The benefits of improved demand forecasting: a case study

Capstone Project

Precious Enahoro

Minerva Schools at KGI

Computational Sciences (Data Science and Statistics) and Business (Scalable Growth)

March 2020

Abstract

Two of the largest South African supermarkets — Shoprite and Pick n Pay — use Enterprise Resource Planning (ERP) systems for demand forecasting but these models are sub-par as they have led to high overstock, understock, wastage and reduced customer satisfaction in the country's supermarket industry. To combat this, enhanced demand forecasting methods that allow for the inclusion of relevant demand-influencing variables that would make the forecasts more robust and accurate are needed in the industry. This is especially important, given an in-depth analysis of the South African supermarket industry that showed the adverse effect of the country's economy on the industry's growth, high bargaining power of customers and stiff rivalry among the biggest supermarkets in the industry.

In this project, two algorithms, Holt-Winters Exponential Smoothing and Prophet were implemented using Walmart demand data as a proxy for South African supermarket data, with the hypothesis that Prophet would perform better, given that it allows for the inclusion of relevant exogenous variables to make the forecast. Cost of Forecast Error (CFE) and Mean Average Error (MAE) were used to evaluate the models. Prophet outperformed Holt-Winters Exponential Smoothing as hypothesized, saving the supermarket $64,060 in costs using the CFE and an MAE of 89, which is about half that of the Holt-Winters model.

To summarise, this paper would give an in-depth analysis into the South African supermarket industry and highlights the potential benefits of adopting improved demand forecasting models in the industry through an implementation and evaluation of the performance of the two forecasting models mentioned above.

Keywords: Demand forecasting, time series, South Africa, supermarkets, industry analysis, Prophet, Holt-Winters

Acknowledgement

I would like to thank God for life, strength, wisdom and all the people He put in my path to help me successfully complete this project.

I would like to specially thank my advisor, Raquel Ribeiro for all the detailed feedback, many office hours, constant encouragement and pushing me to refine this Capstone what it is now.

I would like to thank all the Minerva faculty I reached out to throughout this process for all the help and useful advice.

Finally, I would like to thank my peers, friends and family for all the encouragement, emotional support, feedback and prayers throughout this entire period as it really kept me going.

Table of Contents

1 Introduction

1.1 Why is Time Series Forecasting important for the supermarket industry?

Players in the supermarket industry need to continuously take steps to improve their profitability, either through growing revenue or cutting costs by becoming more operationally efficient. This is imperative because the industry is competitive and typically low margin, meaning that profit is not made by selling individual products but from the volume of products sold in those stores (Bean-Mellinger, 2018; Marketplace, 2013; Reagan, 2013).

Most of the products sold in supermarkets are perishable, and so, the supermarkets have to ensure that they have the right amount of products to satisfy their customers, enough safety stock of products for periods where demand rises higher than expected and ensure they can sell as many of these products before they spoil, to avoid wastage. Arunraj, Ahrens & Fernandes (2016) implemented time series forecasting models to predict the daily sales of bananas in a discount retail store in Germany using relevant exogenous variables like weather, with the aim of reducing food waste. Additionally, most of the players in the industry sell similar items so they have to ensure that products their customers want are always available, as the customers can easily switch to another store that would have those products because of the high bargaining power of customers in the industry (Corporate Finance Institute, 2017). Finally, there are huge costs associated with overstock, which can be categorised as high inventory and storage costs, costs involved in disposing the waste and potential revenue lost, and understock or stockouts, which can be categorised by lost sales, loss of loyal customers and loss of the possible future revenue that would have gotten from the customers. These costs are as high as $1.1 trillion dollars in the global retail industry (Ryan, 2015; Schuman, 2015; WNS DecisionPoint, 2020).

Finding the optimal number of products to stock for each item in the supermarket's product portfolio to satisfy customer demand, maximise revenue and minimise costs incurred is usually very complicated. Variables like the high number of different Stock Keeping Units (SKUs) or products in supermarkets and other variables like promotions, holidays, economic situations and the addition of new products that can either improve and cannibalise the sales of existing products adds another layer of complexity to forecasting and finding this balance.

This is where demand forecasting comes in, as proper demand forecasting has significantly improved retailers performance. For example, Oracle Retail Demand Forecasting (RDF) Cloud Service used automated forecasting to suggest that a major retailer could use 345,000 less SKUs to get the same $480 million revenue, which resulted in significant cost savings (Louis, 2019). More accurate forecasts can also help to reap a $10 million benefit in a billion dollar business (Danziger, 2019). A Mexican company lost about 7.73% of their sales to forecast inaccuracies (Treviño et al., 2014), further highlighting the importance of good demand forecasting. Demand forecasting has historically been done with traditional statistical methods like ARIMA and exponential smoothing, and in recent times, machine learning and deep learning are used to get more accurate forecasts. For example, Alibaba uses IoT for demand forecasting (WNS DecisionPoint, 2020). Adding a variable like weather can increase forecast accuracy by about 20% and Marks and Spencer is taking advantage of this (Danziger, 2019). It is important to note that the classical time series forecasting models do not allow for extra relevant variables to be included in the prediction. This is disadvantageous as there are a lot of variables that can greatly spike or reduce demand for products, for example, social media comments or a pandemic causing people to stock up on food and other essentials. While these might be one-off events, other variables like holidays, weather and the economic situation in a country can also affect demand more regularly and it is important that these variables are included in the forecasting models. This would help to understand the observed time series better and produce the most accurate and robust forecasts possible (Arunraj, Ahrens & Fernandes, 2016; Brownlee, 2016; Louis, 2019; Singh, 2019). Models that allow for these relevant variables to be included while making the demand forecasts are what I refer to as 'improved demand forecasting models'.

Better demand forecasting also helps supermarkets properly schedule employees to ensure that there are enough staff on ground to handle customer demand, especially during peak seasons. Where customer demand is variable, a company having this information would be very essential in planning around employee scheduling. For example, part-time employees can be hired during peak seasons to ensure customers are satisfied but a lot of extra costs are not generated as a result since hiring full time staff would be more costly. Finally, accurate forecasts help supermarkets properly price their products and decide on the best time to carry out advertising and other promotions, in order to generate more revenue. Using other data science techniques like market basket analysis, supermarkets can find out what products are frequently bought together. With that information and knowing when a product is in its peak demand season, the supermarket can create upsell and cross-sale opportunities. Additionally, they can decide the best kind of promotion, pricing, and product shelf placement to implement in order to boost the sales of both products in consideration. Proper forecasting can also help supermarkets know when products would be in low demand, which can help them come up with effective promotion strategies to boost the sales in that period.

1.2 Why focus on the South African supermarket industry?

The global supermarket industry is expected to reach about 12 trillion USD in 2020 and has been showing steady growth in previous years (Grand View Research, 2017; Nicasio, 2019), which shows tremendous promise for the growth of this industry in coming years. The Deloitte's Global Powers of Retailing 2019 Report highlights the top two hundred and fifty retailers in the world and these retailers had a total minimum revenue of $3.7 billion in 2017 (Deloitte, 2019)[1]. According to this report, four of the top five retailers in the world are based in the USA, and only

[1] For the purposes of this project, as our focus is on supermarkets mainly, we will be considering the companies on this list with their dominant operational format being supermarkets, hypermarkets /supercenters/superstores, Cash & Carry/Warehouse Clubs and discount stores.

three African supermarkets were part of the top two hundred and fifty retailers in the world — Shoprite, Spar and Pick n Pay — and they are all based in South Africa (Deloitte, 2019). This highlights that the African supermarket industry needs to catch up with the rest of the world. However, the industry is making steady progress, with a growth rate of 9.8% in 2017, which was the second-highest amongst the other regions (Conway, 2016; Deloitte, 2019).

Since the largest supermarkets in the continent — Shoprite, Spar and Pick n Pay— are based in South Africa (Deloitte, 2019) and have branches in different African countries, they would be instrumental in leading changes in the industry across the continent. South Africa is an interesting case study, given that it is one of the only two countries on the continent where supermarkets account for fifty to sixty percent of all food sales (AT Kearney, 2015; PWC, 2016). The sales in the industry also keep growing, making $44.9 billion overall in 2017 (Bizcommunity, 2019; Ntloedibe, 2018). However, the rate of growth has slowed down in recent years, from about 7.5% in 2011 to about 2.5% currently (CEIC, 2020), due to worsening economic conditions. The industry would greatly benefit from improved demand forecasting, as some supermarkets currently use Enterprise Resource Planning (ERP) systems that are inadequate as they use mostly classical time series forecasting models that do not account for important extra variables that affect demand (Chapman, 2019; Du Toit, 2011; OptiProERP, 2019; Peksa, 2018; PWC, 2012; Willemain, 2019). Improved demand forecasting is also very important here as two million tonnes of food are wasted from retail in South Africa annually, and that loss is valued at about $800 million (Averda, 2019; CSIR, 2019; Oelofse & Nahman, 2012; WWF, 2017). This would be explained in more detail in Chapter 2.

1.3 Capstone Focus

This project is segmented into two parts: (1) an in-depth analysis of the South African supermarket industry, to understand its intricacies and make the case for improved demand forecasting and (2) formulating a case study to show how South African supermarkets can implement better demand forecasting methods, given its benefits.

Improved forecasting methods that include relevant demand influencing variables and other features like weather lead to more robust and accurate forecasts (Arunraj, Ahrens & Fernandes, 2016; Brownlee, 2016; Danziger, 2019; Louis, 2019 Singh, 2019). As mentioned above, more accurate forecasts can help to reap a $10 million benefit in a billion dollar business and has helped improve Marks and Spencer's business (Danziger, 2019), while forecast inaccuracies caused a Mexican company to lose about 7.73% of their sales (Treviño et al., 2014). that allowed for the inclusion of other demand predicting variables was beneficial in reducing costs for various companies. The ERPs currently used by the South African companies do not allow for the addition of the exogenous variables. I hypothesize that better forecasting models which allow for the implementation of extra variables would yield feasible and reasonable benefits for South African supermarkets, as such models–as evidenced in the aforementioned scholarly sources–have been shown to improve customer satisfaction and reduce waste, understock, and overstocking-related costs. My implementation in Chapter 3 aims to test this hypothesis.

In my implementation, I did not use data from a South African supermarket[2] which might initially appear to weaken this analysis. However, I used Walmart data, which is a reasonable substitute, given the shared industry context and that we would expect the data from South African supermarkets would have columns for dates, the stores items were bought from and how much of each item was bought (which we infer as demand), which is a very basic data schema. However, I used some information from the industry analysis in processing the dataset, and we can reasonably expect that data from South African supermarkets would also have annual or weekly seasonality —if the data is of daily granularity — which we also expect to see in the Walmart dataset, though the specific patterns would most likely differ. Additionally, South African consumers and US consumers have similar buying behaviour of supermarket products,

[2] **Note to reader:** I tried looking extensively for this data, but unfortunately, I could not find any I could use for the project.

as both markets prioritise health, wellness, safety and social impact a lot (Deloitte, 2016). Although particular elements from the South African market would not be included in this model, the implementation that would be detailed out in Chapter 3 would demonstrate that it would be worthwhile to think about implementing the best performing model from my analysis, as it is anticipated that the South African supermarkets would reap the benefits and reduced costs that can be gotten from implementing better forecasting methods.

I am carrying out a small - scale implementation that South African supermarkets can deploy at first before implementing on a larger scale. For example, if the supermarket cannot afford improved forecasts for all product lines, they can select the most profitable products or the products that are usually overstocked or understocked, so that the company reaps immediate tangible financial benefits from the implementation of the enhanced forecasts and gets a better idea of the value of the enhanced models. By starting with a smaller implementation, the supermarket can easily see the business value the model brings in a less risky format, as they would not have spent a lot of money implementing something across the entire business that is not yielding the desired results. From there, the company can make strategic decisions about possibly scaling up or outsourcing the forecasting.

2 Analysis of the South African Supermarket Industry

This chapter gives an in-depth analysis of the South African supermarket industry by giving a general overview of the industry, its approach to handling food waste, its customers, how the economic situation affects the industry and Porters' Five Forces analysis. The chapter wraps up with a discussion of what has been done with data science and IT in the industry, and makes the case for improved demand forecasting.

2.1 General Industry Overview

The South African supermarket industry is majorly dominated by the big supermarkets mentioned in the introduction - Shoprite, Spar, and Pick n Pay (Dludla, 2019; Dube & das Nair, 2016). These supermarkets have between seventy to ninety percent of the market share in the industry (Bekker, 2018; Dube & das Nair, 2016; Ntloedibe, 2018) and their popularity amongst low-to middle-income consumers in South Africa keeps growing (Euromonitor, 2019; Farfan, 2019). Supermarkets account for fifty to sixty percent of all food sales (AT Kearney, 2015; PWC, 2016) as mentioned above, compared to the eighty to ninety percent in other African countries except Kenya (AT Kearney, 2015; Grain, 2018; PWC, 2016).The South African supermarket industry made $44.9 billion overall in 2017 (Ntloedibe, 2018) and retail sales are still growing — albeit slowly — due to worsening economic conditions in the country (CEIC, 2020).

South Africa's supermarket industry is well developed, mature and has a solid infrastructure in place, with organized supermarket chains (Euromonitor, 2019; Farfan, 2019; Grain, 2018). The deep understanding of how the supermarket industry functions in other African countries —for example, knowing how to handle supply chain challenges like poor transport — because of similar contexts has enabled South African-based supermarket chains like Shoprite rapidly expand into other African countries (Ntloedibe, 2017; PWC, 2016). However, these expansions have not been all successful as Shoprite left Egypt, Zimbabwe, and Tanzania due to unfavorable market conditions (Grain, 2018). South African supermarkets have also not expanded into countries on other continents (Euromonitor, 2019; Farfan, 2019).

