVAL	Mean daily temperature for the past 14 days	21.33
V26	VIN_av. T (˚C)	QUANTITATIVE	INTERVAL	Mean daily temperature (av. of min and max temp.)	20.52
V27	IN_Split berries*	QUANTITATIVE	RATIO	Number of split berries found	0
V28	IN_Soft berries*	QUANTITATIVE	RATIO	Number of soft berries found	0
V29	IN_Slip skin*	QUANTITATIVE	RATIO	Number of berries found with slip skin	0.9
V30	IN_Neck splits*	QUANTITATIVE	RATIO	Number of berries found with neck splits	0
V31	IN_Marks*	QUANTITATIVE	RATIO	Number of berries found with marks	0
V32	IN_Decayed berries*	QUANTITATIVE	RATIO	Number of decayed berries found	0
V33	IN_Berry drop*	QUANTITATIVE	RATIO	Number of loose berries found in carton	0
V34	IN_Soft berries*_ORD	QUANTITATIVE	ORDINAL (1-7)	ORDINAL	1
V35	IN_Berry drop*_ORD	QUANTITATIVE	ORDINAL (1-7)		1
V36	IN_Marks*_ORD	QUANTITATIVE	ORDINAL (1-7)		1
V37	IN_Split berries*_ORD	QUANTITATIVE	ORDINAL (1-7)		1
V38	IN_Slip skin*_ORD	QUANTITATIVE	ORDINAL (1-7)		2
V39	IN_Neck splits*_ORD	QUANTITATIVE	ORDINAL (1-7)		1
V40	IN_Type of Infestation	QUALITATIVE	NOMINAL	Type of infestation found	Spider web
V41	IN_Packing/handling damage	QUALITATIVE	NOM - Dichotomous	Handing related damage found	No
V42	IN_Berry cracking	QUALITATIVE	NOM - Dichotomous	Berry cracking found	No
V43	IN_Stem Condition	QUALITATIVE	ORDINAL	Bunch rachis condition	FRESH
V44	IN_QC Score colour	QUALITATIVE	ORDINAL	QC score	YELLOW
V45	IN_QC Score	QUANTITATIVE	ORDINAL	QC score (1-3)	2
V46	VOY_Spikes (SUM)	QUANTITATIVE	RATIO	1 observation, for a single ten-minute interval, where the temperature recorded was outside the required range	151
V47	VOY_Breaks (SUM)	QUANTITATIVE	RATIO	9 sequential ten-minute intervals where the recorded temperature was outside the required range	119
V48	VOY_Fruit age at QC*	QUANTITATIVE	RATIO	actual age of fruit at the time of the arrival quality inspection (days)	32
V49	AR_Stem Condition	QUALITATIVE	ORDINAL	Bunch rachis condition	YELLOW
V50	AR_Soft berries ORD	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of soft berries found	5

V51	AR_Soft berries %	QUANTITATIVE	RATIO	percentage of soft berries found in carton	0.2
V52	AR_Shrivelling	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of shrivelled berries found	2
V53	AR_Packing/handling damage	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of packing/handling damage found	3
V54	AR_Marks	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of marked berries found	3
V55	AR_Kind of decay 1	QUALITATIVE	NOMINAL	kind of decay found	Slip skin
V56	AR_Kind of decay	QUALITATIVE	NOMINAL	kind of decay found	Botrytis
V57	AR_Decayed berries*	QUANTITATIVE	RATIO	Number of decayed berries found	4.5
V58	AR_Bruising	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of bruised berries found	3
V59	AR_Berry drop	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of loose berries found	4
V60	AR_Berry drop weight (grams)	QUANTITATIVE	CONTINUOUS	Weight of the loose berries found	
V61	AR_Berry cracking	QUANTITATIVE	ORDINAL (1-7)	Rating of the extent of berries found with cracks	1
V62	AR_QUALITY SCORE COLOURS	QUALITATIVE	ORDINAL	Arrival score in colour (Green, Amber, Red)	AMBER
V63	AR_QUALITY SCORE	QUANTITATIVE	ORDINAL (1-3)	Colour score decobed onto 3-point likert scale (1-3)	2
V64	AR_Original QC score	QUANTITATIVE	ORDINAL (1-6)	Arrival quality score on a 6-point lickert scale	4
V65	AR_Quality rating	QUALITATIVE	ORDINAL	Arrival quality score rating	less than average
V66	AR_ISSUES	QUALITATIVE	NOMINAL	Description of issues found	Decay
V67	AR_STORAGE POTENTIAL	QUALITATIVE	ORDINAL	Potential shelf life	short
V68	AR_Marketing Advice	QUALITATIVE	NOMINAL	Quality advice message	Marketing problems are possible
V69	AR_Claim_type - data base	QUALITATIVE	NOMINAL	Exporter data - QC indicates that arrival condition of the fruit resulted in a lower sales price	QC
V70	AR_claim_code - data base	QUALITATIVE	NOMINAL	Exporter data - QC indicates that arrival condition of the fruit resulted in a lower sales price and cause	QCDEC

