



HELLENIC OPEN UNIVERSITY
SCHOOL OF SOCIAL SCIENCE

Supply Chain Management (SCM)

DISSERTATION

Analyzing sales patterns in a supermarket chain during the pandemic

STUDENT: AGGELOS VERRAS

SUPERVISOR A: DIAMANTIDIS ALEKSANDROS

Contents

LIST OF TABLES.....	3
LIST OF FIGURES.....	3
EXECUTIVE SUMMARY	
ACKNOWLEDGMENTS	
INTRODUCTION	
CHAPTER 1 LITERATURE SURVEY	
1.1 OVERVIEW OF SUPERMARKET CHAIN	
CHAPTER 2 THE SECTOR OF SHOPPING IN AS UPER MARKET CHAIN.....	
2.1 INTRODUCTION OF BAZAAR S.A	
2.2 PRODUCTS OF REVERENCE	
2.3 STRATEGIC ROLE	
CHAPTER 3 METHODOLOGY	
3.1 BASIC TYPE SOF RESEARCH METHODOLOGY	
3.1.1 Research Tools	
3.1.2 Descriptive Statistics	
3.1.3 Moving Average	
3.1.4 Regression Analysis	
3.1.5 Mean Absolute Percent Error	
3.1.6 Mean Absolute Error	
CHAPTER 4 EMPIRICAL ANALYSIS	

EXECUTIVE SUMMARY

This thesis aims to investigate the methods developed to verify preferences and expenditure patterns of shopping habits of people in a supermarket chain. Its intention is to describe the shopping pattern of people and examine the purchasing power and preference according to the social, economic and seasonal shifts. In other words, its aims to distinguish the type of products that people purchase during the year, to assist the supermarket chain to direct the supplies of certain products in a larger amount than usual.

This thesis was based on a quantitative analysis of data set which was provided by the Bazaar S.A. and component end-of monthly sold quantities and rollover requirement of the last 4 years (2019-2022). The survey performed as this Thesis was addresses in a number of 150 retail stores at the territory of Greece and in 9 different categories. For the control of the cases, were used descriptive statistics but mainly regression and EWMA model. The analysis of the results was executed operating the Excel-2022 application and its individual functions.

Using the latest technology time series analysis techniques and statistical models, we outlined features of the examined data (trends, cyclicities, seasonality, etc.). Research data is drawn from existing bibliographies, articles and websites emphasizing the importance of an effective forecasting tool.

ACKNOWLEDGMENTS

For the writing of this essay, I have to thank those who actually helped me and influenced me in their way to its accomplishment.

Particularly, I would like to thank my supervisor Mr. Diamantidis Aleksandros for the work of my supervision from the point of my theoretical search, the formulation of the subject, for the time he devoted and his support until the end of my research.

Also, my greatest thanks go to my pattern in life and in study and my family, who supported me and guided me through my study and growth of my work.

INTRODUCTION

The purchase of different products from a supermarket chain leads to fulfillment of needs in any period of time in a household. It's an area that have been researched a lot for many different reasons. Main reasons performing research in a supermarket chain is the quality of a product, the customer service, and the variety of products (local and import) . The supermarket chains contain all kinds of products such as food, beverages, cleaning products, flour, some supermarkets have electronics appliances for kitchen, books, clothes coffee, snacks etc.

Our study is of high importance as it aims to detect the patterns in shopping philosophy of Greek people during the Covid-19 pandemic. The spread of Covid-19 disease started through the arrival of people from abroad which they didn't know how to protect themselves. That situation, which is unknown to the world and extremely dangerous for the health of the people forces the Governments all over the world to impose an extreme method as a precaution in order to stop the spread of the illness to their countries.

The company from which took our data is BAZAAR S.A. is a medium chain in the supermarket industry with domestic products, products from all over Europe and other places all over the world.

The significance of our research relies in its ability to identify trends in consumption of products during the pandemic of Covid-19 in an extensive range of product categories and retail locations. The business hopes to draw countable conclusions about the market potential of certain items from the examination of historical data. An executive member of the firm provided Excel files that were used for quantitative research to collect primary data for the study. The dissertation will assist businesses understand the requirements and expectations of consumers during that period and take the required steps to boost the efficiency of the supply chain by studying production/demand, trends, and cycles.

For almost 25 years, Bazaar S.A, a part of the Veroukas group of enterprises, has been providing outstanding service to its customers. Bazaar S.A has over 180 locations around Greece, aims to achieve its goals and constantly being at the customer's and professionals in catering providing them with products of high quality from all over the world. Bazaar S.A.'s main goal is to satisfy the demands of its customers consistently and professionally. The Bazaar company's core strategies in this endeavor are its low prices and ongoing promotions. Bazaar's primary goal is to continuously expand and improve its services in response to the demands of its everyday customers.

CHAPTER 1 LITERATURE SURVEY

1.1: Risk of Businesses confronting Natural Disasters

Numerous small enterprises and nonprofit groups in the US experience significant losses each year as a direct result of earthquakes, powerful storms, financial crisis and flooding. Failures of small businesses and nonprofit organizations result in large losses for towns, communities, and cities of various sizes which they are reliable from their local industries in order to maintain their economic status. There are substantial losses for the person as well as the community when even one of them has to close. Owners of small businesses and non-profit organizations are frequently used to living under constant danger. Natural disasters frequently mean the difference between life and death for them.

We concluded that not many risk consultants, small company owners, or experts in natural hazards truly comprehend how natural catastrophes lead to the demise of small enterprises or their excruciatingly drawn-out recovery processes. The majority of owners were uncertain of the best ways to attempt to restore to health.

All other factors being identical, we discovered that the likelihood of an organization closing its doors following an incident was significantly higher when it came to smaller, weaker, and under substantial stress before to the occurrence. Even though they only encountered a little damage, minor businesses and those that were on the verge of collapse frequently ended when the incident happened. Occasionally, the catastrophe just behaved as the straw that broke the camel's back. It gave them a decent excuse to abandon a losing struggle, most likely because the organization would not be able to survive the extended damage, trying calendar weeks followed the tragedy under any circumstances. Many company owners who had small businesses before the natural disaster did reopen and persevere through protracted, agonizing periods of recuperation before eventually giving up the fight due to exhaustion of hope, resources, and fortitude.

Natural disasters can do significant harm to even well-established businesses. Any prolonged term of firm closure may result in a loss of market share. Regaining market share can be quite tricky, even with business interruption and property and liability insurance. Early on in the post-disaster recovery process, the firm owner's perceived skills and experience base appear to be the deciding factors. A complicated self- and social- image of who they are and what they can do restricts owners and operators. Those that take active steps to enhance their company potential are the ones who succeed. As long-term survivors informed us, it is impossible to turn back. You must never stop focusing on the future.

We have also created a new phrase, managerial mitigation, to refer to management strategies used to lower exposure and weakness by applying wise business decisions. These strategies can also involve diversifying the company's customers, shifting the location of its merchandise, maintaining its hard copy and electronic records, and operating many locations. Having many physical locations, conducting business via catalog, or engaging in e-commerce are examples of multiple business outlets. Simple measures like establishing reasonable lease conditions, which permit an organization to evacuate a property if it is unable to achieve required performance standards and the owner is unable

to quickly reclaim it, are examples of managerial mitigation practices. Businesses that use rented premises and have insufficient lease terms dictating who will fix earthquake damage and when they will do it may run into problems. Many company owners in Northridge were prevented by their lease from transferring to another place where they could continue operations, thus they were forced to stay in buildings that had not been restored for a long time.

We've previously talked about the significant psychological damage that managers of small businesses and nonprofit organizations face both during and after a destructive natural disaster. The incident is frequently a 360-degree trauma that affects one's livelihood, family, home, and self-worth. Some were under so much stress that they just decided not to reopen the company following the incident. They just vanished. Others shown incredible resilience, battling stress and strain with a psycho-social resilience that is just astounding. Sadly, stress usually wins out in these situations, the business shutters, and the loss has a lasting impact on the owner or operator's life.

We've concluded that the owner or operator's ability to identify and adjust to the post-event circumstance may be the most crucial factor in the survival equation. Communities never revert to their pre-event state. The aftermath of an incident is never the same. There is a great opportunity for survival and revivability for those who recognize the changes and react accordingly. The deck is stacked against those who want to operate under the outdated paradigm, believing that things will eventually return to as they were before.

We are in a position to translate our understanding of the key variables that distinguish survivors from non-survivors into recommendations for smaller businesses, both profit-seeking and nonprofit. The scene is set for the second phase of the endeavor, which entails creating guidelines for small businesses to increase their odds of surviving a natural disaster and recovering financially after it has occurred.

1.2 Disruption Management in Supply Chain

When supply chain reliability is considered, candidate location issues are significant in the field of operational research. Since supply chain dependability affects many critical elements of any supply chain, particularly money, it is significant to almost every sector. Numerous sectors' supply systems are continuously vulnerable to man-made and natural disturbances. As a result, the supply chain is exposed.

Because of this, supply networks have become more and more vulnerable over time. Natural catastrophes often strike nations all around the world, causing enormous economic harm. Asian nations are renowned for frequently experiencing flooding interruptions. Asian nations—including those in Southeast Asia, Bangladesh, Vietnam, Thailand, Cambodia, Indonesia, and the Philippines—frequently experience flooding. Among the aforementioned nations, Thailand was also affected by disruptive floods in 2011. whereby several supply networks suffered.

NASA (2011) said that Thailand can be considered among the most severely affected countries due to the 2011 floods. Thailand was affected by these flood scenarios in a number of ways. According to World Bank (2012) estimates, Thailand's real GDP growth

decreased in 2011 from 2.90% to 4.10%. The 2011 Thai floods caused an estimated US\$ 45.7 billion in economic losses, according to a World Bank analysis. The interruptions at industrial parks resulted in an 8.60% decline in GDP (real) supported by industries, according to the Japan Economy, Trade, and Industry Ministry (2012). Thailand frequently experiences flood catastrophes, which highlights the significance of maintaining dependable supply systems in Thailand.

Thailand's economy is now a part of global supply chains due to globalization, which is significant for many industries. Thus, the supply chain disruptions brought on by Thailand's periodic floods have an effect on both the global and Thai economies. It may potentially affect Thailand's international ties with other nations. Thailand's supply chain reliability is important to all parties involved, both domestically and globally since the nation has one of Asia's fastest expanding economies.

Thailand's economy is expanding, and its citizens have a keen interest in high-end fashion. Foreigners as well as locals view Thailand as a location for shopping. A strong indicator of this is the abundance of high-end retail options.

Accessible to both tourists and residents Thailand is one of many reasons, including the country's booming tourist industry, the middle class's recent growth, and the promotion of social media, are driving up demand for luxury items among Thai consumers. The European and other western origins of these luxury brands have caused the Thai luxury fashion industry to naturally link with the global supply chain.

1.3 Market analysis and forecast in beverage Industry

Shopping habits have been extensively studied. Over the past century, the retail business has grown significantly, includes a range of smaller retail marketplaces that are now offering newly developed goods, such as products from all over the world which are not so popular in the retail industry and free of preservatives, artificial tastes, fragrances, or colors, which are referred to as the best products and oddest. Bazaar S.A is only one of the many businesses involved in this market area.