Towns and rural areas continue to be key growth areas for the industry in South Africa, with an estimated revenue of $660 million annually (Ntloedibe, 2017). Because of the growth prospects in those areas, larger supermarkets and shopping malls have tried to expand there with some success (Dube, 2017), but smaller stores like Food Lovers Market (Dube & das Nair, 2016) remain strong competitors in that space (Ntloedibe, 2017). This is because the smaller stores offer a long line of credit to their customers (Grain, 2018) and have lower prices — the average price of products from small vendors was seventy six percent lower than from supermarkets (Skinner & Haysom, 2016). However, these smaller stores suffer from inventory issues, financing, long account receivable days due to the long line of credit they give their customers, robbery, high transportation costs, and stiff competition, both from the larger retailers and other small stores (Ntloedibe, 2017).

2.2 Handling food waste in the industry

In South Africa, ten million tonnes of food —which is a third of the amount of food produced annually in the country —goes to waste annually, and this loss is valued at about $4 billion (Averda, 2019; CSIR, 2019; Oelofse & Nahman, 2012; WWF, 2017). Twenty percent of the food wasted comes from retail (CSIR, 2019). To reduce the food waste from this industry, some smaller supermarket chains like Engen are partnering with a tech company to tell customers when discounts on food close to its sell-by date are available through an app and the results so far are promising (BBC News, 2019). Going by the large supermarket chains' investor statements, there is no mention of using technology to combat the issue but they partner with charity organizations to give out food before it spoils as a way to tackle food waste (Pick n Pay, 2019; Shoprite 2019; Spar, 2019).

2.3 The South African Consumer

South African consumers are seeking greater value for their money and are prioritizing convenient and price friendly supermarkets that offer an amazing shopping experience for customers, to shop with (PWC, 2016). According to Nielsen (2016), South African consumers' choices of supermarkets are the most influenced by the store regularly having what they want to buy in stock. This highlights the high costs of understock in this industry. Given South Africa's economic condition — which I explain in greater detail later on in this chapter— it is understandable why consumers are currently greatly influenced by price (Nielsen, 2019). However, it is important to note that this loyalty motivated by price is not long-lasting, so supermarkets should still invest in improving customer experiences and relationships (Consulta, 2019). Doing this has paid off well for Spar as their customers' satisfaction levels keep increasing (Consulta, 2019). South African customers are also turning towards healthier, more organic and natural products, and the supermarket chains are continually expanding their product offerings to meet that demand (Ntloedibe, 2018; PWC, 2016). Additionally, South African supermarkets are employing technology to better connect and engage with customers (PWC, 2016).

Because of the high competition in the industry, most South African food retailers tried to expand their reach by diversifying their product offerings and turning to online shopping. This would increase the complexity of forecasting as there are more variables and products to take into account, and the forecasts have to account for the demand generated through online shopping. The supermarkets have to ensure there are enough products in stock to handle both online and offline demand, while striving to minimise the costs of overstocking and understocking. However, online shopping for groceries has not really taken off yet in the country (Ntloedibe, 2017; PWC, 2016) due to unreliable internet, logistics issues, a high rate of illiteracy, a lack of beneficial legal frameworks and a general distrust of online transactions (Johnson &

Iyamu, 2019; PWC, 2016). However, this channel has great potential (PWC, 2016) and is forecasted to reach about ten percent of retail sales in 2025 (Baudouin & Clémencin, n.d.).

2.4 How the South African economy affects the industry

The South African supermarket industry is highly dependent on how much customers spend and how much the economy grows (Bekker, 2018). The South African supermarket industry made $44.9 billion overall in 2017 (Ntloedibe, 2018) and retail sales are growing (Bizcommunity, 2019). However, the actual level of growth in industry is expected to be slow due to low consumer confidence — expectations about the country’s current and future economic conditions, high levels of debt, a fragile economy, industry maturity and limited retail expansion opportunities (Camera di Commercio Treviso, 2013; Consultancy.co.za, 2020; Euromonitor, 2019; Farfan, 2019; PWC, 2012, 2016). The contribution of the industry to the country’s GDP has reduced from 14.4% in 2012 (Camera di Commercio Treviso, 2013) to 9% in 2016 (Ntloedibe, 2017).

Downward sales are becoming more normal for most supermarkets as customer confidence is at its lowest level since 1982 (Mahlaka, 2017). There is also a high unemployment rate (Bekker, 2018; PWC, 2012) of 29.1% in 2019 (Trading Economics, 2019) and food inflation (Bekker, 2018), despite the country’s estimated GDP of about 358 billion dollars in 2019 (IMF, 2019). These tough economic situations reduce the buying power of South African customers, which would negatively affect the supermarket’s revenue and profits. However, it is predicted that sales and industry growth would improve due to improvement in economic conditions (Global Data, 2018).

Currency devaluation and hyperinflation in Zambia and Zimbabwe reduced Pick n Pay’s earnings from the Rest of Africa in 2019 (Buthelezi, 2020). This shows that economic conditions also affect returns in the supermarket industry in some other African countries. High rents (Consultancy.co.za, 2020; PWC, 2016) and power supply costs due to electricity problems (PWC, 2012) also make business hard to run in South Africa. For example, Shoprite spent about

$682,000 on diesel for their generators in 2014 (PWC, 2016). Access to electricity is very important in this industry because a good number of the products sold by supermarkets have to be kept in fridges to prevent spoilage.

In a bid to be more customer-centric because of the economic situation which made customers more price-conscious, Pick n Pay started a scheme in 2017 where customers could buy goods on credit without interest for 55 days with their shopper cards (Euromonitor, 2019; Farfan, 2019). Supermarkets are also focusing on penetrating the market in the towns and rural areas as mentioned earlier, and targeting customers with lower incomes (Dube, 2015) due to the economic situation in the country. Shoprite's strategy of having different brands to serve all segments of the population is paying off well for them, as there is strong growth opportunity in their Usave chain, which is focused on low-income consumers, compared to the market for Shoprite and Checkers that is more mature and focused on middle-income to high-income consumers (PMA Research, 2017; PWC, 2016). Pick n Pay also wants to further develop their Boxer brand that is aimed at low-income consumers (PWC, 2016).

Because of the industry maturity and other economic issues, there is less room for expansion for existing supermarkets and less space in the market for new entrants. Hence, the proposed best way forward for the supermarkets to grow is to keep expanding into other markets (Consultancy.co.za, 2020).

2.5 Porters' Five Forces

The Porters' Five Forces framework helps to analyse how competitive an industry is, by analysing the five forces listed below that shape every industry and give an insight into how profitable and attractive an industry is. This makes this framework relevant for this analysis.

2.5.1 Bargaining power of suppliers

The bargaining power of suppliers in the South African supermarket industry is low. South African supermarkets source the vast majority of their products locally (Dube, 2015).The large supermarkets maintain and control their own centralized procurement and distribution system, where they supply products to all their branches around the country. Shoprite is leading the way in this approach and Pick n Pay is also making moves to also further optimize their supply chain and distribution infrastructure (Dube, 2015; Ntloedibe, 2018; PWC, 2016). This centralized system helps the supermarkets save on transportation costs.

By using specialized procurement agents (Dube, 2015) and drawing on the economies of scale and scope they get from their size and power (Dube & das Nair, 2016), the larger supermarkets get influence over quantities delivered — they buy from large, established wholesalers that drive efficiency in the supply chain and the supermarkets get significant discounts on bulk orders — and product quality — they place strict demands for their suppliers to fulfill so as not to break any laws and regulations (Dube, 2015). The expected high discounts on orders and extra costs involved in meeting the supermarket's quality demands exclude smaller suppliers in the country from working with large supermarkets (Dube, 2015). This highlights the low bargaining power of suppliers in this industry. Collaboration among South African supermarkets — especially the large ones — is encouraged to further optimize supply chains and get the maximum value from them (Goko, 2015).

2.5.2 Bargaining power of customers

The bargaining power of customers in the South African supermarket industry is high. An industry with high bargaining power of customers is one where the switching costs of the buyers are low and the buyers can get similar products from other players in the industry (Corporate Finance Institute, 2017). Additionally, these industries would have a low profit potential (Corporate Finance Institute, 2017) and would have to ensure they sell a lot of products to get some profit due to price wars between players in the industry to get more customers — this is a defining feature of the supermarket industry (Bean-Mellinger, 2018; Marketplace, 2013; Reagan, 2013).

South African customers have more power to choose out of the many supermarkets and stores selling similar items (PWC, 2012). As highlighted earlier, South African consumers are currently highly influenced by price due to the suboptimal economic condition in the country. However, they are seeking greater value for their money and are prioritising convenience, customer experience and the stores always having what they want to buy in stock (Nielsen, 2016; Ntloedibe, 2018; PWC, 2016). As expected, the South African supermarkets are continuously expanding their product offerings and improving the customer experience in their supermarket chains in order to retain satisfied customers (Ntloedibe, 2018; PWC, 2016). This highlights the high buyer's strength in the industry due to low switching costs between supermarkets, the supermarkets selling similar products and having to set competitive prices to attract customers .

2.5.3 Threat of new entrants

The South African supermarket industry is dominated by few major supermarkets and so, the threat of new entrants is very low, especially given the high barriers to entry for smaller supermarket chains. For example, the major chains negotiated exclusive tenant clauses and leases in major shopping malls around the country where half of all grocery sales in the country are made (Dube & das Nair, 2016; Dludla, 2019). This prevented smaller supermarkets and other specialist stores from getting the space to also set up shop in the malls (Dube & das Nair, 2016; Dludla, 2019) and affected Walmart's move into the country as Massmart in 2014 (Dludla, 2019). However, these larger supermarkets have been given a governmental order to stop enforcing these clauses by September 2020 or regulations concerning this issue would be implemented (Dludla, 2019). Because of the maturity of the South African supermarket industry and other economic issues, there is less room for expansion for existing supermarkets and less space in the market for new entrants (Consultancy.co.za, 2020). Other challenges that new entrants face are access to financing and the high costs that would be needed for advertising and promotions, in order to compete with the incumbent large supermarkets (Dube & das Nair, 2016).

Some new entrants are also not able to thrive as much in the industry. For example, Choppies, a Botswana-based grocery retailer, exited the South African market just four years after entry due to operational issues (Khumalo, 2019). However, Food Lovers Market is a smaller independent South African-based supermarket that is thriving, because it is operating in a gap in the market —fresh fruits and vegetables— that the larger supermarkets did not strongly focus on (Dube & das Nair, 2016). Their flexibility enabled them to cut some procurement costs by getting supplies directly from farmers/producers, which allowed them to charge cheaper and competitive prices for their products (Dube & das Nair, 2016). With greater investments and the use of more technology, they would be able to expand more, get better logistics and better satisfy their customers.

2.5.4 Rivalry among existing firms

There is a fierce rivalry in the industry (Dube, 2015), particularly amongst the largest supermarkets that dominate the space. Supermarkets compete on prices and strive to get prime locations and best talent to curate better shopping experiences, in order to keep customers satisfied and make them keep coming back (PMA Research, 2017). Additionally, these supermarkets sell similar products, which heightens the rivalry.

Shoprite has the largest market share in the industry with a revenue of $10.34 billion in 2017. Spar has the second largest market share in the industry with a revenue of $6.232 billion in 2017, while Pick n Pay's $5.418 billion revenue in 2017 puts the firm in the third largest market share position (Euromonitor, 2019; Farfan, 2019; Mahlaka, 2017; Ntloedibe, 2017; PMA Research, 2017). Shoprite's success is attributed to its expansion to over fifteen African countries (Euromonitor, 2019; Farfan, 2019; Mahlaka, 2017; Ntloedibe, 2017; PMA Research,

2017) and its strategy of having distinct brands that cater to the different market segments in the country. For example, Checkers focuses on the high income customers, U-Save focuses on the low-income customers and Shoprite focuses more on charging low prices that play on customers' price sensitivities (Euromonitor, 2019; Farfan, 2019; Mahlaka, 2017; Ntloedibe, 2017; PMA Research, 2017).

2.5.5 Threat of substitutes

The threat of substitutes is not too apparent for this industry, because we can look at the substitutes from two different angles: informal channels as substitutes for supermarkets i.e kiosks and mom and pop shops, or look at restaurants and fast food outlets as substitutes for supermarkets, if we consider that most people want to go out to eat rather than buy groceries to cook at home (Hays, 2016).

From the first angle, as we highlighted in the earlier sections, this threat is generally low because of the established nature of the larger supermarkets in the country and most people buying food from the chains (AT Kearney, 2015; PWC, 2016). However, in the towns and rural areas, the threat is higher because of the cheaper prices and more generous lines of credit these mom and pop shops offer their customers (Grain, 2018; Ntloedibe, 2017; Skinner & Haysom, 2016). Although two billion rand per month is approximately spent on eating out in South Africa (BusinessTechSA, 2015), this amount is mostly driven by people with higher incomes. Thus, the threat of substitutes in that market segment is higher, especially if we consider convenience. However, if we consider that South African consumers are looking for healthier food (Nielsen, 2016; Ntloedibe, 2018; PWC, 2016), the threat is lower, depending on the restaurant offerings.