Source: Author's own

Appendix F: Account sale export document example

ANTON DÜRBECK gmbh

ACCOUNT SALE

ANTON DÜRBECK GmbH Hessenring 120 DE - 61348 Bad Homburg v.d.H.
HOEKSTRA FRUIT EXPORTERS
(PTY) LTD , HEXRIVER
PO Box 7178
Paarl North 7623
Südafrika

Page:	1
Document Date:	09.05.14
Master Batch No.:	1400541
Voyage No.:	1414039
Vessel departed:	MOL CULLINAN
Vessel arrived:	MOL CULLINAN
Mode of transport	C.I.F.
Departure date:	17.03.14
Arrival Date:	03.04.14
Port:	Rotterdam
Creditor's no.:	614377
Your VAT No.:	
Our VAT No.:	DE 114118807

Result

Batch	Product Brand	Variety Size	Orige	Cat Color	Unit	Quantity	Price	Total Value
14001874	Grapes SUNSHINE	Bonheur L	ZA	I	Kollo 4,5kg Poly duty paid	720		
14001875	Grapes SUPERGRAPE	Bonheur XL	ZA	I	Kollo 4,5kg Poly duty paid	2.720		
Total Result						**3.440**		

Costs

		Total
Forwarding		97,80
Terminal Handling Charge		270,00
Coolstorage		260,00
Merchant haulage	R'dam - de Meern	309,75
Costs for Survey Report		85,00
Laboratory Costs		390,00
Total costs		**1.412,55**
Commission 7%		2.273,71

Total Result EUR

CONTAINER LIST:
DAYU 671388-2

Es gelten unsere allgemeinen Geschäftsbedingungen.

ANTON DÜRBECK GmbH
Geschäftsführer
Karl Dürbeck, Mathias Dürbeck
Telefon (0 61 72) 68 06-0
Telefax (0 61 72) 68 06-800
E-Mail: anton@duerbeck.com

Reg. No.: 0022840160
VAT Reg. No.: DE 114118807
Amtsgericht: Frankfurt
HRB: 9762

Commerzbank AG
BSB: 50050000
Acc.: 420730000
IBAN: DE09500800000420730000
Swift: DRESDEFF

HypoVereinsbank AG
BSB: 50320191
Acc.: 8406510
IBAN: DE05503201910008406510
Swift: HYVEDEMM430

Source: Müssman, 2015:157

Appendix G: Example of a QC where the raw arrival data was extracted from

INBOUND - QUALITY / CONTROL SUMMARY

Our reference : 270834
Motorvessel : Santa Ursula 200B
Origin(s) : South Africa
Inspection date(s) : 11-03-2020, 12-03-2020, 13-03-2020, 16-03-2020

Consignee
Shipper

Grapes, IFG 6 (Sweet Sapphire™), 4.50 kg cartons (paper bags)

Number	Reference	Grower(s)	Packhouse(s)	Brand(s)	Cat	Decay	Issues	Storage	Summary	Size (quantity)	Score	Color
270834-05-04	519101432				1	much found during inspection	Decay, Shrivelling	none	Marketing problems are to be expected.	xl (180)	5	

Color explanation	fit for normal marketing	needs attention	unacceptable

1 / 3

INBOUND - QUALITY / CONTROL REPORT

GENERAL INFORMATION

Our reference	: **270834-05-04**	**Your reference**	: **519101432**
Motorvessel	: **Santa Ursula 200B**	**Consignee**	:
Date	: 11-03-2020	**Location**	:
Shipper	:	**Origin**	: South Africa
Commodity	: Grapes	**Variety**	: IFG 6 (Sweet Sapphire™)
Brand(s)	:	**Quantity**	: 180 cartons
Category	: 1	**Size (quantity)**	: xl (180)
Packhouse(s)	:		
Grower(s)	:		
Pack code(s)	:		
Orchard/field(s)	:		
Container(s)	: MNBU 020773-2		

PACKING AND TEMPERATURE

Pulp temperature : -0.2 °C to +0.3 °C, average 0.0 °C
Recorder(s) : found
Packing : 4.50 kg cartons (paper bags)
Palletizing : fair
Net weights per carton : 3.93 to 4.69 kg, average 4.42 kg
some underweight
Category on carton : present
GGN on carton : present
Shipper's address on carton : present