The significance of our research lies in its ability to identify trends in consumption of products from Supermarkets and other small stores across a wide range of product categories and retail locations. The business hopes to draw quantifiable conclusions about the market potential of certain items from the examination of historical data. The company we investigated provided Excel files that were used for quantitative research to collect primary data for the study. The dissertation will assist the business understand the requirements and expectations of current consumers and take the required steps to boost the efficacy of the supply chain by studying production/demand, trends, and cycles. The societal interpretation of taste will also be investigated in this study.

Initially, the nations and goods that are of utmost relevance to the business in terms of sales and expansion have become apparent. Secondly, it was found that their approach is undoubtedly predicated on models that enable them to forecast their sales. We also

conclude that certain regions have seasonal contracts, some are basic clientele with varying monthly demands, and still others are fresh arrivals. The models' significant outcomes have applications in both practice and research.

1.4 Trends in food retail sales during the Covid-19 Pandemic

All facets of life and the American economy were severely disrupted by the COVID-19 epidemic, including the food and agriculture industries. States implemented extreme public safety measures (such as suspending in-person schools, closing, or limiting restaurants and other enterprises, and issuing stay-at-home orders) in the early weeks after a national pandemic emergency was declared on March 13, 2020. Residents of the United States were much more likely to spend time at home as a result of these actions combined with a voluntary reaction (Gupta et al., 2020). In February 2020, the U.S. economy entered a recession that resulted in sharp rises in unemployment and adjustments to food expenditure (Weinstock, 2021; Dong and Zeballos, 2021; Restrepo et al., 2021).

From the farm to food processors, packers, merchants, and food service facilities, the pandemic disrupted nearly every stage of the food supply chain, particularly for meat products like beef and pig (Chenarides et al., 2021a; Johansson et al., 2021; Lusk et al., 2020; Mallory, 2021). Additionally, data indicated a notable rise in the proportion of Americans (particularly the elderly) who had difficulty finding enough food through at least July 2020 (Ziliak, 2021). In response, officials enhanced the social safety net by increasing food assistance benefits (Bitler et al., 2020; Jones, 2021) and fortifying food supply chains by aiding farmers (Johansson et al., 2021; Giri et al., 2021).

All food categories had positive year-over-year changes in national food retail sales during the 52 weeks that followed the declaration of the pandemic emergency (i.e., the week ending March 15, 2020, through March 7, 2021), though to differing degrees. Just four food retail categories saw year-over-year increases in real dollar value of sales that fell short of the 10.7 percent change in total food retail sales during this period: fruits (9.9%), dairy (9.4%), commercially prepared items (7.5%), and beverages (5.9%). The real dollar value of food retail sales increased year over year in all other categories, with the "other" category seeing the largest gains (12.5%) and fats and oils seeing the largest (20.1%).

The relative composition of sales changed very little 52 weeks before and after the national emergency was declared, according to an analysis of the percentage of actual food retail sales by category. For all but three categories, the absolute magnitude of changes in actual sales share from year to year was less than 0.5% points. Meats, eggs, and nuts saw the biggest share increase: 52 weeks prior to the national emergency declaration, they accounted for 17% of the actual value of all food retail sales; 52 weeks after, that percentage increased to 17.8%. The largest share decreases occurred in commercially prepared items (30.1% versus 29.2% over the 52 weeks before and after the emergency declaration) and beverages (14.5% versus 13.9%), which simultaneously had the smallest

concurrent year-over-year increases in real sales value, and the first and third largest real sales share among all categories.

CHAPTER 2 MARKET ANALYSIS IN GREECE

2.1 An Overview of the Greek Grocery Retail Sector

An extended summary of a study conducted by the Institute of Retail Consumer Goods in Greece (IELKA) is provided in this report. The study's objectives are to measure the impact of the Greek grocery retail sector on the country's economy, as well as to record and examine the industry's key performance indicators, present trends, and future prospects. The grocery retail industry is made up of street markets, which account for a sizeable portion of the market in Greek cities, specialist businesses like butchers, greengrocers, liquor stores, and fish stores, and non-specialized retailers like supermarkets and grocery stores.

The following subjects are included in the analysis: The magnitude of the market

The nationwide network of stores

Employment numbers and prospects

Investments aimed at enhancing productivity and competitiveness

Supporting the country's economy

Comparing and analyzing financial ratios

Index of end-user prices

Development Opportunities

1. Market Size of the Greek Grocery Retail Sector

Depending on the criteria applied to the different types of stores, the grocery retail sector can be examined and recorded at many levels.

Three tiers are defined in this report:

a) Supermarkets (A): The supermarket is a subset of grocery stores that falls under the following definitions: a) offers a wide range of products (at least seven out of ten categories of consumer goods); b) is divided into sections and product categories; c) has a minimum sales area of 200 square meters; d) offers self-service options for customers to make purchases and e) Runs a minimum of two cash registers

Supermarkets can be further classified as hypermarkets, discount shops, cash & carry, etc. depending on their size or other characteristics.

b) Grocery shops (B): These establishments sell a variety of foods that they stock to their patrons. Convenience stores or delicatessens are the names for small grocery stores that primarily offer sandwiches and snack foods; produce markets, or greengrocers in the United States, are the names for these establishments in Britain. Supermarkets are included in the definition of grocery shops (category A), which is restricted to non-specialized retailers.

c) Grocery retailing (C): The total of all retail sales of groceries to the final consumer, including those in categories A and B, is referred to as grocery retailing. Supermarkets, grocery stores, convenience stores, kiosks, minimarkets, liquor stores, bakeries, butchers, greengrocers, fish shops, other grocery specialty stores, street markets, and direct sales from producers to consumers are all included in this broad definition of the sector.

The overall annual turnover of food retail in Greece is estimated by IELKA's study to be between 30 and 32 billion euros . According to IGD, the market might be worth up to 35 billion euros, which is comparable to the combined GDP of Belgium, the Netherlands, and Switzerland. Nonetheless, grocery chains only account for 40% of Greece's 11.5–12.5 billion Euro overall revenue. The fact that Greece has the lowest percentage of food retailing in non-specialized stores in Europe—63% compared to an average of 84% — validates this strong indicator of future expansion potential for the supermarket chains.

Over the past ten years, there has been a notable increase in turnover within the Greek grocery retail industry, particularly in the supermarket sub-sector. More precisely, the turnover of supermarket outlets increased significantly by 60%. This is a result of numerous businesses making significant investments and modernization efforts, foreign supermarket chains entering the local market, changes in Greek consumers' purchasing habits, and a tendency to adopt European market division strategies between specialized and non-specialized stores.

2. Employment Opportunities

While employment in retail commerce expanded by just 4.3 percent overall during the preceding ten years, grocery retail direct employment increased by 30%, one of the largest increases at the business sector level in Greece.

Presently, the food retail industry employs over 190,000 individuals, 40% of whom work for supermarket chains. Unlike employees in other industries, these workers are dispersed across the nation. One of the best-performing economic sectors in preventing both young and long-term unemployment is the grocery retail industry.

It is clear that grocery shops provide job opportunities for a sizable portion of the populace, including non-specialized workers, women, young people, and the elderly, all of whom have higher unemployment rates.

The grocery retailing industry is labor-intensive, with personnel expenditures accounting for the majority of operational expenses. Large Greek supermarket chains' data indicates that these expenses account for about two thirds of operational costs; smaller chains, with fewer marketing and administrative expenditures, might see even greater employment cost ratios.

3. Investments for Productivity and Competitiveness

IELKA's report indicates that the industry has made significant investments in the past ten years. Over 6 billion euros were invested overall between 2002 and 2010.

Interestingly, because of the previously mentioned increased retail networks, these investments are prevalent across the nation and are not subsidized. Supermarkets account for 63% of all investments, although having a relatively small 40% share in the sector's sales. These expenditures fall into a number of areas, but the bulk (about two-thirds) went toward the retail network's technical and organizational improvements with the goal of increasing efficiency in the supply chain, competitiveness, productivity, and customer responsiveness, among other things.

The majority of these expenditures went toward creative projects meant to improve corporate operations and bolster long-term, competitive strategy. A number of these investments were made in partnership with suppliers and took advantage of the openings created by the ECR Hellas organization's programs.

For the grocery retail industry, the approach, and results of a few of the initiatives might be regarded as national and even European best practices. A few of these initiatives are as follows: one of the best out-of-self-prevention systems in the world was developed; another was a collaborative system that managed discount and promotion coupons digitally; still another was an extremely effective electronic invoicing and ordering system; still another was centralized distribution infrastructure; still another was self-checkout cashiers; WMS systems; VMI systems; ERP systems; RFID prototype systems for tracking and tracing; and still another was Master Data Alignment infrastructure. The bigger businesses in the industry also make significant investments in recycling programs, new "Green" storefronts, and sustainability in general.

4. High Contribution to the National Economy

Based on the input-output model, the research measures the grocery retail sector's contribution to the Greek economy by utilizing data from Eurostat and the Hellenic Statistical Authority (EL.STAT).

An economic method that is quantitative in nature and describes the interdependencies between several areas of the national economy is the input-output model. The figure below illustrates how the model assesses three distinct but related forms of economic impacts: direct impacts, which are generated by the sector directly; indirect impacts, which are

generated by other sectors whose economic output is directly and exclusively linked to the grocery retail sectors; and induced or multiplier impacts, which are generated by other sectors of the economy as a result of successive rounds of expenditure/procurement.

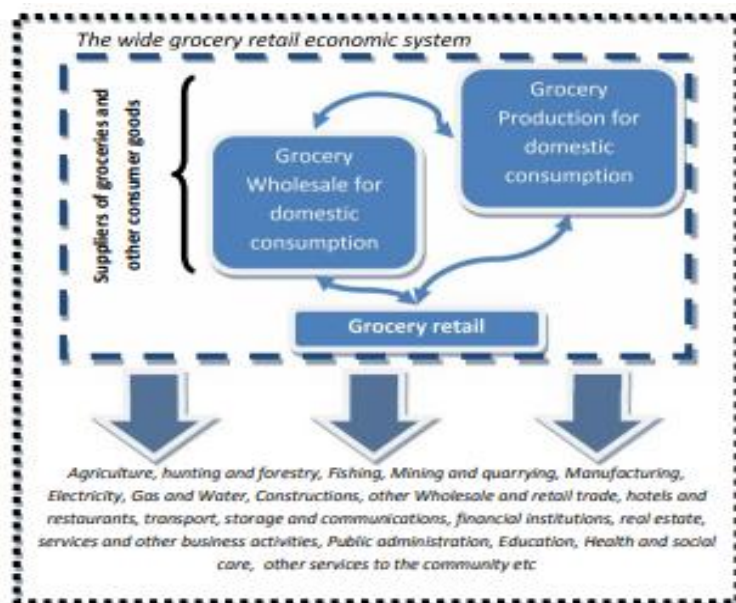


Figure 1: A diagram of a grocery retail system

Consequently, the combination of these three effect kinds represents the sector's overall economic impact. The Greek grocery retailing industry has a significant overall impact, demonstrating the industry's significance for both the GDP and employment of the Greek economy and society. In particular:

- The Gross Value Added (GVA) generated for the Greek GDP is projected to be 14.03 billion euros, or 7.01 percent of the Greek economy's overall GVA.
- A total of 10.89% of Greek employment is found in the grocery retail industry, which employs 521,000 people. Due to the generation of income for the country, the establishment of numerous job possibilities, and the initiation of new company ventures, these numbers demonstrate the significance of the grocery retail sector for the Greek economy and society. With a large contribution to GDP, GVA, and other macroeconomic indices, the grocery retail industry is an essential component of the economy.