2.6 The use of data science and IT in the industry

2.6.1 What has been done?

Few South African retailers use demand forecasts for planning operations and for staffing (PWC, 2012). Shoprite uses data science for customer analytics and has also outsourced some data science tasks to dunnhumby (Businesswire, 2014). Additionally, they use ERP systems for their supply chain management and forecasts (Du Toit, 2011; Shoprite, n.d.). Pick n Pay also carries out forecasts using SAP, an automated ERP system (Pick n Pay, 2018; PWC, 2012; SAPinsider, 2010). Spar recognised that they lacked necessary quality data to properly do forecasting and optimise some business processes which is detrimental (Spar, 2018).

While Shoprite and Pick n Pay have taken good first steps by implementing ERP systems ERP systems do not use the most optimal forecasting algorithms (Chapman, 2019; Willemain, 2019). Most ERP systems use classical time series forecasting algorithms like ARIMA and exponential smoothing models (OptiProERP, 2019; Peksa, 2018) that have been outperformed by other models as we would explore in Chapter 3. Those that use linear regression do not allow for extra relevant features like weather to be added to the models (OptiProERP, 2019; SAP Help Portal, 2020), which are extremely important for getting better forecasts (Danziger, 2019) as we would discuss in more detail in Chapter 3.

2.6.2 Why is improving demand forecasts crucial in this market?

As seen throughout the industry analysis, keeping costs low — especially given the economic situation in South Africa — and satisfying customers is an important task for supermarkets. Because South African customers are price sensitive and highly prioritise stores always having what they want to buy in stock (Nielsen, 2016, 2019), better demand forecasting is needed to reduce the aforementioned high costs of overstock and understock in the industry (PWC, 2012). Having excess inventory leads to more costs for the supermarket, especially if they have to get rid of them without getting the expected profits. Because of the high customer buying power, the availability of other suppliers with the same products and other factors discussed earlier, the cost of understock in a supermarket is extremely detrimental as the company would lose revenue from the lost sales and would most likely lose the potential future sales they would get from the customer to a rival supermarket. If a supermarket is constantly understocked, they would also have to pay more in transportation costs to get supplies overall as they would have to order more frequently to be able to cater to customers' demand.

Because of the economic situation in South Africa, creating useful forecasts has been difficult leading to overstock, understock and wastage in the supermarkets (PWC, 2012). According to Paul Dickson in Goko (2015), bad forecasting in the industry reduces the South African customers' experience because understock would mean they would not be able to get products from their preferred stores and would be more inconvenienced having to go to another supermarket to shop. The improved forecasting methods being proposed would take into account extra relevant variables that would be relevant in properly estimating future demand (Arunraj, Ahrens & Fernandes, 2016; Brownlee, 2016; Louis, 2019; Singh, 2019). In the South African industry, the economic situation is very important to take into account when making forecasts, given its great impact on the industry, so variables like Consumer Price Index (CPI) can be added to the model to make the forecasts more accurate and useful. Knowing that there are tougher times economically would lead to reduced purchasing power and putting that information into the model would help to create more realistic forecasts. In that situation, we would reasonably expect that the demand for essential and cheaper products would increase and there would be a reduction in the demand for luxury goods. So, having this kind of information included in the forecasts would help the supermarket properly plan for the kinds of products to have in stock in a situation like that in order to satisfy customers, and reduce overstock and understock costs. Additionally, holidays and promotions (Nielsen, 2019) are important variables to consider as they greatly affect demand and sales in supermarkets. These forecasts would be extremely beneficial for the supermarkets due to improved customer satisfaction and reduced costs related to overstock and understock.

To summarise and set the stage for the next chapter, the ERP forecasting system being used by South African supermarkets is not optimal. Hence, the proposed improved forecasting method being implemented in Chapter 3 is anticipated to yield more accurate forecasts if implemented in South African supermarkets, given the inclusion of extra relevant features.

3 Time Series Forecasting - Implementation

As highlighted in Chapter 1, despite knowing the many benefits of having more accurate forecasts, South African supermarkets should start with small-scale implementations because it would be better to start with something they can easily implement to see the business value it brings, before making strategic decisions about possibly scaling up or outsourcing, especially when we consider costs.

Data from a South African supermarket would not be used for this implementation. However, daily data about sales of 111 items across 45 Walmart stores from 2012 to 2014 (Walmart + Kaggle, 2014) was used. This dataset is a justifiable substitute as it has a time series component, it is large enough to run my analysis and relevant to the supermarket context. Additionally, it would serve as a representative dataset for the kind of data we expect South African supermarkets to generally have at the bare minimum in general, since the dataset has columns containing how many units of each item were sold in each store, on each day of the year, from 2012 to 2014.

3.1 Exploratory Data Analysis

The **sales** dataset contains daily data of the sale of 111 items across 45 Walmart stores from 2012 to 2014.[3]

[3] The dataset does not include Christmas Day due to missing data.

3.1.1 Data Dictionary

Column	Description	Data Type
date	Dates from 2012-2014	Object
store_nbr	Store number; there are 45 Walmart stores in this dataset	Integer
item_nbr	Item number; there are 111 items in this dataset	Integer
units	Number of units of each item, sold on each day of the year from 2012-2014 in each of the 45 stores. We are inferring that the sales means demand for the product(s).	Integer

Figure 1. Table describing the columns in the **sales** dataset

Most of the variables in the dataset are already in integer form, so they would be appropriate to be used in the algorithms later on. Since we are working on a time series forecasting problem, we would need to convert the **date** column from an object to a DateTime format, so it would be appropriate to be used in the algorithms.

3.1.2 Data Pre-processing and Summary Statistics

```
count    4.612716e+06
mean     9.879201e-01
std      9.880973e+00
min      0.000000e+00
25%      0.000000e+00
50%      0.000000e+00
75%      0.000000e+00
max      5.568000e+03
Name: units, dtype: float64
```

Figure 2. Summary statistics of the numeric variables in the dataset.

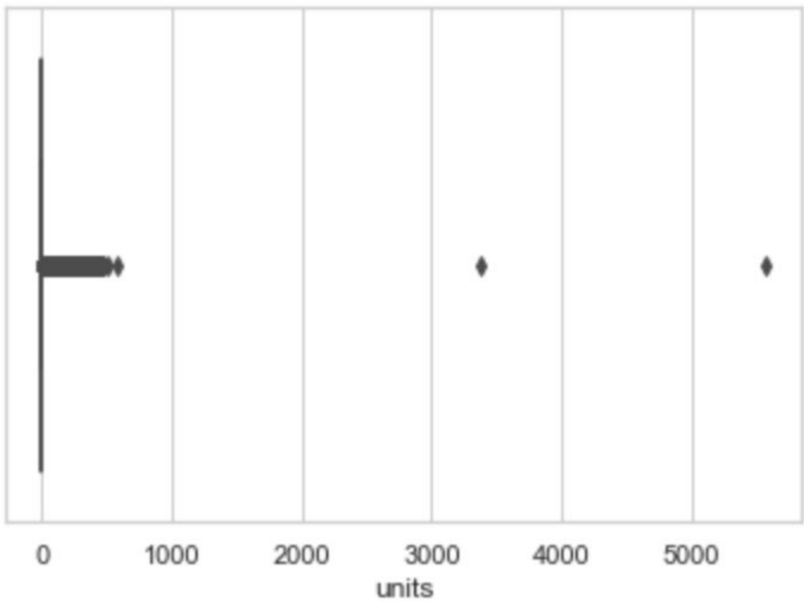

Figure 3. Boxplot showing the distribution of the **units** variable.

From Figure 2, we can see the count, mean, standard deviation, range, and quartiles of the **units** variable, which is the only numeric variable in the dataset. These are descriptive or summary statistics that help to understand the data better and possibly spot any concerning issues. Figure 3 shows the distribution of the **units** variable using a boxplot, which would help us better see the outliers and the huge skew in in the dataset, compared to a histogram, since the number of zeros in the variable were a lot.

Additionally, we can see that there are quite a lot of days in the dataset where items were not sold across the stores, looking at the 25th, 50th and 75th percentiles in Figure 2. This shows that looking at the quartiles would be a better statistic to use to understand the dataset and the distribution of the units instead of a statistic like the mean. The mean tells us that on average, one of each product was sold in each store on every day in the dataset, which would not be an accurate description. Additionally, because the data is heavily skewed as we can infer from the percentiles in Figure 2 and the plot in Figure 3, the mean would not be an accurate statistic to use because the outliers would greatly affect the mean (Laerd Statistics, 2018) ,which we can confirm happened, looking at Figure 2. The median (or 50th percentile) would be a more robust measure of central tendency in this scenario instead (Laerd Statistics, 2018).

While having a lot of days in the dataset where items were not sold across the stores is concerning, it is important to remember only a subset of stores and items sold in those stores are being considered in the dataset. However, it would highlight a big problem if the stores were selling items that no one was buying, as the supermarket would be spending money to keep unprofitable items in stock. Hence, I decided to carry out an analysis to see the best selling and worst selling products, in terms of the amount of the product that was sold during the time period.

	Item_Number	Total number sold
0	21	0.0
1	109	0.0
2	110	0.0
3	101	15.0
4	33	53.0
5	15	178509.0
6	43	450132.0
7	4	659793.0
8	8	690600.0
9	44	777142.0

Figure 4. Table showing the best selling and worst selling items.

As we can see in Figure 4, items **21**, **109** and **110** did not generate any sales in the entire time period. Going by this analysis, these products should be discontinued in the supermarkets. We also see that item **44** is the best selling item. Given that we are advocating for South African supermarkets to start with small scale implementations with the most profitable product(s) in their business, I used that rationale and decided to focus on forecasting the demand for item **44** in

2014. It is important to highlight that this same analysis can be carried out for all the other products in the dataset by just changing the item number in consideration.

My training set is data before the 1st of January 2014, while my test set is data after that date. Because only three years were provided, I decided to make 2012 and 2013 the training set, so that the models had relatively enough data to learn the trends in the dataset to be able to forecast what the demand in 2014 would be. For example, just one year of data would not be enough to see an annual seasonality, if it is present in the data. The test set is generally expected to follow the trend and seasonality identified in the train set, however, there might be some outliers due to randomisation or other external factors that would have caused a rise or decline in the product demand.

To simplify the approach, I formatted the dataset to represent the total number of **Item 44** that were sold across all the 45 stores on each day of the years in consideration. In South Africa, the large supermarkets maintain and control their own centralized procurement and distribution system, where they supply products to all their branches around the country (Dube, 2015; Ntloedibe, 2018; PWC, 2016). This approach influenced my choice in modifying the dataset to be in this format, as I wanted to include some aspects from the industry analysis in this implementation.

date	store_nbr	item_nbr	units
2012-01-01	1	44	0
2012-01-01	2	44	215
2012-01-01	3	44	0
2012-01-01	4	44	0
2012-01-01	5	44	0

→

	date	units
0	2012-01-01	912
1	2012-01-02	655
2	2012-01-03	778
3	2012-01-04	651
4	2012-01-05	988

Figure 5. Image showing how the data structure was modified for the forecasting models.

3.1.3 Time Series Decomposition

Time series decomposition gives a structured framework to think about how to approach a time series forecasting problem (Brownlee, 2017). After decomposing the observed time series, looking at the trend and seasonality of the time series in consideration would inform the most appropriate forecasting models for the problem.

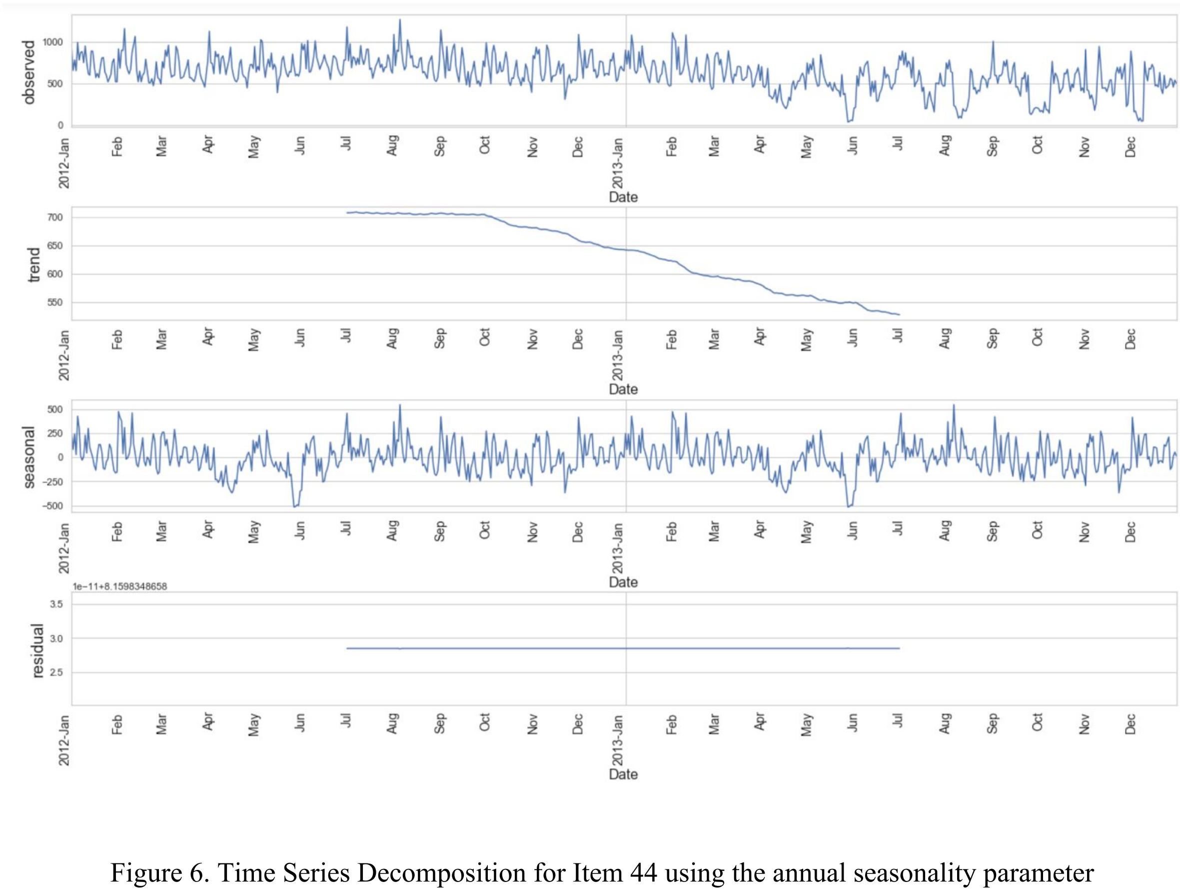

Figure 6. Time Series Decomposition for Item 44 using the annual seasonality parameter