PRODUCT OBSERVATIONS

First impression : acceptable
Texture/consistency : 80% firm, 20% slightly sensitive
Ripeness/maturity : 100% ripe
Colour : purple-black
Stem colour : green/brownish
Brix readings : 15.1 to 20.9%, average 19.4%
Berry size : size xl: 13 to 21 mm, average 18 mm
Bunch size : size xl: medium to extra large
Skindefects/deviations : category 1: ± 5 to 10%, being within the limits for the category indication
Bruising : ± 0 to 2% (to a light extent)
Packing/handling damage : frequent (mainly to a light extent)
Berry drop : ± 0 to 5% loose berries
Berry cracking : not found during inspection
SO2 bleaching : not found during inspection
Shrivelling : ± 5 to 10%
mainly present in every bag
Decay : much found during inspection
up 32, average 24 decayed berries per carton, found in every bag
clusters of nest decay present
General remarks : No remarks

GENERAL IMPRESSION

QUALITY SCORE : **5**
STORAGE POTENTIAL : **none**
ISSUES : **Decay, Shrivelling**
SUMMARY : **Marketing problems are to be expected.**

Scale 1 = excellent 2 = good 3 = acceptable 4 = less than average 5 = poor 6 = total loss

Appendix H: VBA code to copy text from multiple text files to excel

```
Option Explicit
Option Compare Text

Sub Main()
  Dim Path As String, FName As String
  Application.ScreenUpdating = False
  Path = ThisWorkbook.Path & "\"
  FName = Dir(Path & "*.txt")
  Do While FName <> ""
    ProcessTextFile Path & FName
    FName = Dir()
  Loop
  Rows(1).Orientation = 90
  Range("A1").CurrentRegion.EntireColumn.AutoFit
  Application.ScreenUpdating = True
End Sub

Sub ProcessTextFile(ByVal FName As String)
  Dim ff As Integer
  Dim Contents As String
  Dim Lines
  Dim i As Long, f As Long, t As Long, a As Long, b As Long, c As Long
  Dim Section As String, Key As String, Arg

  ff = FreeFile
  Open FName For Binary Access Read Lock Write As #ff
  Contents = Space(LOF(ff))
  Get #ff, , Contents
  Close #ff

  StoreField "", "File", Mid$(FName, InStrRev(FName, "\") + 1)

  Lines = Split(Contents, vbCrLf)
  Section = ""
```

```
  For i = 0 To UBound(Lines)
    f = 1
    t = InStr(f, Lines(i), ":")
    If t = 0 Then
      Section = Trim$(Lines(i))
    Else
      Do While t > 0
        Key = Trim$(Mid(Lines(i), f, t - f))
        b = InStr(t + 1, Lines(i), ",")
        c = InStr(t + 1, Lines(i), ":")
        If b > 0 Then
          a = b - 1
        ElseIf c > 0 Then
          a = InStrRev(Lines(i), " ", c)
        Else
          a = Len(Lines(i))
        End If
        Arg = Trim$(Mid(Lines(i), t + 1, a - t))
        f = a + 1 + IIf(b > 0, 1, 0)
        t = InStr(f, Lines(i), ":")

        If IsDate(Arg) Then
          Select Case Key
            Case "AMH"
              'Ignore, this is a version number
            Case Else
              Arg = CDate(Arg)
          End Select
        ElseIf IsNumeric(Arg) Then
          Select Case Key
            Case "Time"
              'Convert "1234" to Excel time
              Arg = Format(CDbl(Arg) / 2400, "hh:mm")
            Case Else
              Arg = CDbl(Arg)
          End Select
```

```
        End If

        StoreField Section, Key, Arg
      Loop
    End If
  Next
End Sub

Sub StoreField(ByVal Section, ByVal Key, ByVal Arg)
  Dim Header As Range, Dest As Range
  If Section <> "" Then Key = Section & "_" & Key
  Set Header = Rows(1).Find(Key, LookIn:=xlValues,
LookAt:=xlWhole)
  If Header Is Nothing Then
    If IsEmpty(Range("A1")) Then
      Set Header = Range("A1")
    Else
      Set Header = Cells(1, Columns.Count).End(xlToLeft).Offset(,
1)
    End If
    Header = Key
  Else
  End If
  If Dest Is Nothing Then
    Set Dest = Range("A" & Rows.Count).End(xlUp)
    Set Dest = Header.Cells(Dest.Row, 1)
    If Dest.Column = 1 Then Set Dest = Dest.Offset(1)
  End If
  Dest = Arg
End Sub
```

Source: Killer, 2019

Appendix I: Content analysis contingency table from the arrival QC reports

AR_ISSUES (group)	GREEN	AMBER	RED
None	144	2	
Decay	1	146	60
Decay / Shrivelling		11	48
Berry drop / Decay		7	7
Berry cracking / Decay		1	4
Decay / Texture/consistency		1	3
Colour / Decay		3	
Bruising / Decay		2	
Decay / Packing/handling damage		2	
Berry cracking / Decay / Texture-consistency		1	
Decay / Waterberries		1	
Decay / Discoloured paper bags caused by leaking berries/ Shrivelling		1	
Decay / Shrivelling / Weights per carton		1	
Bags easily torn open / Decay		1	
Bunch size / Decay		1	
Decay / Texture		1	
Decay / First impression / Texture/consistency			1
Berry drop / Condensation / Decay /Shrivelling			1
Decay / Stem colour			1
Decay / Sunburn/sunblush			1
Label on punnets	1		
Palletizing	1		
Weights per punnets	1		
Texture/consistency	1	1	
Berry drop / Condensation		1	
Closed zip lock bags / SO2 bleaching		1	
Colour		1	
Bruising / Weights per carton / Weights per punnets		1	
Poorly coloured		1	
Shrivelling		2	
Bruising / Packing/handling damage		2	
Grand Total	149	192	126