5. Grocery retail price analysis

It is believed that the Greek grocery retail industry positively influences ultimate consumer pricing. Based on data from the Greek General Secretariat of Commerce's official price observatory, the bimonthly study by IELKA on the evolution of retail prices in grocery retail in 2010 demonstrates that Greek supermarket chains and their suppliers have made significant efforts to absorb a significant amount of the new heavy taxation in 2010 (caused by the

government measures for the Greek debt crisis⁴), which has actually resulted in a reduction of adjusted product prices. The IELKA index encompasses over 200 product classes and is derived from 996 goods that are available in Greek grocery chains. More particular still:

For the full year 2010, there was a 0.79 percent gain in the total index.

- The adjusted total price index, which excludes the rise in VAT, decreased by 1.61%, indicating that increases in VAT were the main source of this increase. Upon removal of the VAT rise, the average price of 58% of the items and 79% of product categories experienced a decrease in price.
- The process of prices in Greek supermarkets from January 2010 to April 2011, which was substantially less than both the inflation rate (7.22%) and the overall growth in food retailing prices. This include other outlets including specialty shops and street markets (4.59%). Supermarket prices increased by just 1.73% during the same time, but if the new VAT tax rise is taken out of the equation, prices actually decreased by - 1.77%.
- The average costs of Greek supermarket chains are cheaper than those of English (16%) and Spanish (2%) supermarket chains, according to a comparison of a typical supermarket shopping basket between Greece, Spain, and England in April 2011. When the value-added tax (VAT) is subtracted from each nation, the average basket in Greece is less than that of England and Spain by 22% and 7%, respectively.

About IELKA (Institute of Retail Consumer Goods – GR)

Founded in October 2010, IELKA is a non-profit research group that counts the main Greek supermarket chains among its members. The objectives of IELKA are:

To conduct empirical study and advance scientific inquiry on new challenges in food retail

To offer unbiased data about the primary markers and patterns of food retail

To address new issues and difficulties and offer viable, scientifically supported answers

To work together to cooperatively promote research and projects with comparable Greek and international institutes, universities, research groups, etc.

By upholding academic and scientific impartiality and objectivity, IELKA achieves its goal. IELKA employs the following techniques, instruments, and methods:

Research and surveys conducted internally, assigned, or in partnership with external entities
A supervisory body for Greece's grocery retail industry's key metrics and indices

Books, articles, and press releases

Workshops, conferences, seminars, and events.

2.2 THE FUTURE OF RETAIL

One of the most significant areas of the global economy, retail is expanding and developing steadily. It has shown resilient both domestically and internationally, despite the difficulties it has experienced recently due to the pandemic, interruptions in the supply chain, the oil crisis, and inflation. One of the biggest subsectors of the retail trade is the organized retail trade of foodstuffs, which offers tremendous development potential both domestically and internationally.

The structure, scale, development, and financial performance of organized grocery retail in Greece are the main topics of The Future of Retail. A high-level description of the main rivals in the market and their traits follows. Lastly, it discusses the projected trajectory of the worldwide organized retail grocery industry, stressing the dominant trends, the cutting-edge technology, and the structure that has developed in the Greek market.

The organized grocery retail trade in Greece has had average annual turnover growth rate of 7.1% over the previous five years, which is comparable to the entire retail trade, notwithstanding the volatility in the world economy. During the same time frame, Greece's organized grocery retail trade's gross profit margin increased at an average annual rate of 3%. The debt-to-profitability ratio (net debt / EBITDA) for 2021 is lower than it was in previous years, at 3.9 billion, suggesting that companies are controlling and increasing their profitability. There are now 2,700 organized retail food stores in Greece, divided among 53 companies. With a turnover of almost EUR 8.46 billion in 2021, the six biggest rivals in Greece had a 61% rise in their respective revenue from 2017.

The dominant trends worldwide for organized grocery retail businesses are as follows:

The spread of augmented and virtual reality, such the virtual realm known as Metaverse.

The widespread use of artificial intelligence in commercial solutions.

Autonomous product delivery as a less expensive alternative mode of distribution.

Alliances and networks among companies to provide cutting-edge goods and experiences.

The following are the main concerns for organized retail grocery enterprises in Greece:

increasing inflationary trends causing economic uncertainty.

the specifications needed to create supply chain strategies that are competitive.

higher energy expenses.

Greek customers' expectations, which are always shifting.

The following are the main trends in the Greek market:

acquiring small-scale local chains in order to increase their physical presence in the market.

customers' reentry into the private label industry.

Business investment and consumer engagement in ESG initiatives.

Quick commerce is preferred by customers who want a certain selection of things delivered really quickly.

The organized food retail business is expanding significantly in spite of the ever-changing global barriers and difficulties. Adoption of contemporary technology will help both parties greatly and advance the industry. It will be used in supply chain operations, back-office procedures, and direct customer interactions during the purchase process.

2.3 Offline Retail and the COVID-19 Pandemic

Retail refers to the actions that take place when customers are choosing which goods to purchase. Though many studies have attempted to explain and identify the elements that influence consumer behavior, it is difficult to measure or describe what happens depending on consumers' ideas. This is known as consumer behavior. Numerous societal, cultural, and personal aspects influence consumer behavior. Social elements are the most erratic and challenging to manage among them. The climate and pandemic illnesses are two of them. Regarding the former, customer propensity to visit stores and make retail purchases is influenced by temperature and fine dust concentrations. Another significant societal aspect influencing consumer behavior is the presence of pandemics. South Korea was impacted by the severe acute respiratory syndrome (SARS) outbreak that struck China in November 2002, resulting in 8465 confirmed cases and 801 fatalities. Hong Kong and South Korea both suffered economic losses at the time in every industry. After the SARS pandemic seven years ago, the new wine-origin influenza A 2009 epidemic in South Korea resulted in more financial losses, particularly in the travel industry.

Beginning in December 2019, the COVID-19 pandemic reached South Korea in January 2020. Shin explained how COVID-19 resulted in a decline in sales, and how that decline was connected with the prior sales scale. The COVID-19 pandemic reduced customers' desire to make in-person purchases, which had a significant negative influence on retail sales at big-box stores, duty-free stores, and urban retail spaces. The CCSI, which was 70.8 in South Korea in April 2020—lower than it was during the 2008 financial crisis (77.9)—reflected this. In an effort to stop the COVID-19 virus from spreading, a social distance policy was put in place. This, together with people's fear of infection, caused people to stay indoors, which in turn caused a decline in offline retail sales.

2.4 Sales Forecasting

Logistics management includes retail, and like all logistics management, demand unpredictability is a significant obstacle. In the retail sector, sales forecasting is the most effective way to deal with demand unpredictability. It can help managers manage the product inventory efficiently, preventing losses brought on by either an excess or insufficient stock of products. Accurately predicting future data is challenging because retail sales data, which are

time-series data, differ from the data often utilized for general regression analysis. Several statistical techniques may be used for sales forecasting, which uses historical sales data trends to predict future sales, in order to increase accuracy. The two forecasting analysis models that are most represented are ARIMA and ETS. They are frequently used in the retail sector to anticipate and evaluate retail sales in the restaurant, fashion, and entertainment retail sectors. This study used the ARIMA and ETS forecasting models to predict retail sales at shopping malls during the COVID-19 epidemic.

The COVID-19 pandemic clearly had a detrimental effect on offline retail sales, but the amount of data that could be used to support assessments of offline retail sales was constrained by the global implementation of lockdown and social distancing measures. This study forecasted the sales in 2022 by retail category by analyzing offline retail sales data collected prior to and during the COVID-19 pandemic. The entertainment and food and beverage retail sectors will continue to be most affected by any more COVID-19 outbreaks and the execution of the ensuing social distancing policies. Even if COVID-19 and social distancing regulations persisted, the fashion sector was predicted to recover mostly from the effects of the pandemic and post-COVID-19 sales levels in 2022. The COVID-19 pandemic had little effect on the sports and cosmetics retail categories, with the exception of the month when the epidemic occurred. It is worth mentioning that within a single shopping mall, the COVID-19 epidemic had varying degrees of influence on different retail categories. Various facets of the theory of consumer behavior have an influence on each category and should be examined. As a result, each retail category had to develop a unique plan to manage and combat the pandemic's effects. Retail categories related to dining out had a special difficulty and required the most assistance to combat the COVID-19 pandemic's effects.

These days, offline stores are more than just places to buy things; their settings are designed to meet the requirements and preferences of potential clients. Offline retailers must meet two essential criteria in order to interact with customers in a meaningful way: (1) matching demands and preferences of the customer and customizing the purchasing experience; and (2) evoking certain emotions in the customer through the consumer's experience. This study shows that, even inside a single building, the effects of COVID-19 differed by store type. This is related to how customers behave in every retail sector and how the epidemic has caused impulsive or panicked purchases. Future research should consider panic buying, which is motivated by fear and anxiety in situations like pandemics, natural catastrophes, or protracted strikes. Additionally, the varying effects on different retail categories highlight how crucial it is to create focused measures to mitigate the effects of a pandemic, especially for those that depend on customers dining out. Overall, this analysis shows that different offline retail shop types were affected differently by the COVID-19 pandemic, resulting in differing post-pandemic sales projections for each retail category. These findings have several important implications for academics, one of which being the necessity of developing a sophisticated research approach for the offline retail industry.

CHAPTER 3: METHODOLOGY

The scope of the current research is to examine the appropriate method to predict possible statistical features, possible predictable sales, and possible connection across the Covid-19 limitations.

Moreover, it aims to focus and describe research methodology that includes data collection method and results research methods to give managers and executives instructions by choosing methods, tools, and techniques to help the company recognize the needs/expectations of current costumers and take the required actions to fulfill the petition of customers and to increase the productivity of supply chain.

Finally, this strive is in fact descriptive research in the sector of Super Market industry which has taken place through quantitative data investigation to provide perceptions and understanding in relation to the subject under examination data of 9 different categories of product.

BASIC TYPES OF RESEARCH METHODOLOGY

Research tools

As per Moon, forecasting is a management process that requires meticulous organization, with special attention given to the people, procedures, and tools that make up forecasting management, just like any other management process. (Michele S. Garver, Carlo D. Smith, John T. Mentzer, and Mark A. Moon, 2018).What kind of forecasting technique is most appropriate for a given circumstance depends on a number of variables. Forecasting approaches fall into two categories: quantitative and qualitative. To identify temporal demand trends and derive quantitative inferences from data, we employed a variety of quantitative methodologies in our study.