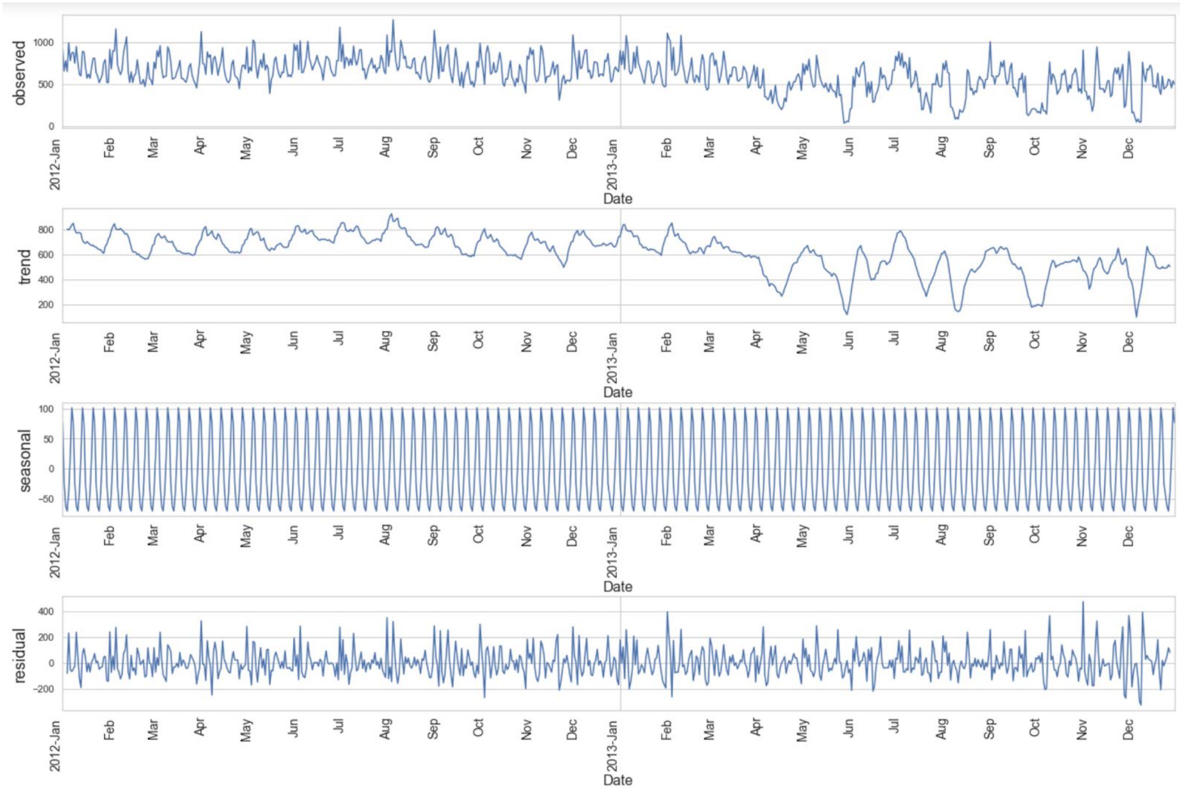

Figure 7. Time Series Decomposition for Item 44 with a weekly seasonality parameter

The seasonal_decompose function (Statsmodels, 2019) carries out a time series decomposition of the data as seen in Figures 6 and 7 above, and it gives a visual representation of four parts of the time series. The **observed** panel shows the actual data. The **trend** panel shows the general tendency for the data to linearly or nonlinearly increase, decrease or stay constant over time. The **seasonal** panel shows if there are some patterns in the time series that repeat themselves over a specified time period. The function has a parameter that allows the user to input the desired frequency or seasonality that they feel the data would have, while looking at the observed data. The choice of this seasonality parameter is dependent on the granularity of the data; since the data in consideration here is daily, we can assume that the data would have both annual and weekly seasonalities, because it would make sense to assume that most people would have consistent weekly shopping habits — for example, the most sales of luxury and non-essential products would be made on weekends if we consider people have more free time and would want to indulge a bit to relax, while the sales of essential goods might be spread more evenly across the week. So, we put that into the function and see whether those assumptions are correct or not. Figure 6 shows a time series decomposition under the assumption that the data has annual seasonality i.e. the demand pattern repeats each year, while Figure 7 shows a time series decomposition under the assumption that the data has weekly seasonality i.e. the demand pattern repeats each week. The **residual** panel shows the remaining part of the time series that was not captured by the trend or seasonality, possibly due to randomisation.

Looking at the panels in Figures 6 and 7, there is a decline in 2013 demand, and this can be due to economic conditions, price wars with other supermarkets, other products coming into the market that would cannibalise the sales of Item 44 or a switch to other competing products if this product does not reflect the current state of customer's preferences in the market e.g. if this product is less healthy and there is a general switch towards healthier products in the market, there would be reduced demand for the product in coming years. These are situations that are likely to occur in South African supermarkets, which adds support to the suitability of this dataset for this implementation.

Evaluating the time series decomposition

The **trend** panel in Figure 6 shows that the time series has no trend at first and then a decreasing trend over time. We can also see a clear annual seasonality as the pattern of sales in one year is similar to the pattern of sales in the following year. On the other hand, the weekly seasonality decomposition shows some seasonality and the trend does not show a clear direction in the increase or decrease of sales, whether linear or exponential. Looking at the residuals —which show the remaining data that was not captured by the trend and seasonality measures— in Figures 6 and 7, the annual seasonality decomposition does a better job at helping us understand the time series better as it has little to no residuals, compared to the weekly seasonality decomposition which has larger residuals. This shows that incorporating annual seasonality assumptions in the models should be prioritised.

3.1.4 Exploratory Data Analysis to hypothesize what Item 44 could be

Looking at the total sales for each day of the week and the total sales in each month could give insights that could help in hypothesising what Item 44 could possibly be, by looking at what days and months the product was sold most in, in total in the training set.

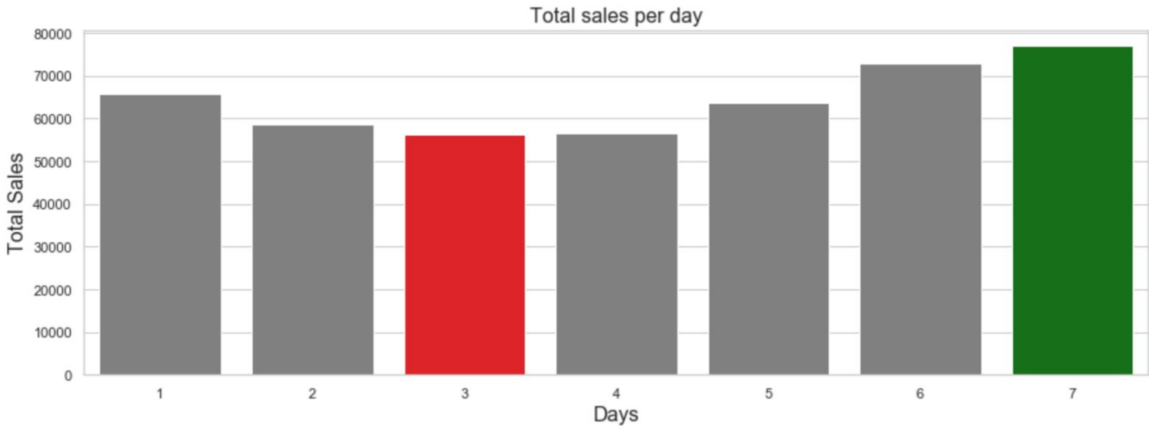

Figure 8. Bar chart showing the total sales made on each day from 2012-2013.

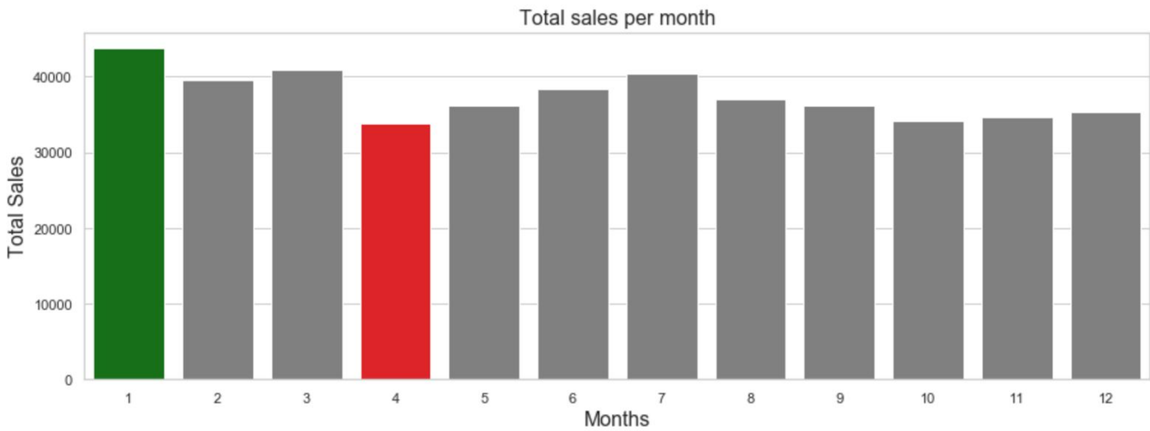

Figure 9. Bar chart showing the total sales made in each month from 2012-2013.[4]

[4] For Figures 8 and 9, the green bar shows the day/month with the highest total sales while the red bar shows the day/month with the highest total sales.

From Figures 8 and 9, we can see that Item 44 sells the most on the weekends and sells the most in January, followed by March and July. My hypothesis is that Item 44 might be junk food, for example, crisps. I can reasonably assume that they would be sold more on weekends if we consider that people would buy them for ‘cheat days’ or as snacks when they have gatherings with friends during the weekend. Throughout the year, they are sold in good amounts, which makes sense because people would always snack. They would be sold more in January if we consider that people might be buying them as comfort food because of the cold weather and possibly stocking up before the Super Bowl in February. I assume people would buy relatively less junk food in April in preparation for the summer, where everyone tries to look their best physically, and then during the summer months, they would ease up on that and focus on having fun, eating and relaxing with friends, especially around the 4th of July holiday.

3.2 Forecasting Model Selection

In the existing scholarly literature, different classical or traditional forecasting models have been used to forecast demand in the supermarket or food retail industry. For example, the *AutoRegressive Integrated Moving Average* (ARIMA) model was used to predict the daily demand for onions and potatoes in an Indian wholesale vegetable market (Shukla & Jharkharia, 2013) and the demand for a group of perishable dairy products, alongside the *Holt-Winters Triple Exponential Smoothing* model (Da Veiga et al., 2014). The *Seasonal AutoRegressive Integrated Moving Average* (SARIMA) model was used to forecast the wholesale monthly prices for tomato in Turkey (Adanacioglu & Yercan, 2012) and a hybrid SARIMA model with a neural network to forecast demand in a Chilean supermarket (Aburto & Weber; 2003, 2007). An *Auto-Regressive Moving Average with eXogenous regressors* (ARMAX) model was used to forecast daily demand for beer in the Slovenian market (Bratina & Faganel, 2008). A *Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors* (SARIMAX) model was used to forecast the daily sales of bananas in a discount retail store in Germany (Arunraj, Ahrens & Fernandes, 2016).

I implemented the following two models, with the aim of comparing the baseline model which is one of the models South African supermarkets usually use for their forecasting (Holt-Winters Triple Exponential Smoothing) with a machine learning based forecasting model that allows for the inclusion of relevant variables (Prophet). The choices of parameters for each model is discussed in the code notebook.

1. Holt-Winters Triple Exponential Smoothing

This is the baseline model that would serve as a basis to compare the performance of the proposed model to be implemented. The Holt-Winters Triple Exponential Smoothing model can be used for datasets with both trend and seasonality (Brownlee, 2018a; Hyndman & Athanasopoulos, 2018), which we can see our model has from Figure 6. This is better than an ARIMA model, which would have to be changed to a SARIMA model before it can be used for data with seasonality. This is why I chose this model as the classical model to implement, out of the other classical models that are found in the South African supermarkets' ERP systems.