Source: Author's own

Appendix J: The total temperature spikes and breaks for each container

Container number	Total spikes	Total breaks
MNBU9060175	16	2
KKFU6784741	20	4
MNBU3403894	21	4
SUDU8035960	28	4
MORU1110397	21	5
MWCU6741346	15	7
TTNU8248774	22	8
NYKU7124867	22	10
TCLU1233073	33	11
MSWU9068409	30	14
CAIU55018596	31	14
CXRU1349590	31	14
CRSU6151021	47	15
TTNU8245712	32	16
TRLU1682793	34	16
TTNU8537308	34	16
CXRU1222482	33	17
MORU1142178	35	19
MNBU0338157	37	20
SUDU5054837	45	21
CRSU6138457	36	24
MNBU3851728	41	25
TTLU1094886	43	27
BMOU9777185	45	27
MNBU3694608	56	27
TCLU1325837	45	29
SZLU9711738	49	29
MORU1317933	55	32
MWCU5296776	51	35
MNBU3997409	52	35
SEGU9091778	52	36
TEMU9344851	63	37
TLLU1120199	58	38
SZLU9234640	56	40
MSWU1034336	61	40
MNBU3601544	82	40
CAIU5445937	63	47
CRLU1414583	88	48
TTNU8341927	65	49
MWCU6821647	73	49

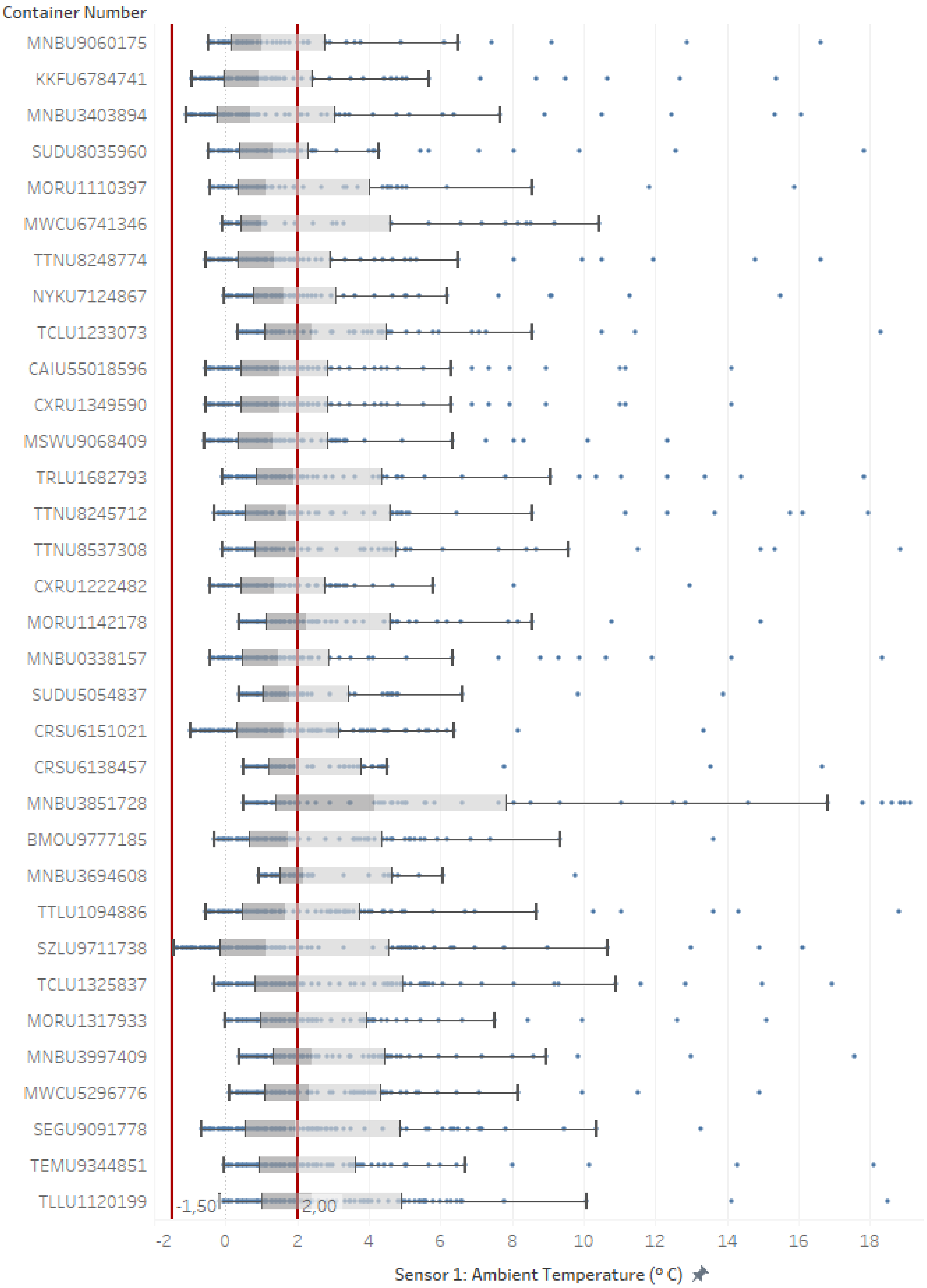
Container number	Total spikes	Total breaks
PONU4954907	80	50
BMOU9844534	68	52
TCLU1363246	76	52
KKFU6575274	77	53
MNBU3746698	81	54
SEGU9237008	79	55
SZLU9443212	76	56
CAIU5466971	90	57
TEMU9169300	81	59
TTNU8318486	84	60
TRIU8070199	88	60
SUDU6197690	88	61
TLLU1122925	86	62
TEMU9571367	87	63
SUDU5155494	139	64
TTNU8298841	81	65
MNBU0343831	85	66
MORU1317949	103	67
MNBU3160578	97	68
MNBU4093005	102	74
KKFU6756035	84	75
MWCU6980675	91	75
NYKU7149340	91	75
MNBU0071693	103	76
TRIU8027464	108	76
CXRU1261555	102	78
MNBU0560349	101	80
MNBU3195553	143	82
SUDU6294030	114	90
TCLU1274540	108	92
NYKU7139208	132	97
TEMU9403717	132	97
TTNU8158726	139	101
SZLU9586024	119	102
KKFU6760740	127	103
TTNU8019968	139	103
MORU1146893	133	105
SUDU8144385	146	105
MNBU0583581	122	106
MORU0601523	122	106