Trend, seasonality, and noise are identified and predicted through the application of time series statistical methods. A trend is a pattern that moves as a straight line or curve along a slope, either gradually or decreasingly. In actuality, the trend component indicates whether and by how much sales are increasing or decreasing. Seasonality refers to patterns and cycles that recur across time in the same manner. For instance, actual seasonality asks what impact, if any, the summer season has on sales.

Lastly, randomness represents a fluctuating pattern in the demand history that is not discernible from the other time series components.

A daily sales prediction is actually necessary in the retail business to help shop managers place exact orders without sacrificing waste or stock-outs. On the other hand, time series that show daily sales of consumable products at retail establishments are typically quite volatile and skewed, and they also vary with time. Although these are significant restrictions, forecasting frequently ignores them. To address these problems, a time series forecasting model that considers forecast uncertainty as well as the impact of outside factors on sales, such as weather, holidays, festivals, price reductions, and seasonal variations by day of the week and month of the year, is required. There aren't any techniques that are obviously better than the particular data over others. Pros and drawbacks apply to each strategy. In this study, we employ descriptive statistics to identify the most important variables that will provide us with the most relevant insights. We next run two models, regression and EWMA, to determine which model best fits our data.

3.1. Descriptive statistics

In order to meaningfully analyze, characterize, arrange, or summarize data so that, for instance, patterns may appear in the data, descriptive statistics is helpful. Descriptive statistics, however, do not enable us to draw inferences about hypotheses we may have entertained or to draw conclusions from data beyond our analysis. They are only a means of characterizing our data.

Following data collection in quantitative research, answers are described by statistical analysis, which might include the relationship between two variables or the average of one variable (like sales).

Three primary categories of descriptive statistics exist:

1. The frequency-measurement distribution
2. The central tendency, which calculates the mean, median, and mode; and 3. The variability, which includes the variance, standard deviation, and range, which calculate the values' spread.

Descriptive statistics were employed in our study to help us comprehend the research in a concise summary by enabling us to use the data for preliminary interpretation.

3.2. Exponentially Weighted Moving Average

Examining the qualities of a suitable model to comprehend, recognize, and project past patterns into the future was the primary goal of this study. Using the EWMA forecasting approach, an analyst may give previous data varying weights. Furthermore, the EWMA model succeeds in accurately capturing the short-term dynamics of a time series.

Lambda, known as the smoothing parameter. The decision-maker simply needs to provide the value of a single input parameter, $0 < \lambda < 1$, which represents the appropriate trade-off between historical and current understanding. The parameter determines the current observation's significance in the EWMA computation. The more closely the EWMA resembles the original time series, the greater the value of lambda. EWMA's secondary goal is to monitor volatility changes; hence, for small λ values, recent observations have an immediate impact on the estimate, and for λ values closer to one, the estimate responds more gradually to recent changes in the underlying variable's returns. The EWMA formula is presented as follows:

$$EWMA(t) = \lambda * EWMA(t-1) + (1-\lambda) * sales$$

3.3. Regression analysis

Regression analysis, according to Moon, is a statistical technique that is helpful when demand is being influenced by a measuring component other than time. (Michele S. Garver, Carlo D. Smith, John T. Mentzer, and Mark A. Moon, 2018). Two different types of variables exist. The primary variable you are attempting to comprehend, or forecast is called the dependent, and the independent variables are made up of external and internal elements that you assume may have an effect on your dependent variable.

The following are identified and explained by time series analysis:

1. Any regularity or systematic change in the data series that results from seasonality—the "seasonal."

2. Trends in cycles.
3. Data trends.
4. The trends' rates of growth.

Regression analysis is a dependable technique for determining which variables affect a certain issue, which factors are important, which factors are negligible, and how these factors interact. Regression model yields four key variables: residuals, p-value, adjusted and predicted R-squared. These variables give us crucial information.

In general, you want to select the models with better predicted and adjusted R-squared values. These statistics are intended to circumvent a significant issue with standard R-squared, which might lead you to describe an excessively complicated model because it grows each time a predictor is added. The adjusted R squared might drop with low-quality predictors and only rises when the additional term enhances the model more than would be predicted by chance.

Regression terms with low p-values are statistically significant. The process of adding all potential predictors to the model and then methodically eliminating each term with the highest p-value one at a time until you are left with just significant predictors is known as "reducing the model."

The difference between the actual value and the predicted value that the model predicts for a certain observation is how the residuals are computed. Because they show how much a model explains the variance in the observed data, then are helpful in regression. The most straightforward model that yields random residuals is a strong contender to be a reasonably accurate and objective model.

3.4. Mean Absolute Percent Error

The Mean Absolute Percent Error is a useful metric to assess the accuracy of forecasts and evaluate the overall performance of processes. It is calculated by dividing the total absolute value of percent errors from previous periods by the number of periods for which errors have been tracked. (Mark A. Moon, Carlo D. Smith, Michel S. Garver, John T. Mentzer, and T2018) The mistake is presented as a percentage. The more accurate the forecast, the lower the MAPE. is also helpful for regression as it gauges the precision of a forecasting system.

3.5. Mean Absolute Error

The MAE statistic quantifies the discrepancies between two sets of data. The total of the absolute differences between the actual and expected values is what it is, in fact. Because the faults are smoother in this one, we choose the one with the lowest MAE.

CHAPTER 4 EMPIRICAL ANALYSIS

4.1 DATA COLLECTION PROCESS

The Bazaar firm delivered the data under a non-disclosure agreement. Excel data was used to gather information and the related product based on the factors we selected. The information covers a period of four years, specifically from 2019 to 2022. Each year, the product quantities that were sold were gathered, and we maintain monthly entries.

Preprocessed data was transferred into a format that could be used with quantitative forecasting methods. We planned to test a range of model parametrizations, such as regression devices and exponential moving average EWMA, among others.

4.2 PRELIMINARY ANALYSIS

4.2.1 Data selection

We had to pick those that we could utilize for our study and present useful results and ideas to the Bazaar firm because the data that was provided for all items, respectively, was restricted because we had to look at amounts for the years 2019 to 2022. Because of this, we used Microsoft Excel program to arrange the data in sheets on which we labeled according to the year we refer to from 2019 through 2022. We constructed a table per year which is consist of the categories of interest which are Stationery, Antiseptics (Gel-suspension), Chlorinated Cleaners, Alcohol & Alcohol Lotions, Coffee, Snacks, Flour, Dairy, and Alcohol Beverages and sales per year corresponding to volume or value.

We are doing our analysis on different sheets according to the product category and we will conclude with our analysis.

4.2.2 Descriptive statistics

Using the Excel application's Descriptive Statistics from Data Analysis function, I selected all of the data for each year independently. I have included a representation of the year 2019 in table below. (Data that may be supplied upon request is available for the remaining categories and their sales.)

Table 1: Descriptive Statistics for year 2019

	January	February	March	April	May	June	July	August	September	October	November	December
Mean	227,866.05 Mean	230,759.32 Mean	235,304.26 Mean	240,446.20 Mean	233,812.18 Mean	219,933.82 Mean	226,270.83 Mean	208,551.54 Mean	234,723.22 Mean	261,217.57 Mean	263,325.98 Mean	294,118.85 Mean
Typical Error	79,583.77 Error	83,292.09 Error	81,965.05 Error	83,745.48 Error	79,178.85 Error	73,929.89 Error	76,911.97 Error	70,573.42 Error	79,482.38 Error	92,249.84 Error	94,145.18 Error	105,307.65 Error
Median	199,999.02 Median	165,270.32 Median	190,916.03 Median	184,141.36 Median	190,814.37 Median	170,606.22 Median	162,256.47 Median	153,539.71 Median	172,294.02 Median	195,128.17 Median	197,421.27 Median	202,184.12 Median
Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode	0 Mode
Standard Deviation	238,751.32 Deviation	249,876.26 Deviation	245,895.14 Deviation	251,236.45 Deviation	237,536.54 Deviation	221,789.68 Deviation	230,735.90 Deviation	211,720.27 Deviation	238,447.13 Deviation	276,749.52 Deviation	282,435.53 Deviation	315,922.94 Deviation
Sample Variance	57,002,192,786.74 Variance	62,438,145,835.83 Variance	60,464,417,968.33 Variance	63,119,751,451.70 Variance	56,423,607,875.34 Variance	49,190,663,703.92 Variance	53,239,057,817.92 Variance	44,825,474,410.76 Variance	56,857,035,995.46 Variance	76,590,299,186.38 Variance	79,769,626,470.52 Variance	99,807,302,144.49 Variance
Kurtosis	-0.64 Kurtosis	-0.04 Kurtosis	-0.38 Kurtosis	-0.26 Kurtosis	-1.36 Kurtosis	-1.84 Kurtosis	-1.87 Kurtosis	-1.83 Kurtosis	-1.62 Kurtosis	-0.63 Kurtosis	-0.14 Kurtosis	-0.44 Kurtosis
Skewness	0.78 Skewness	0.95 Skewness	0.82 Skewness	0.86 Skewness	0.55 Skewness	0.43 Skewness	0.46 Skewness	0.47 Skewness	0.51 Skewness	0.80 Skewness	0.93 Skewness	0.87 Skewness
Range	649,505.93 Range	707,785.68 Range	691,246.82 Range	714,980.49 Range	617,182.16 Range	531,164.12 Range	544,186.22 Range	492,057.93 Range	598,953.55 Range	761,875.30 Range	799,237.31 Range	882,026.92 Range
Minimum	82.62 Minimum	253.02 Minimum	132.26 Minimum	80.97 Minimum	100.58 Minimum	62.70 Minimum	100.23 Minimum	127.29 Minimum	211.03 Minimum	154.23 Minimum	106.59 Minimum	122.49 Minimum
Maximum	649,588.55 Maximum	708,038.70 Maximum	691,379.08 Maximum	715,061.46 Maximum	617,282.74 Maximum	531,226.82 Maximum	544,286.45 Maximum	492,185.72 Maximum	599,164.58 Maximum	762,029.53 Maximum	799,343.90 Maximum	882,149.41 Maximum
Sum	2,050,794.43 Sum	2,076,833.90 Sum	2,117,738.54 Sum	2,164,015.79 Sum	2,104,309.64 Sum	1,979,314.38 Sum	2,036,437.45 Sum	1,876,963.83 Sum	2,112,509.01 Sum	2,350,958.15 Sum	2,369,933.83 Sum	2,647,069.63 Sum
Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count	9 Count

The score that falls precisely in the middle of the range of values is called the median. In the table above we can see that median is higher in December than the other months which is 202,184,12€ and also have the maximum Mean which give us the intuition that sales for this Month is more than others which is reasonable because in December we have the holiday of Christmas. In this period people are ordering more products to enjoy a Christmas table that will give them enjoyment in their families, and they are celebrating with others eating all together homemade foods and desserts.

The standard deviation, which gauges how widely apart the data are from the mean, is another useful statistic. It is helpful for comparing data sets. A high standard deviation suggests that the data are moving more quickly. It also serves as a gauge for the chance that an investment will deviate from the projected rate of return. A smaller standard deviation indicates less volatility, whereas an increased standard deviation indicates greater risk associated with the investment.