In the Holt-Winters model, a future prediction is a weighted sum of past observations, but more weight is given to the more recent observations and exponentially smaller weights are applied as the observations go further back in time (Brownlee, 2018a; Hyndman & Athanasopoulos, 2018). In terms of parameters, the Python package being implemented automatically optimises the optimal parameter settings based on the dataset, but they would still be outlined here. The alpha is the learning parameter that determines whether the model pays more attention to the more recent past observations (if the alpha is large) or the ones further back in time (if the alpha is small) (Brownlee, 2018a). The beta parameter is to control how much the trend affects the future predictions while the gamma controls how much the seasonality affects the future predictions (Brownlee, 2018a). The major cons of this model are that it does not allow for the addition of relevant extra variables or features to improve the model forecasts and it cannot handle data with dual seasonality, except we use its sister model, TBATS (De Livera et al., 2011; Hyndman, 2012; Hyndman & Athanasopoulos, 2018).

2. Prophet

Prophet is an open-source forecasting software released by Facebook as a solution to automatic forecasting techniques being quite hard to tune, or too inflexible to include additional relevant assumptions, and the lack of analysts with in-depth knowledge about time series methods but have deep domain knowledge in organizations (Taylor and Letham, 2017). Hence, Prophet is beneficial for the business setting because it can be used in different contexts, has easily adjustable and interpretable model parameters for users with relevant domain knowledge about the time series, and is flexible enough to allow for the inclusion of relevant exogenous variables to enable better forecasts, as highlighted before (Taylor and Letham, 2017). This reason, in addition to others I would discuss later in this section, is why I chose to implement this model in this project.

To give more detail about the model itself, Prophet forecasts time series data using a decomposable time series model (Harvey & Peters, 1990) using trend, seasonality (yearly, weekly, and daily) and holiday effects. The model also handles outliers and missing data well. This model is conceptually similar to a Generalized Additive Model (GAM), where a class of regression models is combined, with non-linear smoothing techniques applied to the regressors (Taylor and Letham, 2017). The use of a similar GAM formulation in the model makes fitting the model and the inclusion of other relevant exogenous variables as extra regressors to the model simple to do (Taylor and Letham, 2017). This is very advantageous as it makes the forecasts more reliable and accurate, and forecasts are produced in a timely fashion, which would be useful in a business setting where speed and accuracy are of the essence. Prophet models are fitted in Stan (a Bayesian inference programming language) and the model frames the time series forecasting problem as a curve-fitting exercise, which differs from other classical models (Taylor and Letham, 2017).

Prophet has outperformed different classical models like exponential smoothing, TBATS - a classical time series model useful for modeling daily data that has both annual and weekly seasonality (De Livera et al., 2011; Hyndman, 2012; Hyndman & Athanasopoulos, 2018), ARIMA and SARIMA models in scenarios such as bitcoin forecasting, forecasting the daily streamflow in a river in the US and forecasting for the air pollution levels of Bhubaneswar City (Taylor and Letham, 2017; Tyralis & Papacharalampous, 2018; Yenidogan et al., 2018; Acharaya et al., 2019). The sources serve as support to my hypothesis that Prophet would produce the most accurate forecast in this implementation.

Another plus to using this model for my implementation is that compared to ARIMA models, Prophet can model all kinds of seasonalities within the model, without needing to add them in the exogenous variables and has a faster fitting time. Additionally, Prophet adds another layer of originality to this project as there is no scholarly paper that highlights how Prophet can be used for forecasting in the food retail or supermarket industry, based on my research.

3.3 Feature Engineering

Feature engineering is an essential portion of modelling because extra features can be added to the model that can help to explain the variation in the data. These added variables are relevant in helping to understand the observed time series better and produce the most accurate and robust forecasts possible (Arunraj, Ahrens & Fernandes, 2016; Brownlee, 2016; Louis, 2019; Singh, 2019).

These exogenous variables can be information about promotions, weather (Louis, 2019), holidays, pay days, a location's economic situation, outlier events like strikes and pandemics (Binh Pham, 2016; Insight Works, 2016; Lokad, 2019; Reade, 2015). These are variables that can influence demand by causing spikes or sharp declines in sales at different times of the year and so, it would be extremely relevant to capture the effects of those variables while making the forecast (Arunraj, Ahrens & Fernandes, 2016).

Exogenous Variables Included

1. Weather - Average daily temperature (in Fahrenheit) and average amount of daily rainfall (in inches) recorded at the stores that were used in the data, across the two years in the dataset.
2. US holidays - this was encoded as a binary feature, indicating whether that day was a holiday or not.
3. Feature engineering was used to create some relevant features to add to the model.
   a. **months in the year**, **day of week, day of the month-** date-time features that were one-hot encoded so they could be properly used in the models.
   b. **is_weekend** - this was encoded as a binary feature, indicating whether that day was a weekend or not.
   c. There is usually a trend of the current observations being correlated with observations at previous time steps in time series, and so we can add them as features to the model to improve our forecasts (Brownlee, 2016; Singh, 2019). We can use autocorrelation plots to determine the appropriate lags to use to create features (Singh, 2019). Looking at Figure 10 below, we see that there are many significant correlations up until the first 30 days in our dataset. Significant correlations are those greater than the confidence intervals highlighted in blue. Because removing a month's work of data —due to the lags and NA values created as a result— would not be significant, I chose to use a lag of 30 days to create features for each entry in the dataset. So, for example, for the 31st of January, **lag_1** would be the demand on the 30th of January, **lag_2** would be the demand on the 29th of January and so on.
   d. Additionally, for each day, I added a feature each for the **minimum**, **maximum** and **average** value of the demand from the thirty previous days because of the lag chosen. For example, for the 8th of January, I would take the minimum, maximum and average of the demand from the 1st to 30th of January of that year.

To prevent data leakage here because pd.rolling includes the value in consideration in the window, I added a .shift ( ) function.

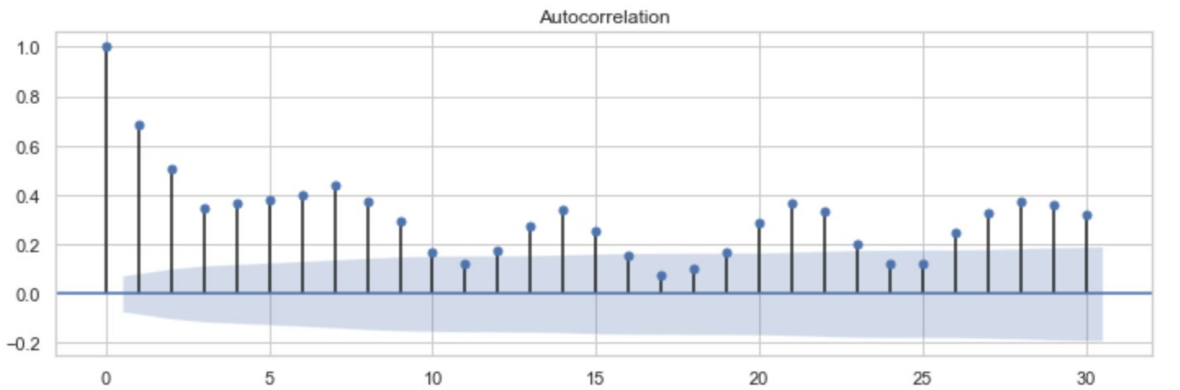

Figure 10. Autocorrelation plot for the training set that helps to decide appropriate lags

3.4 Evaluation Metrics

To evaluate the performance of forecasting models, metrics like Mean Average Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Squared Error and Mean Average Error (MAE) are usually used in the scholarly literature (Hyndman & Koehler,2005; Hyndman & Athanasopoulos, 2018). However, for this project, I chose to use Cost of Forecast Error (CFE) and Mean Average Error (MAE) as my evaluation metrics as they are the most appropriate metrics to use in this context.

3.4.1 Why other metrics would not be suitable

The MSE and RMSE tend to penalize and magnify errors and is dependent on the scale of the data being considered (Hyndman & Koehler, 2005; Hyndman & Athanasopoulos, 2018).

Compared to the MAE, they are more sensitive to outliers (Armstrong, 2001) and also harder to interpret.

The MAPE is scale-independent and easy to understand, as it tells us the average percentage error in our forecasts when compared to the actual observation (Hyndman & Koehler,2005; Hyndman & Athanasopoulos, 2018). However, MAPE puts a heavier penalty on negative errors than on positive errors and if any of the values in the data is zero, then the MAPE would be undefined (Hyndman & Koehler, 2005; Hyndman & Athanasopoulos, 2018). In our dataset, as seen in Figure 2, and in the supermarket industry in general, there could be days where there would be no demand for certain products, which makes the MAPE unsuitable to use as an evaluation metric in this context. The symmetric MAPE (sMAPE) was proposed as a solution to the issues MAPE has that was highlighted above, but sMAPE can have negative values, which makes it unsuitable (Hyndman & Koehler, 2005; Hyndman & Athanasopoulos, 2018).

If the range in the dataset was larger or we were comparing datasets of different scales, especially since some of the values were zero, the Mean Average Scaled Error (MASE) would have been the most appropriate metric to use in this context. With MASE, we divide the MAE of our model's forecast by the training set MAE from the baseline model (Hyndman & Koehler,2005; Hyndman & Athanasopoulos, 2018). In this way, we can directly compare model results to each other as a MASE value greater than one means the model is worse than the naive/ baseline model and should not be considered further, and vice-versa (Hyndman & Koehler,2005; Hyndman & Athanasopoulos, 2018). I chose not to include the MASE in my evaluation because using the CFE and MAE, we can easily see which models performed better than the other(s), and if the models performed better than the baseline model.

It is important that relevant and easily interpretable metrics are used to evaluate the forecasting models, especially because of the business context in consideration.

3.4.2 Why Cost of Forecast Error is a relevant metric

The main metric I would use to evaluate the performance of the models is Cost of Forecast Error (CFE). This metric gives an estimated dollar value of how much forecast inaccuracy costs the supermarket, whether through overstock or understock costs. This metric is especially useful for the business audience who would like to see in dollar values what the improved forecasts would add to their business, as cash or added costs is a more tangible metric to evaluate the quality of forecasts.

To get this Cost of Forecast Error, I assume each product costs **$3.50**, which is a sensible estimate for the price of a bag of chips (Walmart, 2020), which is a product that we can assume Item 44 could possibly be. My second assumption is that the cost of overstock would be the sales the supermarket would have made from the goods (had they not spoiled) plus an extra cost of disposing the goods or giving them to charity. Hence, the cost of overstock is **$5**. My third assumption is that the cost of understock would be the sales the supermarket would have made from the sales of the goods, had they been in stock plus the cost of losing a customer due to the unavailability of the goods. The cost of losing a customer is estimated as the future amount we would have gotten from them if they had continued coming to the store and buying the product. I estimated this to be them buying the product 3 more times in the time period in consideration, and the price of the good is slightly discounted to be $3, because money in the future is worth less in the present. Hence, the cost of understock would be **$12.5**.

I could have also included inflation rates in my estimates, but I wanted to simplify the assumptions as much as possible. Understock is worse for the supermarkets due to the high competition and availability of suppliers that sell the same products in the industry, as we discovered in Chapter 2, and the magnitude in my estimates reflect that. However, a flaw that one might argue is that this metric would be biased in favour of a model that would try to overstock rather than understock if it cannot predict the right amount of stock.

3.4.3 Why Mean Average Error is a relevant metric

The MAE is scale dependent, easy to understand and compute, less sensitive to outliers compared to other scale dependent metrics like RMSE and MSE and is useful for both a technical and business audience (Hyndman & Koehler, 2005; Hyndman & Athanasopoulos, 2018). Though the MAE is scale dependent, it is still okay to use it in this case because most of the training data is in the same scale in the hundreds and only a few in the thousands, as seen in Figure 6 earlier, and we are using the same training and test data for all the models.

With the MAE, we directly have an idea of how many units our forecasts are off by on average, by taking the average of the differences between the predictions and the actual values for each day in the test set. This can inform the supermarket on how much safety stock is needed to keep on hand in the store, in order to ensure we adequately meet customers demands as much as possible especially when demand is very variable. However, we need to ensure this number is as low as possible, especially for perishable goods, so that we do not overstock on them as we would lose money if they spoil before being sold. So, deciding on what to stock, the supermarket can look at shorter term forecasts and plan accordingly, so that the daily, weekly and perhaps, quarterly forecasts are as accurate as possible.

The best performing model would have the lowest CFE and MAE, as this would show the model produces the most accurate forecasts and causes the firm to lose the least money due to overstock and/or understock, compared to the others.