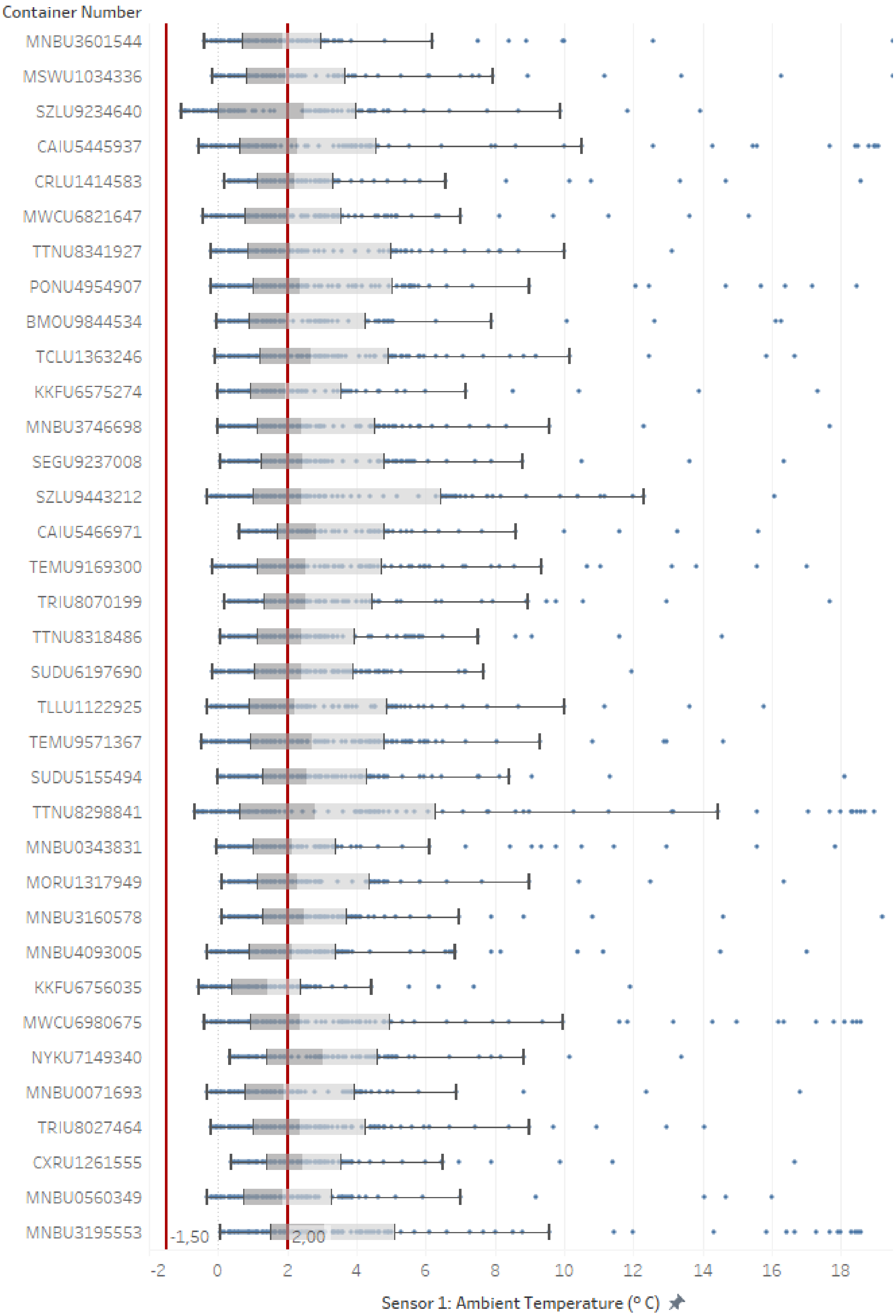
Container number	Total spikes	Total breaks
KKFU6773259	160	106
SZLU9635788	131	107
MORU1307874	133	108
MORU1111135	143	111
TCLU1241464	139	115
MORU5813715	140	117
TCLU1352895	189	117
TEMU9185430	151	119
SUDU8145695	175	123
MNBU9039861	164	124
TTNU8000514	166	126
MNBU3465324	193	131
MORU1130726	172	132
MNBU3087116	170	138
TGHU9918930	182	142
SUDU5098305	177	144
MNBU3822170	175	151
TTNU8003241	178	152
CHIU9036741	223	152
MNBU0113356	191	159
TRIU8010590	203	159
CXRU1058677	179	162
KKFU6744097	199	162
SUDU6198314	212	170
SUDU6021798	208	183
SUDU6202918	208	183
MNBU3815269	215	186
TRIU8975485	273	195
MNBU3544125	233	198
MNBU0207732	310	199
SUDU6069364	219	200
KKFU6760077	251	206
BMOU9753162	268	230
SUDU8053665	288	230
MNBU3708050	319	255
SUDU6058061	341	266
SUDU8176103	318	278
TRLU1687835	303	280
MNBU0318361	331	294
SZLU9696799	450	383
KKFU6724716	445	421
MWCU6818197	615	481