4.3 EXPONENTIAL SMOOTHING

4.3.1 Exponentially Weighted Moving Average analysis

I conducted the test for this model by selecting 60% of the data. After running the EWMA model, I determined the value of λ that resulted in a reduced MAPE. So, using that λ , I rerun the EWMA model, this time using the remaining 40% of the data to get the MAPE and determine whether the model fits well enough to produce a forecast.

In order to make a graph and choose the value λ that optimizes model's performance I selected a range of λ . I especially chose prices from 0.01 to 1. Then I calculated MAE & MAPE for each lambda. I ran the model EWMA for each product separately. In order to contract the table that would provide me with all the above-mentioned results I set one column as a month index

starting from 1 and finishing to the last observation in ascending order. Number 1 is not related to the month January but with the relevant month and year where the data starts. Then I created another column where I named it Sales in EUR and I inserted the given data. In the third column I used the type of EWMA as shown below:

$$EWMA(t) = \lambda * EWMA(t-1) + (1-\lambda) * sales$$

The first cell, however, is the exception, where I set it to equal actual sales in EUR. The forecasting error, which is the rule real sales minus EWMA(t) outcomes, was then determined in a different column that I generated and applied to each observation in turn. Next, I made a new column using the Excel application's =ABS(CELL) method to get the absolute error. This was a crucial step since the MAE for the particular λ we chose is determined by averaging the absolute forecasting error.

Next, I located the MAPE, where I split the predicting error by the actual sales in a new column. Lastly, we obtained the MAPE by averaging this column. We made a table for each λ in order to determine which λ would result in the lower MAPE. After selecting this λ , I ran the model on the remaining 40% of the data. Every product that was approved for each month was put through this procedure again. Lastly, I'll give an example from a single category, such as Stationary. (Data that may be supplied upon request is available for the remaining categories.)

I created the following tables as a result:

Table 2: EWMA Model for $\lambda=0.01$ for Stationery

Month Index	Sales in EUR	EWMA (Fast) ($\lambda=0.01$)	0.01	Forecasting Error $\lambda=0.01$	Absolute Forecasting Error	MAE	MAPE
1	199,999.08	199,999.08		0.00			
2	165,270.32	165,617.61		-347.29	347.29	202.98	0.21%
3	190,916.03	190,663.05		252.98	252.98		0.13%
4	184,141.36	184,206.58		-65.22	65.22		0.04%
5	190,814.37	190,748.29		66.08	66.08		0.03%
6	170,606.22	170,807.64		-201.42	201.42		0.12%
7	162,256.47	162,341.98		-85.51	85.51		0.05%
8	153,539.71	153,627.73		-88.02	88.02		0.06%
9	172,294.02	172,107.36		186.66	186.66		0.11%
10	153,539.71	153,725.39		-185.68	185.68		0.12%
11	197,421.27	196,984.31		436.96	436.96		0.22%
12	202,184.12	202,132.12		52.00	52.00		0.03%
13	195,176.14	195,245.70		-69.56	69.56		0.04%
14	201,871.38	201,805.12		66.26	66.26		0.03%
15	289,569.82	288,692.17		877.65	877.65		0.30%
16	177,783.55	178,892.64		-1,109.09	1,109.09		0.62%
17	172,857.58	172,917.93		-60.35	60.35		0.03%
18	174,314.68	174,300.71		13.97	13.97		0.01%
19	174,580.00	174,577.21		2.79	2.79		0.00%
20	151,508.97	151,739.65		-230.68	230.68		0.15%
21	179,724.46	179,444.61		279.85	279.85		0.16%
22	204,035.12	203,789.21		245.91	245.91		0.12%
23	190,338.10	190,472.61		-134.51	134.51		0.07%
24	198,071.02	197,995.04		75.98	75.98		0.04%
25	191,021.99	191,091.72		-69.73	69.73		0.04%
26	195,102.62	195,062.51		40.11	40.11		0.02%
27	210,897.80	210,739.45		158.35	158.35		0.08%
28	218,512.80	218,435.07		77.73	77.73		0.04%
							0.11%

Table 3: Results MAE & MAPE for Stationery

A/A	λ	MAE	MAPE
1	0.01	202.98	0.11%
2	0,10	2,029.75	1.06%
3	0,20	4,018.20	2.10%
4	0,30	6,096.15	3.18%
5	0,40	7,930.26	4.14%
6	0,50	10,098.93	5.27%
7	0,60	12,321.71	6.43%
8	0,70	14,440.56	7.53%
9	0,80	16,575.53	8.66%
10	0,90	18,869.39	9.96%
11	1.00	21,695.51	11.95%

In addition to the table, the graph that I can derive from this data also clearly shows the minimum λ . I created the following graph by choosing every value for λ and the appropriate MAPE.

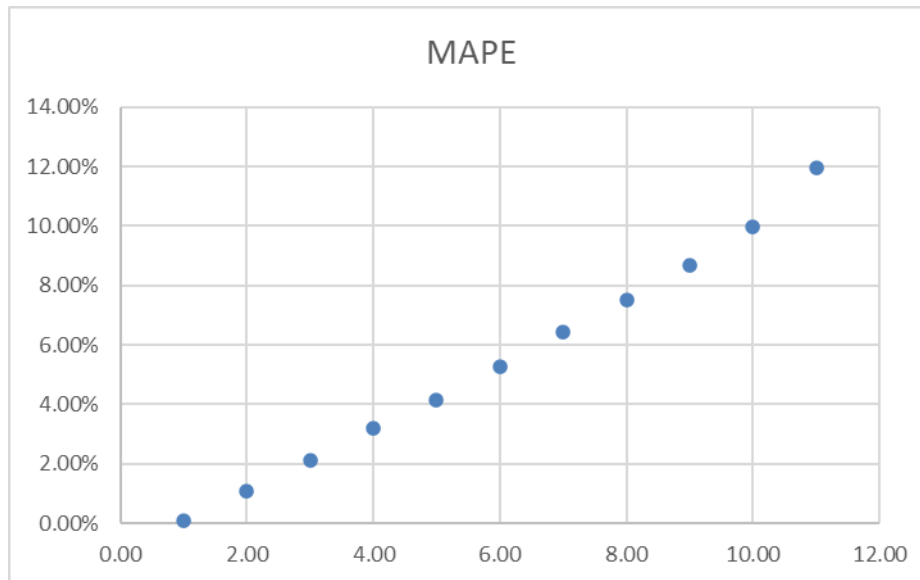


Figure 2: MAPE for Stationery

This model looks as if it is a good fit model. Since EWMA is a recursive function, the prior observation is used to calculate the current observation. To be more precise, we multiply the prior period's projection by λ and then add the actual sales multiplied by $1-\lambda$. (Data that may be supplied upon request is available for the remaining categories.)

Therefore, we used the regression model, which is deemed suitable for identifying patterns and seasonality by statistical analysts and theorists.

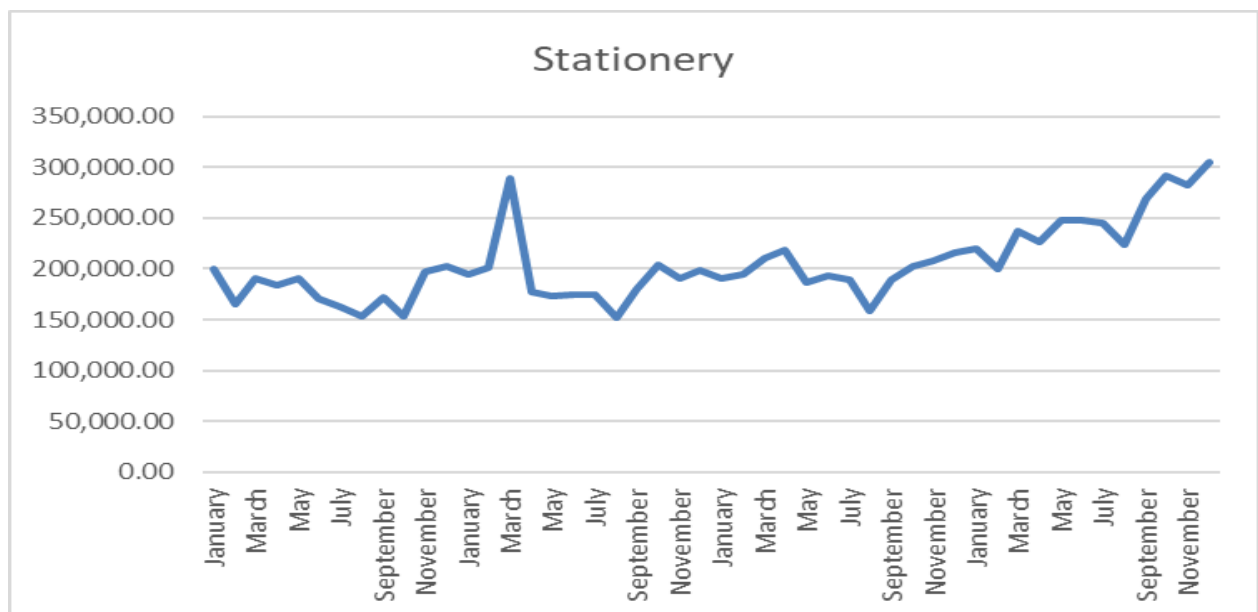
4.3.2 Regression Analysis

Regression is a suitable model for detecting trends and seasonality, as I have indicated. Prior to beginning to build the model, I selected all of the sales from a single product (particularly in this instance, I'm still using Bulgaria as an example) and the pertinent observation index I had made. I arrived at the table below, which has a sample size of $T=48$:

Table 4: Sales for Stationery for years 2019 to 2022

A/A		Month	Stationery (kitchen paper-toilet paper): Value
1	2019	January	199,999.08
2		February	165,270.32
3		March	190,916.03
4		April	184,141.36
5		May	190,814.37
6		June	170,606.22
7		July	162,256.47
8		August	153,539.71
9		September	172,294.02
10		October	153,539.71
11		November	197,421.27
12		December	202,184.12
13	2020	January	195,176.14
14		February	201,871.38
15		March	289,569.82
16		April	177,783.55
17		May	172,857.58
18		June	174,314.68
19		July	174,580.00
20		August	151,508.97
21		September	179,724.46
22		October	204,035.12
23		November	190,338.10
24		December	198,071.02
25	2021	January	191,021.99
26		February	195,102.62
27		March	210,897.80
28		April	218,512.80
29		May	186,166.58
30		June	192,801.94
31		July	189,081.48
32		August	158,372.38
33		September	189,103.38
34		October	201,990.17
35		November	208,517.00
36		December	216,320.43
37	2022	January	220,139.18
38		February	200,321.50
39		March	237,333.16
40		April	226,295.40
41		May	247,640.97
42		June	247,202.49
43		July	244,880.53
44		August	223,329.56
45		September	268,431.12
46		October	292,171.58
47		November	281,877.99
48		December	304,753.66

I selected the suggested charts, particularly the line chart, from the Excel application's ribbon, and I came up with the following graph:



At first glance we can notice that there is no seasonality for these products because they are products that people use it every day. We can see that sales are moving smoothly and could say that are quite predictable.

We can now use the regression model to statistically determine the non-seasonality we saw in the previous graph. Using a sample size of $T=48$ ($48 \times 60\% = 29$ observations), I utilized 60% of the total observations. I made a column called "month index" and used the formula `=MONTH (CELL WITH RELEVANT MONTH)` to compute the amount of data from the Excel application. Next, I added a new column and called it "constant term." I used the number 1 as a constant term for all of the data in this column.