3.5 Results

Results

Model Name	CFE	MAE
Holt-Winters'	$ 315,945	177
Prophet	$ 251,885	89

Figure 11. Table showing the results from the models, looking at the CFE and MAE metrics

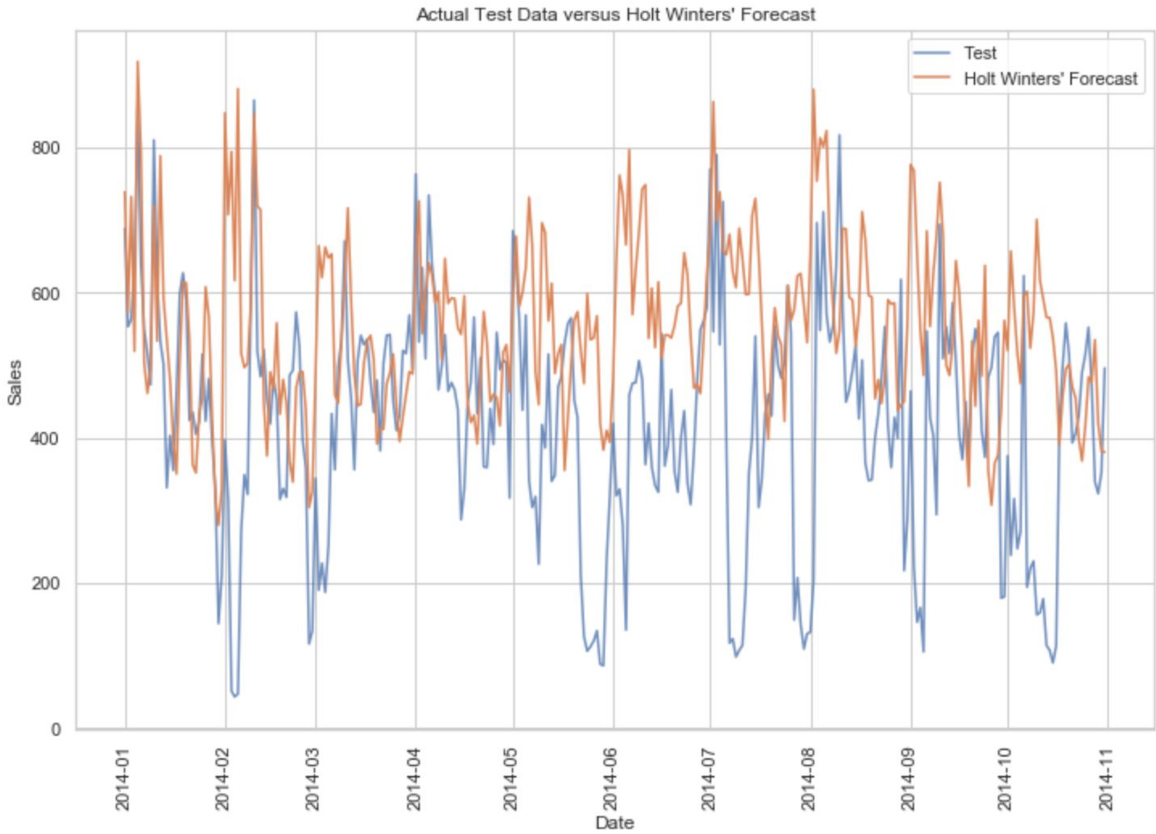

Figure 12. Plot showing Holt-Winters predictions against the actual test data

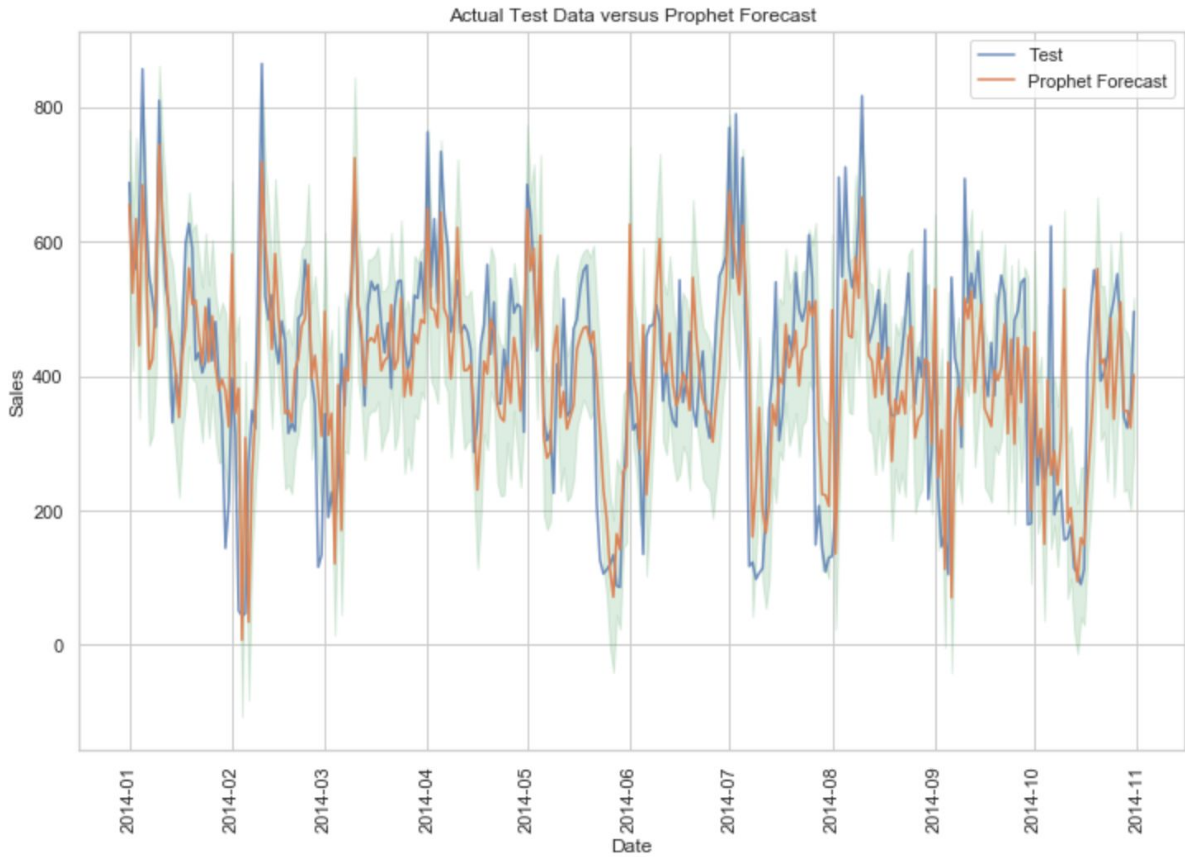

Figure 13. Plot showing Prophet's predictions against the actual test data

Interpretation of results

Before this section begins, it is important to note that it is very unlikely for demand forecasts to be completely accurate (Chapman, 2019), as some of the data movements might be random and we might not have used all the relevant predictor variables in our model.

Looking at Figure 11, 12 and 13, we can see that as hypothesized, Prophet greatly outperformed the Holt-Winters model on the two metrics of CFE and MAE, which goes to show how adding relevant features can greatly improve our model accuracy. If the Holt-Winters model was better, the supermarket could have saved **$ 64,060** more, due to less costs relating to overstock and/or understock. To rephrase, the implementation of the Prophet model to forecast the demand of this product would have saved the supermarket **$ 64,060.** The Mean Average Error would also have been reduced by half, which is quite significant as well.

These results make sense as we would expect that the use of the extra features would improve Prophet's accuracy. The results are also not perfect, which is a good sign, because if they were perfect, it would have signified some data leakage from the data processing section. The biggest gap between Holt-Winters forecast and the actual test data is on the 5th of February, which shows that there is some information in the extra variables that would have helped Holt-Winters properly predict that day's forecast, for example, temperature, rain or the day being a weekend. That day is predicted quite well by the Prophet forecast, although it might signify that that day is very special or an outlier, given that the confidence interval drops below zero there. We can interpret that to mean returned produce or a recommendation for the supermarket not to have stock on that day due to a drastic drop in demand. The biggest gap between Prophet forecast and the actual test data is on the 6th of September, which is another day forecasted terribly well by the Holt-Winters model. This signifies that there are some days that demand changes drastically from what has happened in the past which is likely to happen in the real world. The models are unable to account for this drastic change due to lack of extra information, or due to the drastic drop being a chance event.

Feature Importance

	component	beta_coefficients
36	day_of_week_6	0.81
39	is_weekend	0.81
0	day_of_month_1	0.60
1	day_of_month_10	0.42
26	day_of_month_5	0.32
...	...	...
28	day_of_month_7	-0.18
25	day_of_month_4	-0.22
27	day_of_month_6	-0.26
33	day_of_week_3	-0.31
31	day_of_week_1	-0.83

Figure 14. Table showing the importance of the features used in Prophet's predictions

In a regression, beta coefficients show the strength of each independent variable in influencing the changes in the dependent variables (Statistics How To, 2016). As Prophet uses a class of regression models, beta coefficients are appropriate to be used in calculating how Prophet determined the importance of each feature in demand. From Figure 14, we can see that the increase in demand was always higher during the weekend or on Saturday. While our exploratory data analysis showed Sunday as the day with the highest sales, Prophet's feature importance is not too far off by highlighting the weekend and Saturday as the most important variables that are related with an increase in demand. Hence, this is expected. On the flip side, Prophet determined that for lower demand , it was most likely to occur on Monday or Wednesday, which is reasonable, given that our exploratory data analysis showed Wednesday as the day with the lowest sales in Section 3.1.4. The most surprising variables showing up are the days of the month, especially as we did not consider them in the exploratory data analysis. It is also surprising that this product's demand is shown not to be strongly influenced by holidays or weather in the Prophet model.

Evaluating Residuals

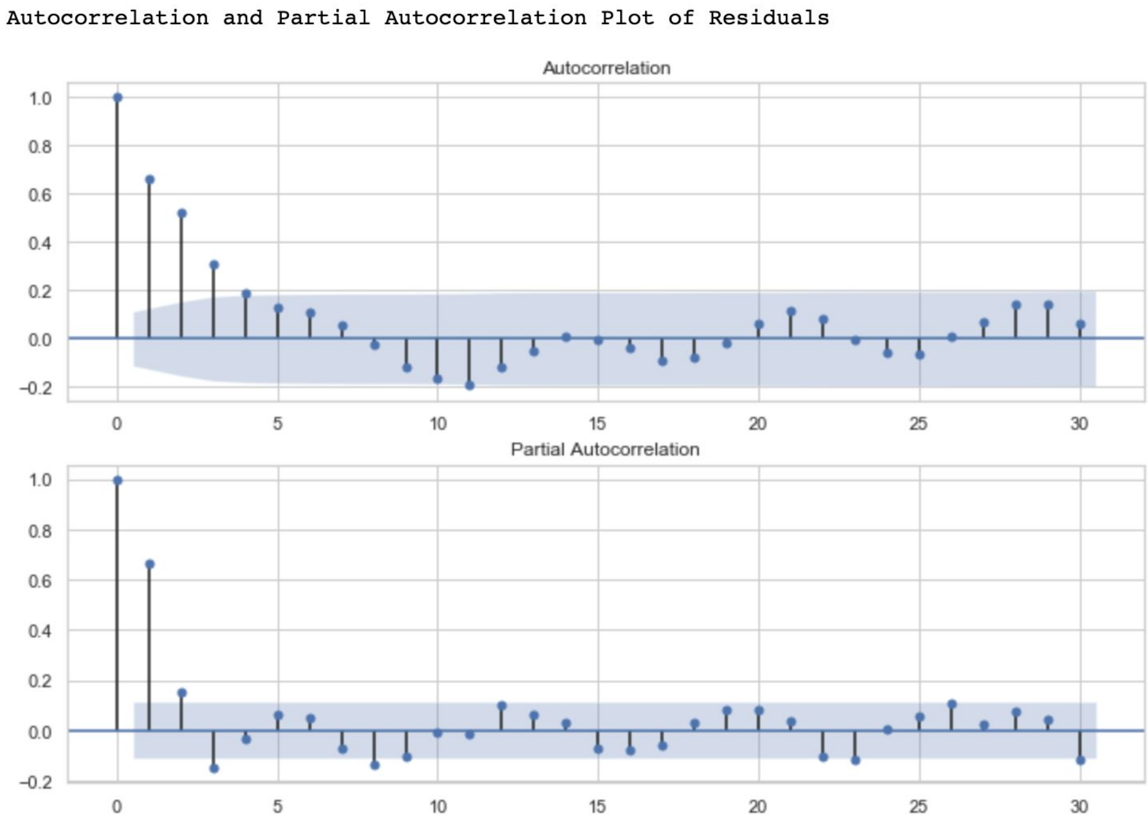

Figure 15. Graph showing the autocorrelation and partial autocorrelation plot of Holt-Winters residuals