Container number	Total spikes	Total breaks
SZLU96866430	785	616
MORU1308551	1125	686
SUDU8163699	840	704
TTNU8345877	871	833
SZLU9667023	927	879
SUDU8235192	1231	1202
SUDU8087916	1386	1295
KKFU6961750	1653	1587
AKLU6902135	2479	2451
BMOU9749584	2486	2461
SZLU9231744	2865	2843
SZLU9616371	3050	3014
TRIU8098766	4031	3998

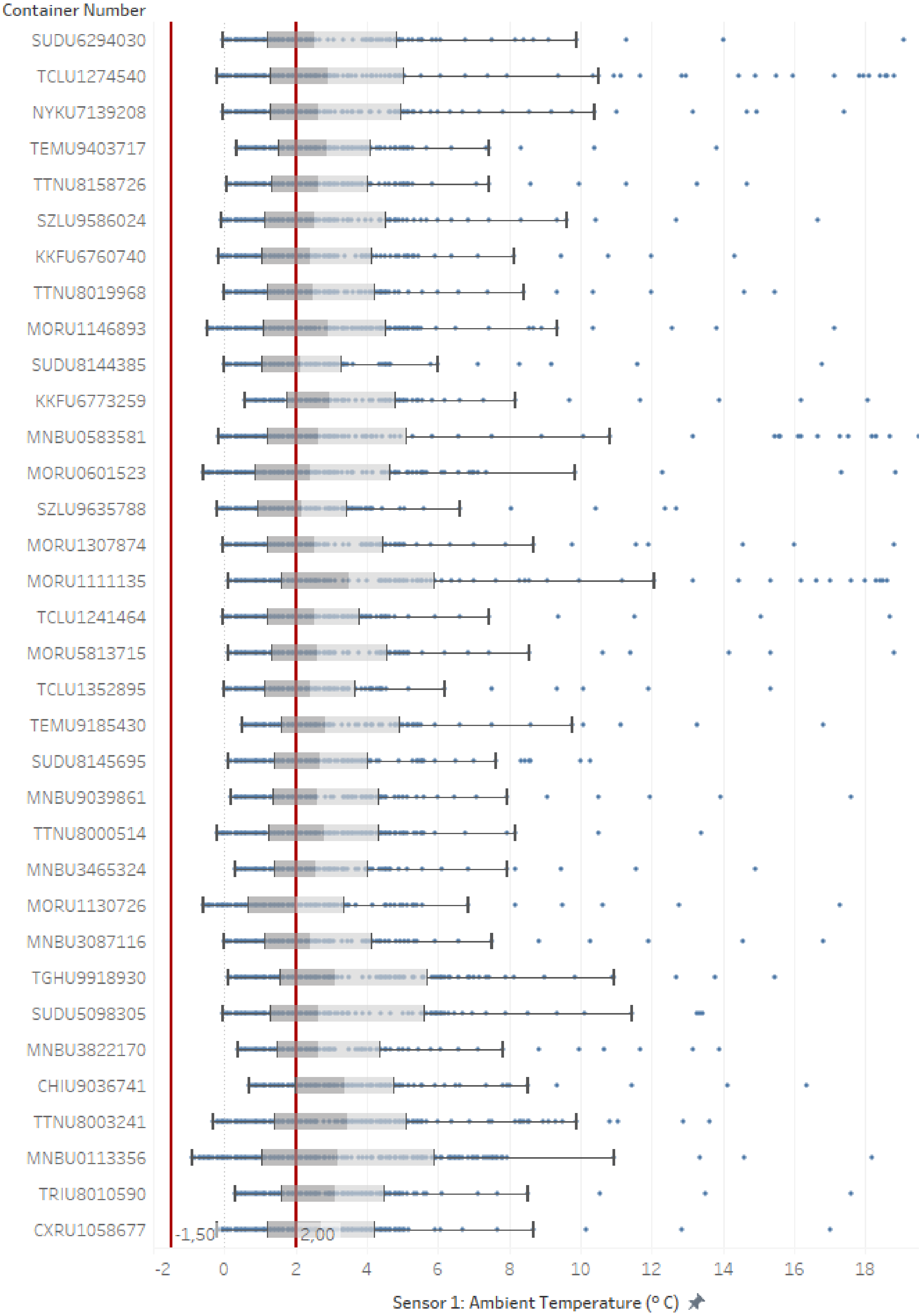
Source: Author's own

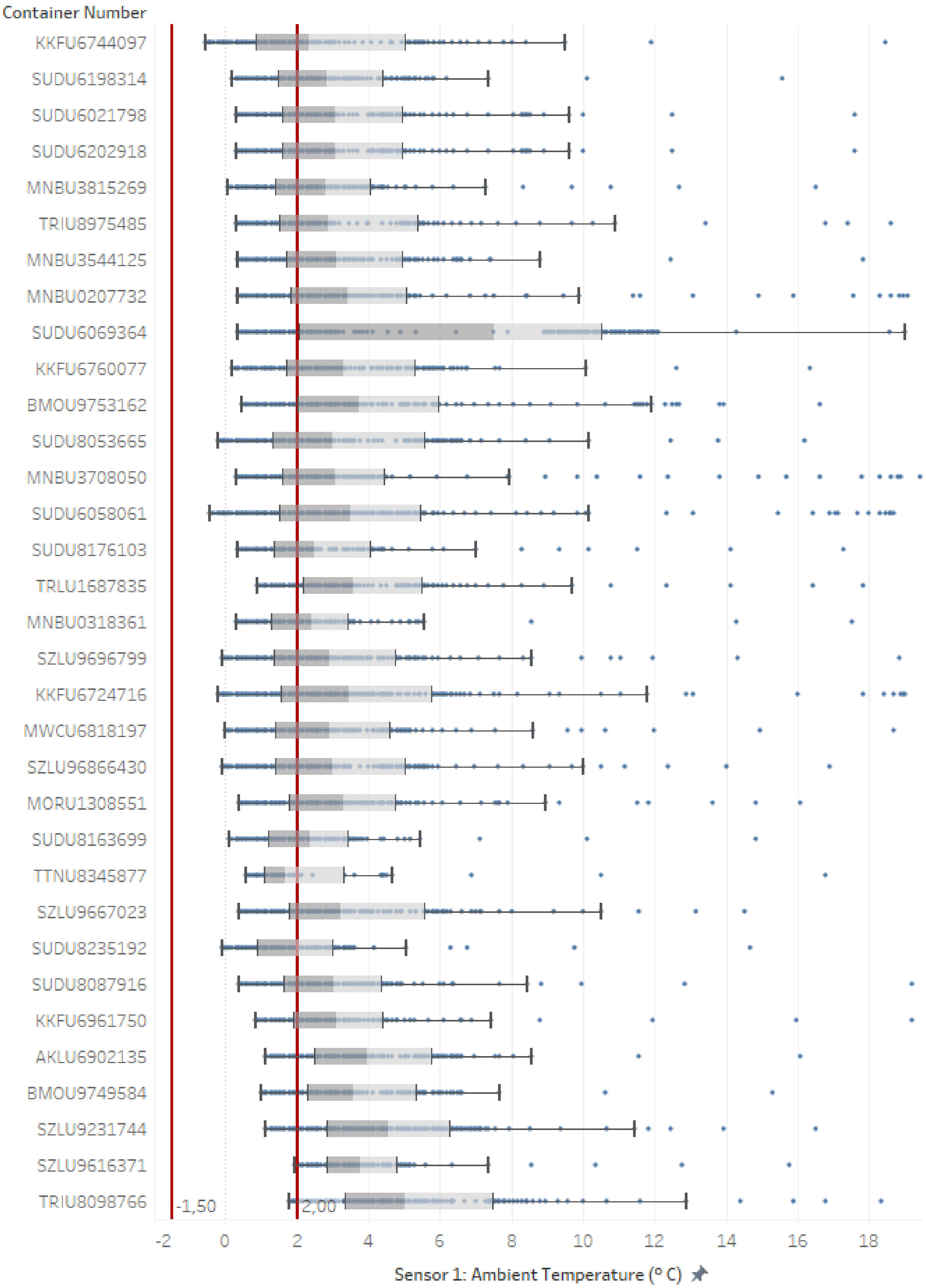
Appendix K: Box-and-whisker plots of the temperature breaks in the supply chain for each container