Table 5: Regression With Dummies for Stationery for Years 2019 to 2022

Year	Month	Date	Stationery Sales	Month Index	Constant Term	Linear Trend	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2019	January	Jan-19	199,999.08	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	February	Feb-19	165,270.32	2	1	2	1	0	0	0	0	0	0	0	0	0	0
	March	Mar-19	190,916.03	3	1	3	0	1	0	0	0	0	0	0	0	0	0
	April	Apr-19	184,141.36	4	1	4	0	0	1	0	0	0	0	0	0	0	0
	May	May-19	190,814.37	5	1	5	0	0	0	1	0	0	0	0	0	0	0
	June	Jun-19	170,606.22	6	1	6	0	0	0	0	1	0	0	0	0	0	0
	July	Jul-19	162,256.47	7	1	7	0	0	0	0	0	1	0	0	0	0	0
	August	Aug-19	153,539.71	8	1	8	0	0	0	0	0	0	1	0	0	0	0
	September	Sep-19	172,294.02	9	1	9	0	0	0	0	0	0	0	1	0	0	0
	October	Oct-19	153,539.71	10	1	10	0	0	0	0	0	0	0	0	1	0	0
	November	Nov-19	197,421.27	11	1	11	0	0	0	0	0	0	0	0	0	1	0
	December	Dec-19	202,184.12	12	1	12	0	0	0	0	0	0	0	0	0	0	1
2020	January	Jan-20	195,176.14	1	1	13	0	0	0	0	0	0	0	0	0	0	0
	February	Feb-20	201,871.38	2	1	14	1	0	0	0	0	0	0	0	0	0	0
	March	Mar-20	289,569.82	3	1	15	0	1	0	0	0	0	0	0	0	0	0
	April	Apr-20	177,783.55	4	1	16	0	0	1	0	0	0	0	0	0	0	0
	May	May-20	172,857.58	5	1	17	0	0	0	1	0	0	0	0	0	0	0
	June	Jun-20	174,314.68	6	1	18	0	0	0	0	1	0	0	0	0	0	0
	July	Jul-20	174,580.00	7	1	19	0	0	0	0	0	1	0	0	0	0	0
	August	Aug-20	151,508.97	8	1	20	0	0	0	0	0	0	1	0	0	0	0
	September	Sep-20	179,724.46	9	1	21	0	0	0	0	0	0	0	1	0	0	0
	October	Oct-20	204,035.12	10	1	22	0	0	0	0	0	0	0	0	1	0	0
	November	Nov-20	190,338.10	11	1	23	0	0	0	0	0	0	0	0	0	1	0
	December	Dec-20	198,071.02	12	1	24	0	0	0	0	0	0	0	0	0	0	1
2021	January	Jan-21	191,021.99	1	1	25	0	0	0	0	0	0	0	0	0	0	0
	February	Feb-21	195,102.62	2	1	26	1	0	0	0	0	0	0	0	0	0	0
	March	Mar-21	210,897.80	3	1	27	0	1	0	0	0	0	0	0	0	0	0
	April	Apr-21	218,512.80	4	1	28	0	0	1	0	0	0	0	0	0	0	0
	May	May-21	186,166.58	5	1	29	0	0	0	1	0	0	0	0	0	0	0
	June	Jun-21	192,801.94	6	1	30	0	0	0	0	1	0	0	0	0	0	0
	July	Jul-21	189,081.48	7	1	31	0	0	0	0	0	1	0	0	0	0	0
	August	Aug-21	158,372.38	8	1	32	0	0	0	0	0	0	1	0	0	0	0
	September	Sep-21	189,103.38	9	1	33	0	0	0	0	0	0	0	1	0	0	0
	October	Oct-21	201,990.17	10	1	34	0	0	0	0	0	0	0	0	1	0	0
	November	Nov-21	208,517.00	11	1	35	0	0	0	0	0	0	0	0	0	1	0
	December	Dec-21	216,320.43	12	1	36	0	0	0	0	0	0	0	0	0	0	1
2022	January	Jan-22	220,139.18	1	1	37	0	0	0	0	0	0	0	0	0	0	0
	February	Feb-22	200,321.50	2	1	38	1	0	0	0	0	0	0	0	0	0	0
	March	Mar-22	237,333.16	3	1	39	0	1	0	0	0	0	0	0	0	0	0
	April	Apr-22	226,295.40	4	1	40	0	0	1	0	0	0	0	0	0	0	0
	May	May-22	247,640.97	5	1	41	0	0	0	1	0	0	0	0	0	0	0
	June	Jun-22	247,202.49	6	1	42	0	0	0	0	1	0	0	0	0	0	0
	July	Jul-22	244,880.53	7	1	43	0	0	0	0	0	1	0	0	0	0	0
	August	Aug-22	223,329.56	8	1	44	0	0	0	0	0	0	1	0	0	0	0
	September	Sep-22	268,431.12	9	1	45	0	0	0	0	0	0	0	1	0	0	0
	October	Oct-22	292,171.58	10	1	46	0	0	0	0	0	0	0	0	1	0	0
	November	Nov-22	281,877.99	11	1	47	0	0	0	0	0	0	0	0	0	1	0
	December	Dec-22	304,753.66	12	1	48	0	0	0	0	0	0	0	0	0	0	1

I then performed the regression analysis to arrive at a function that included both independent and dependent variables. I selected the data analysis from the Excel 2016 program, specifically from the tools. I selected the regression tool from the analytical tools, and I set the sales column as the dependent variable and the column's linear trend and all of the dummies as the independent variables. I arrived at the summary output that follows:

Table 6: Summary Output for Stationery

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.757352311							
R Square	0.573582524							
Adjusted R Square	0.20808183							
Standard Error	24033.02011							
Observations	27							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	12	10876912890	906409407.5	1.569306251	0.208908681			
Residual	14	8086204780	577586055.7					
Total	26	18963117671						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Constant Term	165553.8819	16011.90665	10.33942338	6.16846E-08	131211.7577	199896.0062	131211.758	199896.0062
LINEAR TREND	866.9329111	629.0512255	1.378159482	0.189782358	-482.2477833	2216.113605	-482.247783	2216.113605
FEB	9723.830646	19400.7171	0.501209857	0.624011806	-31886.56912	51334.23042	-31886.5691	51334.23042
MAR	51903.34107	19420.17666	2.672650304	0.018205333	10251.20469	93555.47745	10251.2047	93555.47745
APR	12275.97284	18481.49736	0.664230424	0.517333978	-27362.89668	51914.84236	-27362.8967	51914.84236
MAY	6745.831046	21797.57599	0.309476203	0.761517495	-40005.31976	53496.98185	-40005.3198	53496.98185
JUN	-3496.626865	21760.41185	-0.160687532	0.874635122	-50168.06852	43174.81479	-50168.0685	43174.81479
JUL	2667.682982	12925.15547	0.206394653	0.839454115	-25054.01841	30389.38437	-25054.0184	30389.38437
AUG	-25166.60269	21740.565	-1.157587334	0.266400918	-71795.47709	21462.27171	-71795.4771	21462.27171
SEP	-2548.635598	21757.93198	-0.11713593	0.90841618	-49214.75847	44117.48728	-49214.7585	44117.48728
OCT	-637.393509	21793.44977	-0.029247022	0.977080419	-47379.69445	46104.90744	-47379.6945	46104.90744
NOV	13587.94358	21847.02983	0.621958394	0.543965788	-33269.27516	60445.16232	-33269.2752	60445.16232
DEC	18968.89567	21918.5397	0.865426982	0.40139165	-28041.6965	65979.48783	-28041.6965	65979.48783

And the Regression is the following:

$$Y = 165553.88 + 866.93 \cdot t + 9723.83 \cdot \text{FEB} + 51903.34 \cdot \text{MAR} + 12275.97 \cdot \text{APR} + 6745.83 \cdot \text{MAY} + 3496.63 \cdot \text{JUN} + 2667.68 \cdot \text{JUL} + 25166.60 \cdot \text{AUG} + 2548.64 \cdot \text{SEP} + 637.39 \cdot \text{OCT} + 13587.94 \cdot \text{NOV} + 18968.90 \cdot \text{DEC}$$

Table 7: Representation of regression function for Stationery

Constant Term	165553.8819
LINEAR TREND	866.9329111
FEB	9723.830646
MAR	51903.34107
APR	12275.97284
MAY	6745.831046
JUN	-3496.626865
JUL	2667.682982
AUG	-25166.60269
SEP	-2548.635598
OCT	-637.393509
NOV	13587.94358
DEC	18968.89567

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.757352311
R Square	0.573582524
Adjusted R Square	0.20808183
Standard Error	24033.02011
Observations	27

Each piece of information has the following meanings:

Several R. The degree to which two variables have a linear relationship is determined by the correlation coefficient. The correlation coefficient's absolute value represents the strength of the association and can take any value between -1 and 1. The link is stronger the higher the absolute value:

1 denotes a solid, fulfilling connection.

A significant negative association is indicated by a -1.

0 denotes no link whatsoever.

R Square: As a measure of the quality of fit, it is the Coefficient of Determination. It displays the number of points that lie on the regression line. To be more specific, the R^2 value is the

sum of the squared deviations of the original data from the mean, which is derived from the total sum of squares.

R^2 is rather weak in our example, coming up at 0.57 (rounded to two digits). This indicates that the regression analysis model fits 57% of our values. Put otherwise, the independent variables (x-values) account for 57% of the variance in the dependent variables (y-values). R^2 of 57% or more is often regarded as a not so good match.

R Squared adjusted: Adjusted for the number of independent variables in the model, it is the R^2 . This number should be used for multiple regression analysis instead of R^2 .

The standard error: The smaller the value, the more confident you may be in your regression equation. This is another goodness-of-fit metric that illustrates the accuracy of your regression analysis. The Standard Error is an absolute metric that indicates the average distance that the data points fall from the regression line, whereas R^2 indicates the proportion of the dependent variables variation that is explained by the model.

Observations: It is simply the number of observations in our model.

I then added a new column and gave it the label "estimated sales." Using the formula SUBPRODUCT, I was able to get the estimated sales (I selected the regression and selected the pertinent constant term, linear term, and all of the dummies).