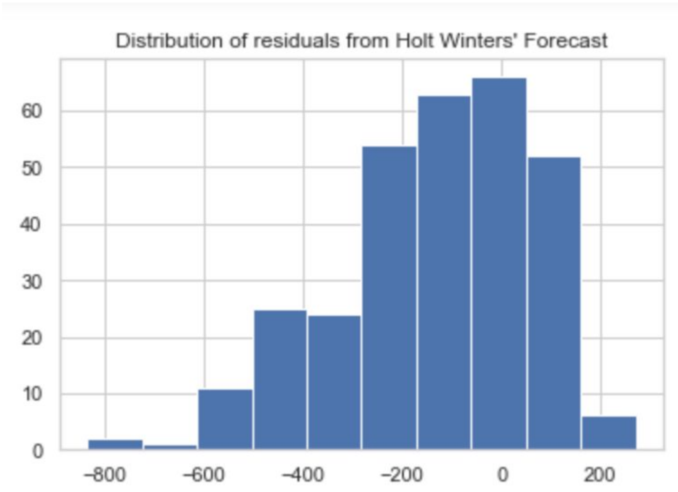

Figure 16. Graph showing the distribution of Holt-Winters residuals

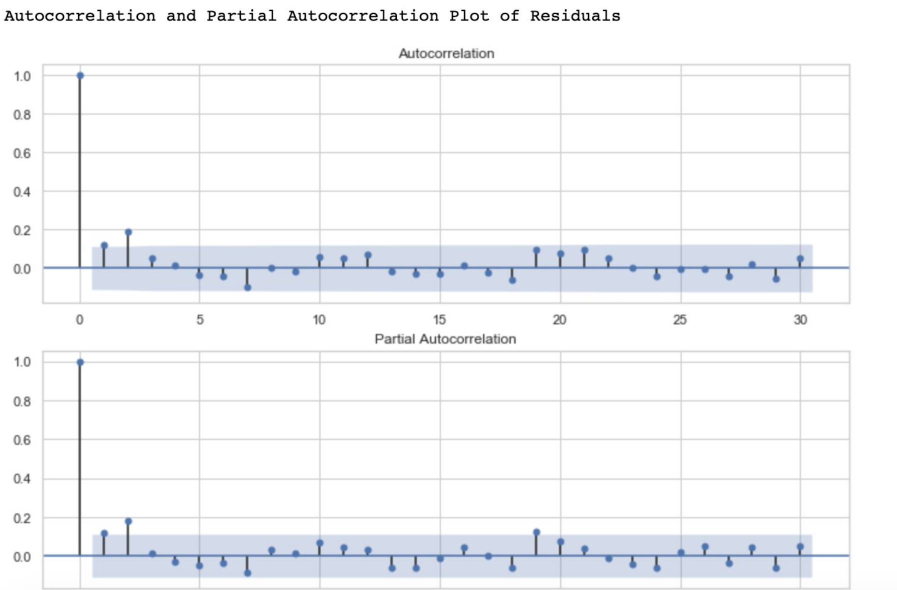

Figure 17. Graph showing the autocorrelation and partial autocorrelation plot of Prophet's residuals

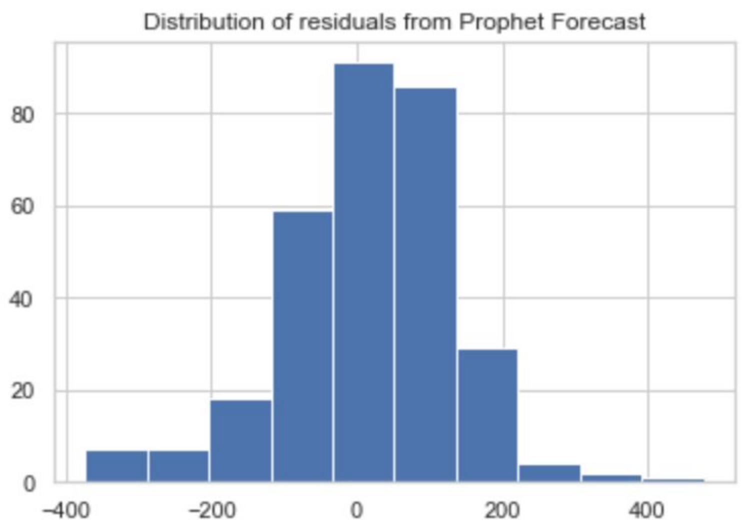

Figure 18. Graph showing the distribution of Prophet's residuals

Evaluating the residuals of models is another way to critically analyse if the models are performing optimally. The residuals are the difference between the actual data and the forecasts. According to Hyndman & Athanasopoulos, (2018), a good forecasting method would yield uncorrelated residuals (which we can see with the autocorrelation plots) because it would show that there is some information left in the residuals that should have been used in making the forecasts, and they should have a zero mean to show the models are not biased. It would also be good for them to be normally distributed.

Looking at Figures 15 and 17, we can see that the Holt-Winters model yielded uncorrelated residuals, compared to the Prophet model, showing that there was some uncaptured information left by the Holt Winters model that should have been used for the forecast. Prophet's use of exogenous variables helped it capture that information. Looking at Figures 16 and 18, we can see that the Holt-Winters model is biased as its distribution of residuals does not have a zero mean and is skewed, while Prophet's residuals, on the other hand, have a zero mean and are normally distributed, showing a lack of bias in the model.

4 Future Work & Conclusion

4.1 Future Work

In this implementation, we did not have information about other possibly relevant variables that would affect the demand of items and be useful in understanding the trend of demand such as the economic situation (which can be inferred using metrics like CPI) or promotions. If we had this information, it would be valuable in greatly improving the accuracy of the forecast (Arunraj, Ahrens & Fernandes, 2016; Brownlee, 2016; Louis, 2019; Singh, 2019), and would be extremely relevant for the South African case. This is because their supermarket industry is greatly influenced by economic conditions and they use a lot of promotions as well. So, to extend this project, this analysis would be re-run including other relevant variables, for other products and on South Africa supermarket demand data to see whether the insights remain the same. Qualitative forecasting to include non-quantitative but relevant insights would also be great additions when making decisions using these forecasts, which are meant to serve as a guide. However, the results are expected to be similar in terms of the best performing model.

Secondly, I would create an optimisation model for prescriptive analytics after demand forecasting is carried out. Knowing the forecasted demand for the product is one thing, but the company has to then make decisions on what to do next, given this information. This is especially crucial if the demand for the product is very variable for example, or if qualitative forecasting shows an increasing or decreasing trend in the sales of the product in the near future. This optimisation model would help the supermarket decide the optimal amount of product to stock, the optimal safety stock to have and how much of the product to order at a particular time and the best supplier to order from, considering costs and lead time. Additionally, it would help in creating proper merchandising, pricing and revenue management strategies to boost revenue and customer satisfaction.

I would also implement more complex models like neural networks and long short-term memory networks, if I had more data to see if they can perform better as they can understand deeply complex relationships in the data although they run the risk of overfitting (Pesch and Senge, 2019). Having said that, South African supermarkets should not attempt to deploy these kinds of models until they become more analytically mature because it would be harder to understand more technical models at their current stage, except they outsource it. This reason, and the lack of enough data, is why I chose not to implement them in my Capstone. Additionally, neural networks might not be the best models to use for time series forecasting forecasting as they are not as accurate (possibly due to overfitting in the training set as highlighted earlier) and require a lot of time to properly pre-process, compared to other statistical time series models or regression-based machine learning algorithms (Brownlee, 2018b; Smyl, 2020).

Finally, I would conduct a similar industry analysis for supermarkets in the US as they had four out of the top 5 retailers in the world as mentioned in Chapter 1. This analysis would help to understand the peculiarities of this industry in the US that make it quite successful, and how it is similar to and differs from the South African supermarket industry. By this analysis, further recommendations on how South African supermarkets can improve can be provided, taking into account the difference in contexts.

4.2 Conclusion

The main hypothesis that framed this paper and Capstone was that better forecasting models which allow for the implementation of extra variables would yield feasible and reasonable benefits for South African supermarkets, as such models–as evidenced in the scholarly sources highlighted throughout the paper–have been shown to improve customer satisfaction and reduce waste, understock, and overstocking-related costs.

Chapter 1 made the case for why improved demand forecasting was necessary for supermarkets, because of saved costs, increased customer satisfaction, better marketing and HR planning. The chapter also talked about cases where money had been saved through better forecasting and where money had been lost due to forecast inaccuracies. Additionally, the chapter highlighted why South Africa would be a relevant and interesting case study, given that the supermarkets were the largest on the continent, the industry was growing ─ albeit hampered due to worsening economic conditions─and hence, would be instrumental in driving changes in the industry across the continent. The South African supermarkets also used Enterprise Resource Planning (ERP) systems for demand forecasting, which were sub - optimal, and as such, were prime candidates for the use of improved demand forecasting systems. Though specific data was not used in this analysis, having the basic data schema and similar buying behaviour between US and South African consumers made the Walmart proxy data a good substitute.

Chapter 2 went through an in-depth analysis of the industry, highlighting some strategies supermarkets took to succeed despite the worsening conditions in the industry and the competitive landscape of the industry, which served to make the case for the need to implement improved forecasting. Chapter 3 went through a deep dive into the implementation and analysis of the results of the two models - Prophet and Holt-Winters, where we saw Prophet outperform Holt-Winters as hypothesized, due to the presence of exogenous variables.

Though South African supermarket data was not used, the project has shown why it would be clearly beneficial for those firms to implement the improved forecasting models, given the results of the implementation and the industry analysis. It is anticipated that the supermarkets would get significant cost savings, reduce waste and improve their customer satisfaction ratings through the implementation of these improved forecasting models, similar to the results we got here. This was just a small implementation and subset of data used, and so it would be great for the supermarkets to try out these enhanced models and see what they get from it. There will also not be a single best model for all products, so it is important to train for the different products and see what works best (Pesch & Senge, 2019).

5 References

Acharaya et al. (2019). Time Series based Air Pollution Forecasting using SARIMA and Prophet Model | Proceedings of the 2019 International Conference on Information Technology and Computer Communications. Acm.Org. https://dl.acm.org/doi/10.1145/3355402.3355417

Averda. (2019, February 19). Fighting food waste in households - Averda South Africa. Averda. https://averda.co.za/news/fighting-food-waste-households/

Arunraj, N., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. International Journal of Operations Research and Information Systems. 7. 1-21. 10.4018/IJORIS.2016040101.

BBC News. (2019, August 29). Cutting food waste in South Africa. BBC News; BBC News. https://www.bbc.com/news/av/business-49515666/cutting-food-waste-in-south-africa

Bean-Mellinger, B. (2018). What Is the Profit Margin for a Supermarket? Chron.Com. https://smallbusiness.chron.com/profit-margin-supermarket-22467.html

Bekker, D. (2018). Wholesale and Retail of Food :: South Africa. Whoownswhom.Co.Za. https://www.whoownswhom.co.za/store/info/4594?segment=Food+%26+Beverage

Binh Pham. (2016, May 16). Production demand forecasting: How to do it right? TRG Blog; TRG International. https://blog.trginternational.com/bid/192827/production-demand-forecasting-how-to-do-it-right

BizCommunity. (2019, May 7). South African retail supply shows steady growth. Bizcommunity.Com; BizCommunity. https://www.bizcommunity.com/Article/196/168/192863.html

Brownlee, J. (2016, December 13). Basic Feature Engineering With Time Series Data in Python. Machine Learning Mastery. https://machinelearningmastery.com/basic-feature-engineering-time-series-data-python/

Brownlee, J. (2017, January 29). How to Decompose Time Series Data into Trend and Seasonality. Machine Learning Mastery. https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/

Brownlee, J. (2018a, August 19). A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python. Machine Learning Mastery. https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/

Brownlee, J. (2018b, October 30). Comparing Classical and Machine Learning Algorithms for Time Series Forecasting. Machine Learning Mastery. https://machinelearningmastery.com/findings-comparing-classical-and-machine-learning-methods-for-time-series-forecasting/

BusinessTechSA. (2015). How much South Africans spend on takeaways. Businesstech.Co.Za. https://businesstech.co.za/news/business/182033/how-much-south-africans-spend-on-takeaways/

Businesswire. (2014, February 3). dunnhumby Chosen by Shoprite as Customer Science Partner. Businesswire.Com. https://www.businesswire.com/news/home/20140203005166/en/dunnhumby-Chosen-Shoprite-Customer-Science-Partner

Camera di Commercio Treviso. (2013). OVERVIEW OF THE SOUTH AFRICAN RETAIL MARKET. https://tb.camcom.gov.it/uploads/CCIAA/Corsi/Atti/2013_11_07/OverviewOFTHESOUTHAfrica.pdf

CEIC. (2020, March 18). South Africa Retail Sales Growth. CEICdata.com. https://www.ceicdata.com/en/indicator/south-africa/retail-sales-growth

Chapman, M. (2019, December 20). 8 Inventory Demand Forecasting Techniques You Must Try in 2020. EazyStock. https://www.eazystock.com/uk/blog-uk/8-best-inventory-demand-forecasting-techniques/

Checkers South Africa. (2019). *Checkers | Buy Groceries Online with the Sixty60 App* [Video]. Retrieved from https://www.youtube.com/watch?v=VWySakronKQ

Consulta. (2019, March 6). SA’s top 5 supermarket brands battle it out for the hearts and minds of South African consumers – SA-csi 2018. Consulta Blog; Consulta Blog. https://blog.consulta.co.za/sas-top-5-supermarket-brands-battle-it-out-for-the-hearts-and-minds-of-south-african-consumers-sa-csi-2018/

Consultancy.co.za. (2020, January 22). South Africa’s retail sector is lagging behind the continent. Consultancy.Co.Za; Consultancy.co.za. https://www.consultancy.co.za/news/1989/south-africas-retail-sector-is-lagging-behind-the-continent

Conway, J. (2016). Grocery retail: worldwide sales value by region 2016-2021 | Statista. Statista; Statista. https://www.statista.com/statistics/754878/grocery-retail-worldwide-sales/

Corporate Finance Institute. (2017). Bargaining Power of Buyers - How Buyers Exert Negotiating Power. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/knowledge/strategy/bargaining-power-of-buyers/

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.