Container Number
MNBU3601544
MSWU1034336
SZLU9234640
CAIU5445937
CRLU1414583
MWCU6821647
TTNU8341927
PONU4954907
BMOU9844534
TCLU1363246
KKFU6575274
MNBU3746698
SEGU9237008
SZLU9443212
CAIU5466971
TEMU9169300
TRIU8070199
TTNU8318486
SUDU6197690
TLLU1122925
TEMU9571367
SUDU5155494
TTNU8298841
MNBU0343831
MORU1317949
MNBU3160578
MNBU4093005
KKFU6756035
MWCU6980675
NYKU7149340
MNBU0071693
TRIU8027464
CXRU1261555
MNBU0560349
MNBU3195553
-1,50
2,00
-2
0
2
4
6
8
10
12
14
16
18
Sensor 1: Ambient Temperature (° C)


Container Number
SUDU6294030
TCLU1274540
NYKU7139208
TEMU9403717
TTNU8158726
SZLU9586024
KKFU6760740
TTNU8019968
MORU1146893
SUDU8144385
KKFU6773259
MNBU0583581
MORU0601523
SZLU9635788
MORU1307874
MORU1111135
TCLU1241464
MORU5813715
TCLU1352895
TEMU9185430
SUDU8145695
MNBU9039861
TTNU8000514
MNBU3465324
MORU1130726
MNBU3087116
TGHU9918930
SUDU5098305
MNBU3822170
CHIU9036741
TTNU8003241
MNBU0113356
TRIU8010590
CXRU1058677
-1,50
2,00
-2
0
2
4
6
8
10
12
14
16
18
Sensor 1: Ambient Temperature (° C)



Source: Author's own

Appendix L: 2022 Hex River vine census and year-on-year change

CULTIVAR	HECTARES PLANTED 2021	HECTARES PLANTED 2022	% YoY* CHANGE	AGE 0-2	AGE 3-9	AGE 10-15	AGE 16+
Crimson Seedless	2 443	2 394	-2%	8%	47%	28%	17%
IFG 68-175 (Sweet Celebration™)	376	432	15%	37%	60%	3%	1%
Sugrathirtyfive (Autumncrisp®)	340	420	24%	44%	55%	0%	1%
IFG Ten (Sweet Globe™)	311	354	14%	30%	67%	2%	2%
IFG Seventeen (Sweet Joy™)	277	280	1%	2%	98%	0%	0%
Sugranineteen (Scarlotta Seedless®)	271	263	-3%	4%	59%	36%	1%
Sheegene 20 (Allison™)	232	221	-5%	4%	88%	8%	0%
Sugrasixteen (Sable Seedless®)	210	211	1%	11%	61%	21%	7%
Starlight	168	166	-1%	3%	49%	47%	0%
Red Globe	147	145	-1%	5%	6%	44%	45%
Arrafifteen (Arra Sweeties™)	142	145	2%	4%	92%	4%	0%
Arratwentynine (Arra Passion Fire™)	140	140	0%	27%	73%	0%	0%
Joybells	133	119	-10%	6%	90%	1%	2%
Tawny Seedless	124	110	-11%	1%	99%	0%	0%
Sundance Seedless	104	88	-15%	0%	39%	44%	17%
Autumn Royal	97	87	-11%	0%	12%	81%	7%
IFG Six (Sweet Sapphire™)	91	61	-32%	0%	98%	2%	0%
Sheegene 21 (Ivory™)	34	61	81%	60%	39%	0%	0%
Regal Seedless	59	58	-3%	4%	0%	4%	92%
Sugraone (Superior Seedless®)	65	54	-17%	2%	5%	65%	28%
Other	641	584	-9%				
GRAND TOTAL	**6 406**	**6 394**	**0**	**846**	**3 485**	**1 280**	**782**
% Split per age tier				13%	55%	20%	12%

Source: SATI, 2022