I ended up the following table:

Table 8: Estimated Sales from the regression model for Stationery

Month	Date	Stationery Sales	Residuals	Month Index	Constant Term	Linear Trend	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
January	Jan-19	199,999.08	-11,741.26	1	1	1	0	0	0	0	0	0	0	0	0	0	0
February	Feb-19	165,270.32	-29,141.99	2	1	2	1	0	0	0	0	0	0	0	0	0	0
March	Mar-19	190,916.03	2,843.77	3	1	3	0	1	0	0	0	0	0	0	0	0	0
April	Apr-19	184,141.36	14,179.99	4	1	4	0	0	1	0	0	0	0	0	0	0	0
May	May-19	190,814.37	3,347.37	5	1	5	0	0	0	1	0	0	0	0	0	0	0
June	Jun-19	170,606.22	-12,033.63	6	1	6	0	0	0	0	1	0	0	0	0	0	0
July	Jul-19	162,256.47	6,216.97	7	1	7	0	0	0	0	0	1	0	0	0	0	0
August	Aug-19	153,539.71	1,486.38	8	1	8	0	0	0	0	0	0	1	0	0	0	0
September	Sep-19	172,294.02	-20,046.11	9	1	9	0	0	0	0	0	0	0	1	0	0	0
October	Oct-19	153,539.71	8,743.18	10	1	10	0	0	0	0	0	0	0	0	1	0	0
November	Nov-19	197,421.27	7,258.15	11	1	11	0	0	0	0	0	0	0	0	0	1	0
December	Dec-19	202,184.12	18,352.13	12	1	12	0	0	0	0	0	0	0	0	0	0	1
January	Jan-20	195,176.14	14,456.61	1	1	13	0	0	0	0	0	0	0	0	0	0	0
February	Feb-20	201,871.38	59,108.60	2	1	14	1	0	0	0	0	0	0	0	0	0	0
March	Mar-20	289,569.82	-13,917.23	3	1	15	0	1	0	0	0	0	0	0	0	0	0
April	Apr-20	177,783.55	-14,179.99	4	1	16	0	0	1	0	0	0	0	0	0	0	0
May	May-20	172,857.58	-3,347.37	5	1	17	0	0	0	1	0	0	0	0	0	0	0
June	Jun-20	174,314.68	-10,113.29	6	1	18	0	0	0	0	1	0	0	0	0	0	0
July	Jul-20	174,580.00	-6,216.97	7	1	19	0	0	0	0	0	1	0	0	0	0	0
August	Aug-20	151,508.97	-1,486.38	8	1	20	0	0	0	0	0	0	1	0	0	0	0
September	Sep-20	179,724.46	20,046.11	9	1	21	0	0	0	0	0	0	0	1	0	0	0
October	Oct-20	204,035.12	-8,743.18	10	1	22	0	0	0	0	0	0	0	0	1	0	0
November	Nov-20	190,338.10	-7,258.15	11	1	23	0	0	0	0	0	0	0	0	0	1	0
December	Dec-20	198,071.02	3,794.79	12	1	24	0	0	0	0	0	0	0	0	0	0	1
January	Jan-21	191,021.99	-2,715.35	1	1	25	0	0	0	0	0	0	0	0	0	0	0
February	Feb-21	195,102.62	-29,966.61	2	1	26	1	0	0	0	0	0	0	0	0	0	0
March	Mar-21	210,897.80	11,073.46	3	1	27	0	1	0	0	0	0	0	0	0	0	0

I then added a new column and called it "Residuals." After deducting the projected sales from the actual sales, I was left with the following table:

Table 9: Residuals from the regression model for Stationery

Month	Date	Stationery Sales	Estimated Sales	Residuals	Month Index	Constant Term	Linear Trend	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
January	Jan-19	199,999.08	177,011.58	-11,741.26	1	1	1	0	0	0	0	0	0	0	0	0	0	0
February	Feb-19	165,270.32	220,058.02	-29,141.99	2	1	2	1	0	0	0	0	0	0	0	0	0	0
March	Mar-19	190,916.03	181,297.59	2,843.77	3	1	3	0	1	0	0	0	0	0	0	0	0	0
April	Apr-19	184,141.36	176,634.38	14,179.99	4	1	4	0	0	1	0	0	0	0	0	0	0	0
May	May-19	190,814.37	167,258.85	3,347.37	5	1	5	0	0	0	1	0	0	0	0	0	0	0
June	Jun-19	170,606.22	174,290.10	-12,033.63	6	1	6	0	0	0	0	1	0	0	0	0	0	0
July	Jul-19	162,256.47	147,322.74	6,216.97	7	1	7	0	0	0	0	0	1	0	0	0	0	0
August	Aug-19	153,539.71	170,807.64	1,486.38	8	1	8	0	0	0	0	0	0	1	0	0	0	0
September	Sep-19	172,294.02	173,585.82	-20,046.11	9	1	9	0	0	0	0	0	0	0	1	0	0	0
October	Oct-19	153,539.71	188,678.09	8,743.18	10	1	10	0	0	0	0	0	0	0	0	1	0	0
November	Nov-19	197,421.27	194,925.97	7,258.15	11	1	11	0	0	0	0	0	0	0	0	0	1	0
December	Dec-19	202,184.12	176,824.01	18,352.13	12	1	12	0	0	0	0	0	0	0	0	0	0	1
January	Jan-20	195,176.14	187,414.77	14,456.61	1	1	13	0	0	0	0	0	0	0	0	0	0	0
February	Feb-20	201,871.38	230,461.22	59,108.60	2	1	14	1	0	0	0	0	0	0	0	0	0	0
March	Mar-20	289,569.82	191,700.78	-13,917.23	3	1	15	0	1	0	0	0	0	0	0	0	0	0
April	Apr-20	177,783.55	187,037.57	-14,179.99	4	1	16	0	0	1	0	0	0	0	0	0	0	0
May	May-20	172,857.58	177,662.05	-3,347.37	5	1	17	0	0	0	1	0	0	0	0	0	0	0
June	Jun-20	174,314.68	184,693.29	-10,113.29	6	1	18	0	0	0	0	1	0	0	0	0	0	0
July	Jul-20	174,580.00	157,725.94	-6,216.97	7	1	19	0	0	0	0	0	1	0	0	0	0	0
August	Aug-20	151,508.97	181,210.84	-1,486.38	8	1	20	0	0	0	0	0	0	1	0	0	0	0
September	Sep-20	179,724.46	183,989.01	20,046.11	9	1	21	0	0	0	0	0	0	0	1	0	0	0
October	Oct-20	204,035.12	199,081.28	-8,743.18	10	1	22	0	0	0	0	0	0	0	0	1	0	0
November	Nov-20	190,338.10	205,329.17	-7,258.15	11	1	23	0	0	0	0	0	0	0	0	0	1	0
December	Dec-20	198,071.02	187,227.20	3,794.79	12	1	24	0	0	0	0	0	0	0	0	0	0	1
January	Jan-21	191,021.99	197,817.97	-2,715.35	1	1	25	0	0	0	0	0	0	0	0	0	0	0
February	Feb-21	195,102.62	240,864.41	-29,966.61	2	1	26	1	0	0	0	0	0	0	0	0	0	0
March	Mar-21	210,897.80	207,439.34	11,073.46	3	1	27	0	1	0	0	0	0	0	0	0	0	0

I came at the two graphs below based on the tables above. I selected the estimated and actual sales for each of the 27 observations to represent the first graph, and I selected the line charts from the Excel application's ribbon. I produced the graph below at the end:

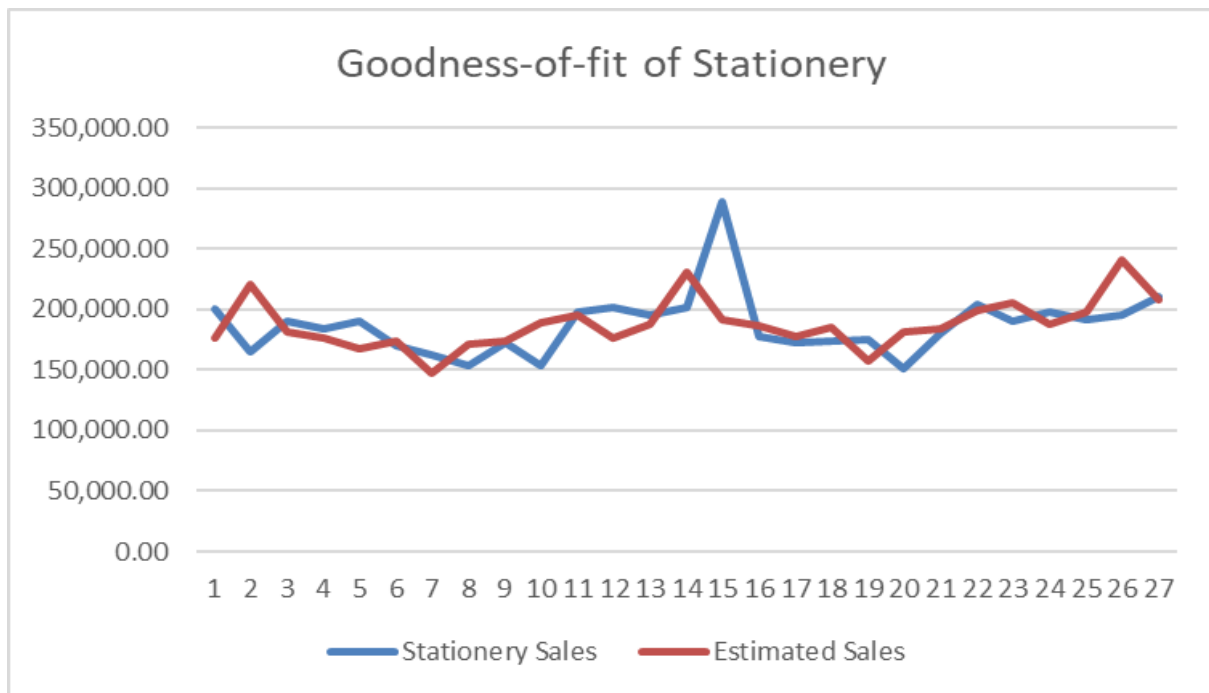


Figure 3: Goodness of fit for Stationery

Upon initial observation, it appears that projected sales correspond partially with actual sales. Furthermore, the fact that the pattern alterations stay consistent makes the seasonality virtually certainly present.

I selected the column residuals of all 27 data to depict the second graph, and we selected line charts from the ribbon in the Excel program. Actually, the residual plot is usually used to validate the model and identify regression issues. I produced the graph below at the end:

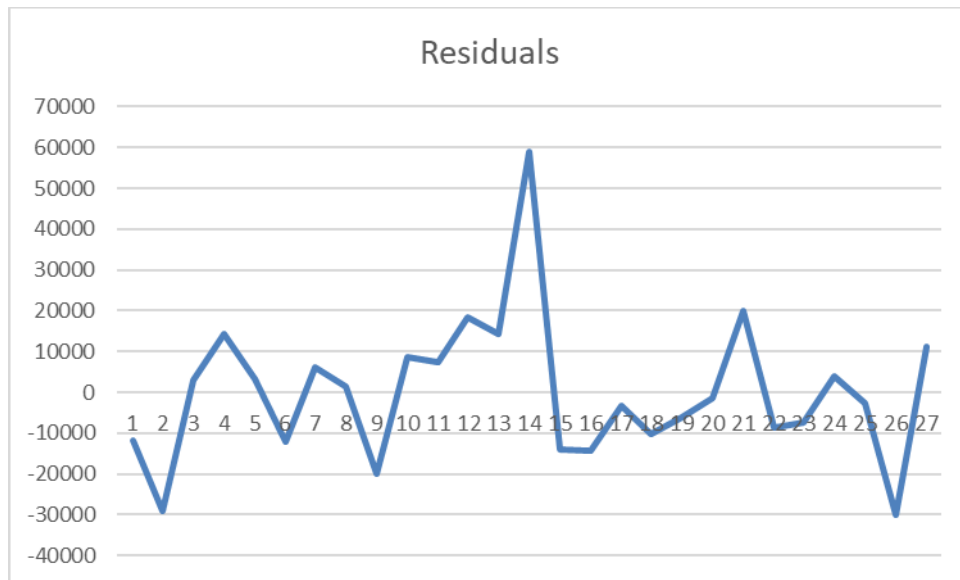


Figure 4: Residuals for Stationery

A rather arbitrary pattern may be seen in the residual graphic. It is initially evident that the residuals are negative, followed by positive, negative, and finally positive. This erratic pattern suggests that the data are somewhat well-fitted by a linear model.