Danziger, P. N. (2019, October 8). What Do You Call A Holiday Retail Forecast That Ignores The Weather? Wrong. Forbes. https://www.forbes.com/sites/pamdanziger/2019/10/08/whats-a-holiday-retail-forecast-called-the-ignores-the-weather-wrong/

Deloitte. (2019). Global Powers of Retailing 2019. Retrieved 12 December 2019, from https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Consumer-Business/cons-global-powers-retailing-2019.pdf

Dube, S., & das Nair, R. (2016, February 19). BARRIERS TO ENTRY IN THE SOUTH AFRICAN SUPERMARKET INDUSTRY. CCRED - A Leading University Research Group. https://www.competition.org.za/review/2016/2/19/barriers-to-entry-in-the-south-african-supermarket-industry

Euromonitor. (2019). Retailing in South Africa | Market Research Report | Euromonitor. Euromonitor.Com. https://www.euromonitor.com/retailing-in-south-africa/report

facebook. (2017, June 29). Stationary and non-stationary timeseries · Issue #234 · facebook/prophet. GitHub. https://github.com/facebook/prophet/issues/234

facebook. (2017, November 16). Measuring feature importance for additional regressors · Issue #361 · facebook/prophet. GitHub. https://github.com/facebook/prophet/issues/361

facebook. (2018, June 21). Regressor Influence · Issue #585 · facebook/prophet. GitHub. https://github.com/facebook/prophet/issues/585

facebook. (2019, February 14). Getting regression coefficients for additional regressors · Issue #848 · facebook/prophet. GitHub. https://github.com/facebook/prophet/issues/848

Farfan, B. (2017). South Africa's Largest Retailers from Steinhoff to Woolworths. The Balance Small Business; The Balance Small Business. https://www.thebalancesmb.com/largest-african-retailers-2892272

Food and Grocery Retail Market Analysis Report by Type, By Region And Segment Forecasts, 2011 - 2020. (2018). Reportbuyer.Com. https://www.reportbuyer.com/product/5491690/food-and-grocery-retail-market-analysis-report-by-type-by-region-and-segment-forecasts-2011-2020.html

Global Data. (2018). Retailing in South Africa, Market Shares, Summary and Forecasts to 2022. GlobalData Report Store. https://store.globaldata.com/report/gdrt0304mr--retailing-in-south-africa-market-shares-summary-and-forecasts-to-2022/

Goko, C. (2015). Retailers need to look to 'collaboration on supply-chain issues.' Supermarket.Co.Za. https://www.supermarket.co.za/news-article.asp?ID=5450&CatTags=14-Supply%20chain

Grain. (2018). SUPERMARKETS OUT OF AFRICA! Food systems across the continent are doing just fine without them. https://www.grain.org/media/W1siZiIsIjIwMTgvMTEvMDgvMDdfMDBfMThfMjMxX1N1cGVybWFya2V0c19vdXRfb2ZfQWZyaWNhX0VOXzAzLnBkZiJdXQ

Grand View Research. (2017). Food & Grocery Retail Market Size, Share | Industry Report, 2011 - 2020. Grandviewresearch.Com. https://www.grandviewresearch.com/industry-analysis/food-grocery-retail-market

Hays, G. (2016, January 15). Paper or Plastic: The Hidden Competitors Threatening Grocery Retailers. Business.Com; business.com. https://www.business.com/articles/grocery-retailers-hidden-competitors/

Hyndman, R., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. Otexts.Com. https://otexts.com/fpp2/

Hyndman, R. J. (2013). Forecasting with daily data. Robjhyndman.Com. https://robjhyndman.com/hyndsight/dailydata/

Hyndman, R. J., & Koehler, A. B. (2005). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

IBM Institute for Business Value. (n.d.). Just add weather. Retrieved March 1, 2020, from https://www.ibm.com/downloads/cas/MBYJELNN

IMF. (2019). Report for Selected Countries and Subjects. Imf.Org. https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/weorept.aspx?pr.x=29&pr.y=10&sy=2017&ey=2021&scsm=1&ssd=1&sort=country&ds=.&br=1&c=199&s=NGDPD%2CPPPGDP%2CNGDPDPC%2CPPPPC%2CPCPIPCH&grp=0&a=

Johnson, O., & Iyamu, T. (2019). Framework for the adoption of e-commerce: A case of South African retail grocery sector. The Electronic Journal of Information Systems in Developing Countries, 85(5). https://doi.org/10.1002/isd2.12095

Khumalo. (2019, August 14). Choppies to exit SA market, chair to step down. Fin24. https://www.fin24.com/Companies/Retail/choppies-to-exit-sa-market-20190814

Laerd Statistics. (2018). Mean, Mode and Median - Measures of Central Tendency - When to use with Different Types of Variable and Skewed Distributions. Laerd.Com. https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php

Lokad. (2019). Forecasting technology FAQ - Inventory Optimization Software. Lokad.Com. https://www.lokad.com/forecasting-technology-faq

Louis, N. (2019). Retailers Turn to Latest Oracle Demand Forecasting Service to Optimize Inventory. Globalbrandsmagazine.Com. https://www.globalbrandsmagazine.com/retailers-turn-to-latest-oracle-demand-forecasting-service-to-optimize-inventory/

Mahlaka. (2017). SA’s retail nightmare. Supermarket.Co.Za. https://www.supermarket.co.za/news-article.asp?ID=7020&CatTags=1

Marketplace. (2013, September 12). Groceries: a low margin business, but still highly desirable - Marketplace. Marketplace. https://www.marketplace.org/2013/09/12/groceries-low-margin-business-still-highly-desirable/

Nicasio, F. (2019). How to Optimize Your Inventory: A Guide to Getting Stock Levels “Just Right” - Vend Retail Blog. Retrieved 12 December 2019, from https://www.vendhq.com/blog/optimize-your-inventory-get-it-just-right/

Nielsen. (2016, February 11). SA’s new retail reality means “one size fits all” strategies no longer cut it - Nielsen. Bizcommunity.Com; BizCommunity. https://www.bizcommunity.com/Article/196/168/153163.html

Nielsen. (2019, October 16). South Africa is the second-most price sensitive country in the world. Nielsen.Com. https://www.nielsen.com/za/en/insights/article/2019/south-africa-is-the-second-most-price-sensitive-country-in-the-world/

Ntloedibe, M. (2017). South Africa Retail Food Industry. https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Retail%20Foods_Pretoria_South%20Africa%20-%20Republic%20of_11-7-2017.pdf

Ntloedibe, M. (2018). The South African Retail Foods Industry. https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Retail%20Foods_Pretoria_South%20Africa%20-%20Republic%20of_6-27-2018.pdf

Oelofse, S., & Nahman, A. (2012, August 9). Estimating the magnitude of food waste generated in South Africa. ResearchGate; ResearchGate. https://www.researchgate.net/publication/230645334_Estimating_the_magnitude_of_food_waste_generated_in_South_Africa

OptiProERP. (2019, June 4). Intelligent Forecasting, Manufacturing Forecasting Software | OptiProERP. ERP for Manufacturers | Manufacturing Software | OptiProERP. https://www.optiproerp.com/blog/forecasting-made-easy-for-manufacturers-with-intelligent-erp-software/

Peksa, J. (2018, December 14). Extensible Portfolio of Forecasting Methods for ERP Systems: Integration Approach. ResearchGate; Information Technology and Management Science. https://www.researchgate.net/publication/329657548_Extensible_Portfolio_of_Forecasting_Methods_for_ERP_Systems_Integration_Approach

Pesch, R., & Senge, R. (2019). From Whiteboard to Production System Demand Forecasting System for an Online Grocery Retailer. https://cdn.oreillystatic.com/en/assets/1/event/300/From%20whiteboard%20to%20production_%20A%20demand%20forecasting%20system%20for%20an%20online%20grocery%20shop%20Presentation.pdf

Pick N Pay. (2019). 2019 INTEGRATED ANNUAL REPORT. Retrieved 12 December 2019, from http://www.picknpayinvestor.co.za/downloads/investor-centre/annual-report/2019/iar-2019.pdf

PMA Research. (2017). Competition Increases in South African Retail Environment. Pma.Com. https://www.pma.com/content/articles/2017/03/competition-increases-in-south-african-retail-environment/

Prophet. (n.d.). Prophet. https://facebook.github.io/prophet/

PWC. (2012). South African retail and consumer products outlook - 2012-2016. https://www.pwc.co.za/en/assets/pdf/retail-and-consumer-products-outlook-2012-2016.pdf

PWC. (2016). So much in store:Prospects in the retail and consumer goods sector in ten sub-Saharan countries. https://www.pwc.co.za/en/assets/pdf/retail-in-africa.pdf

Reagan, C. (2013, June 6). What's Behind the Rush Into the Low-Margin Grocery Business. CNBC; CNBC. https://www.cnbc.com/id/100794988

Reuters Editorial. (2019, November 25). UPDATE 1-South Africa's big supermarkets told to drop exclusivity clauses in mall leases. U.S.; Reuters. https://www.reuters.com/article/safrica-retail/update-1-south-africas-big-supermarkets-told-to-drop-exclusivity-clauses-in-mall-leases-idUSL8N28540Z

Ryan, T. (2015). Retailers suffer the high cost of overstocks and out-of-stocks – RetailWire. Retailwire.Com. https://retailwire.com/discussion/retailers-suffer-the-high-cost-of-overstocks-and-out-of-stocks/

SAP Help Portal. (2020). Sap.Com. https://help.sap.com/viewer/2cfbc5cf2bc14f028cfbe2a2bba60a50/2.0.04/en-US/d5f360d9dbed412a82778c8d5c500ac1.html

SAPinsider. (2010). Pick n Pay Benefits from a New Retailing Solution. Wispubs.Com. https://sapinsider.wispubs.com/Assets/Case-Studies/2010/July/Pick-N-Pay-Benefits-From-A-New-Retailing-Solution

Schuman, E. (2015, June 15). What is behind retail's $1.1 trillion inventory losses? Computerworld. https://www.computerworld.com/article/2934353/what-is-behind-retails-1-1-trillion-inventory-losses.html

Shoprite 2018 Financial Statements. (2019). Retrieved 12 December 2019, from https://www.shopriteholdings.co.za/content/dam/MediaPortal/documents/shoprite-holdings/FinancialStatements/6043_Shoprite_AFS_2018E.pdf

Shoprite Holdings | Supply Chain Management. (2019). Shopriteholdings.Co.Za. https://www.shopriteholdings.co.za/trade-partners/supply-chain-management.html

Shoprite SA. (2019). Retrieved 12 December 2019, from https://play.google.com/store/apps/details?id=za.co.shoprite.shopriteeezicoupon&hl=en_GB

Singh, A. (2019, December 9). 6 Powerful Feature Engineering Techniques For Time Series Data. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/12/6-powerful-feature-engineering-techniques-time-series/

Smyl, S. (2020). When Neural Networks are not the Best Tool for Time Series Forecasting. Gallery.Video. http://cfany.gallery.video/cfasouthafrica/detail/videos/curated/video/5996911947001/when-neural-networks-are-not-the-best-tool-for-time-series-forecasting?autoStart=true

Spar International Annual Review. (2019). Retrieved 12 December 2019, from https://spar-international.com/wp-content/uploads/2019/05/SPAR_International_Annual_Report_2018.pdf

Standardized Beta Coefficient: Definition & Example - Statistics How To. (2016, February 16). Statistics How To. https://www.statisticshowto.datasciencecentral.com/standardized-beta-coefficient/

Statsmodels. (2019). statsmodels.tsa.seasonal.seasonal_decompose — statsmodels. Statsmodels.Org. https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html

Taylor, S. J., & Letham, B. (2017). Forecasting at scale. https://doi.org/10.7287/peerj.preprints.3190v2

The changing drivers that influence consumer behaviour | Consumer Business | Deloitte Southern Africa. (2016, October 14). Deloitte. https://www2.deloitte.com/za/en/footerlinks/pressreleasespage/the-changing-drivers-that-influence-consumer-behavior.html

The Future of Demand Forecasting. (2019). Retrieved 12 December 2019, from https://progressivegrocer.com/future-demand-forecasting

Trading Economics. (2019). South Africa Unemployment Rate | 2000-2019 Data | 2020-2022 Forecast | Calendar. Tradingeconomics.Com. https://tradingeconomics.com/south-africa/unemployment-rate

Treviño, A., Sergio, M., Contreras, A., & Valdez, S. (2014). LOSS OF SALES DUE TO INCORRECT FORECASTING OF DEMAND. http://www.web.facpya.uanl.mx/vinculategica/Revistas/940-964%20LOSS%20OF%20SALES%20DUE%20TO%20INCORRECT%20FORECASTING%20OF%20DEMAND.pdf

Tyralis, H., & Papacharalampous, G. A. (2018). Large-scale assessment of Prophet for multi-step ahead forecasting of monthly streamflow. Advances in Geosciences, 45, 147–153. https://doi.org/10.5194/adgeo-45-147-2018

Walmart. (2020). Chips on Sale! Walmart.Com. https://www.walmart.com/browse/food/chips-sale/976759_976787_1001390_6288637_8326100

Walmart + Kaggle.(2014). Walmart Recruiting II: Sales in Stormy Weather. Retrieved 12 December 2019, from https://www.kaggle.com/c/walmart-recruiting-sales-in-stormy-weather

Willemain, T. (2019, March 13). Consider this if considering ERP for demand planning and forecasting. Smart Software. https://smartcorp.com/blog/erpforecasting/

WNS DecisionPoint. (2020). The IoT Revolution in Retail. Wns.Co.Za. https://www.wns.co.za/insights/articles/articledetail/367/the-iot-revolution-in-retail

WWF. (2017). The truth about our food waste problem. Wwf.Org.Za. https://www.wwf.org.za/?21962/The-truth-about-our-food-waste-problem

Yenidogan, I., Cayir, A., Kozan, O., Dag, T., & Arslan, C. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. 2018 3rd International Conference on Computer Science and Engineering (UBMK). https://doi.org/10.1109/ubmk.2018.8566476