I made a new column that I termed "Time trend component" in order to identify patterns. I utilized the function =SUMPRODUCT (the constant term and linear trend from the regression I discovered initially) from the Excel program to locate it. I completed it for 27 observations of data, and the result was the following table:

Table 10: Time trend component for Stationery

Month	Date	Stationery Sales	Month Index	Estimated Sales	Constant Term	Linear Trend	Time Trend
January	Jan-19	199,999.08	1	177,011.58	1	1	420,477.66
February	Feb-19	165,270.32	2	220,058.02	1	2	428,827.34
March	Mar-19	190,916.03	3	181,297.59	1	3	415,741.62
April	Apr-19	184,141.36	4	176,634.38	1	4	404,335.74
May	May-19	190,814.37	5	167,258.85	1	5	401,664.22
June	Jun-19	170,606.22	6	174,290.10	1	6	388,519.32
July	Jul-19	162,256.47	7	147,322.74	1	7	353,233.21
August	Aug-19	153,539.71	8	170,807.64	1	8	368,033.35
September	Sep-19	172,294.02	9	173,585.82	1	9	389,597.84
October	Oct-19	153,539.71	10	188,678.09	1	10	385,966.80
November	Nov-19	197,421.27	11	194,925.97	1	11	436,128.24
December	Dec-19	202,184.12	12	176,824.01	1	12	422,820.13
January	Jan-20	195,176.14	1	187,414.77	1	13	426,422.91
February	Feb-20	201,871.38	2	230,461.22	1	14	476,196.60
March	Mar-20	289,569.82	3	191,700.78	1	15	525,164.60
April	Apr-20	177,783.55	4	187,037.57	1	16	408,747.12
May	May-20	172,857.58	5	177,662.05	1	17	394,476.63
June	Jun-20	174,314.68	6	184,693.29	1	18	402,996.97
July	Jul-20	174,580.00	7	157,725.94	1	19	376,325.94
August	Aug-20	151,508.97	8	181,210.84	1	20	376,771.81
September	Sep-20	179,724.46	9	183,989.01	1	21	407,797.47
October	Oct-20	204,035.12	10	199,081.28	1	22	447,231.40
November	Nov-20	190,338.10	11	205,329.17	1	23	439,814.27
December	Dec-20	198,071.02	12	187,227.20	1	24	429,476.22
January	Jan-21	191,021.99	1	197,817.97	1	25	433,037.96
February	Feb-21	195,102.62	2	240,864.41	1	26	480,197.03
March	Mar-21	210,897.80	3	207,439.34	1	27	462,596.14

I came to the following graph based on the table mentioned above. We selected the actual sales and time trend component of each of the 27 observations, as well as the line charts from the Excel 2016 application's ribbon, to depict this graph. I produced the graph below at the end:

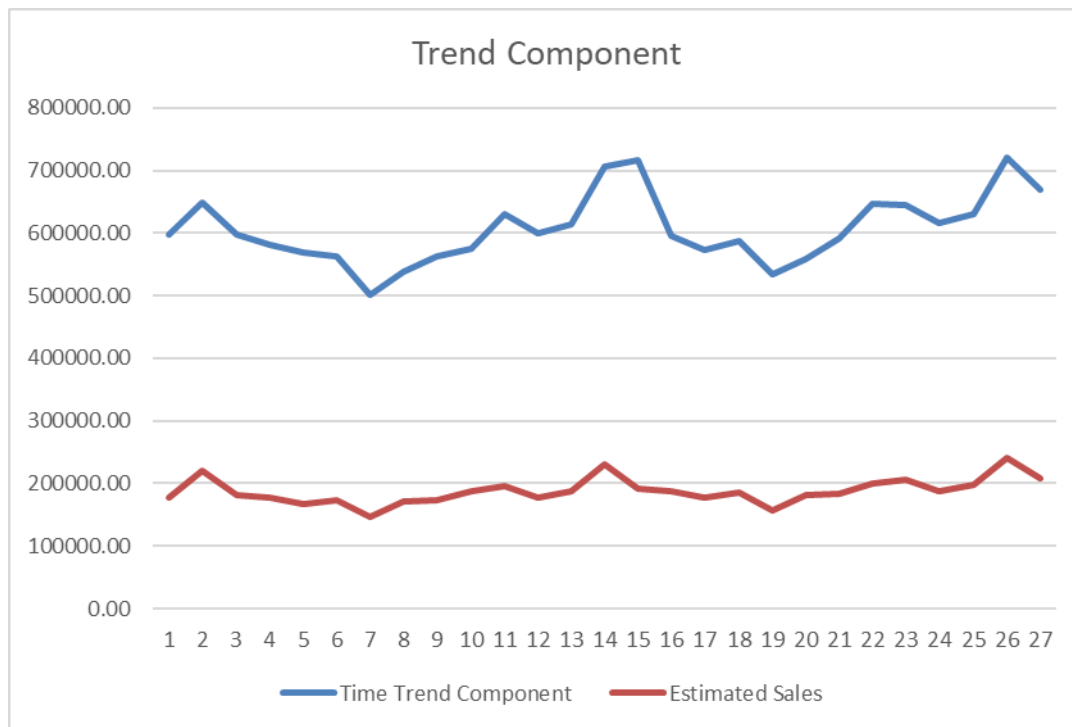


Figure 4: Time Trend Component for Stationery 2019-2021

You can forecast possible future developments for the market by using trends. Because of the pattern's up and down slope, we can tell that it is an unstable trend. Data grows along with time.

As a result, I added a new column and called it "Seasonal Component." I used the type estimated sales minus time trend component to find the values, and the result was the following table:

Table 11: Seasonal Component for Stationery

Estimated Sales	Time Trend	Constant Term	Linear Trend	Time Trend Component	Seasonal Component
177011.58	420,477.66	1	1	597491.24	-420479.66
220058.02	428,827.34	1	2	648888.36	-428830.34
181297.59	415,741.62	1	3	597043.20	-415745.62
176634.38	404,335.74	1	4	580975.12	-404340.74
167258.85	401,664.22	1	5	568929.08	-401670.22
174290.10	388,519.32	1	6	562816.41	-388526.32
147322.74	353,233.21	1	7	500563.96	-353241.21
170807.64	368,033.35	1	8	538850.00	-368042.35
173585.82	389,597.84	1	9	563193.66	-389607.84
188678.09	385,966.80	1	10	574655.89	-385977.80
194925.97	436,128.24	1	11	631066.22	-436140.24
176824.01	422,820.13	1	12	599657.14	-422833.13
187414.77	426,422.91	1	13	613851.69	-426436.91
230461.22	476,196.60	1	14	706672.81	-476211.60
191700.78	525,164.60	1	15	716881.38	-525180.60
187037.57	408,747.12	1	16	595801.69	-408764.12
177662.05	394,476.63	1	17	572156.67	-394494.63
184693.29	402,996.97	1	18	587709.26	-403015.97
157725.94	376,325.94	1	19	534071.87	-376345.94
181210.84	376,771.81	1	20	558003.64	-376792.81
183989.01	407,797.47	1	21	591808.48	-407819.47
199081.28	447,231.40	1	22	646335.68	-447254.40
205329.17	439,814.27	1	23	645167.43	-439838.27
187227.20	429,476.22	1	24	616728.43	-429501.22
197817.97	433,037.96	1	25	630881.93	-433063.96
240864.41	480,197.03	1	26	721088.44	-480224.03
207439.34	462,596.14	1	27	670063.48	-462624.14

I arrived at the following graph from the table above. I selected the actual sales and seasonal component of each of the 48 observations, as well as the line charts from the Excel 2016 application's interface, to depict this graph:

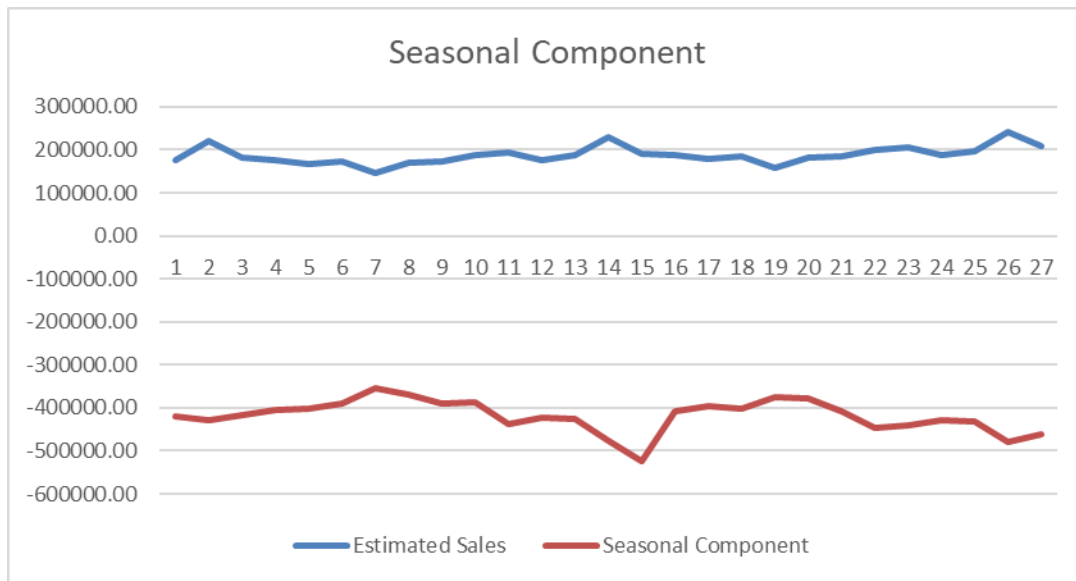


Figure 5: Seasonal Component for Stationery 2019-2021

Seasonality of a time series is the term used to describe the seasonal component. It is one of the variations that shows roughly consistent variations in timing, direction, and size throughout time. We may infer that there is a partial seasonality from this graph.

I followed the same steps and performed the first regression, but this time I used the final 19 of the 48 observations. I arrived at the table that follows:

Table 12: Forecasted Sales for regression model for Stationery

May	May-21	186,166.58	5	184,919.49
June	Jun-21	192,801.94	6	188,017.93
July	Jul-21	189,081.48	7	184,996.72
August	Aug-21	158,372.38	8	158,866.68
September	Sep-21	189,103.38	9	196,782.96
October	Oct-21	201,990.17	10	215,096.59
November	Nov-21	208,517.00	11	213,213.21
December	Dec-21	216,320.43	12	228,552.76
January	Jan-22	220,139.18	1	220,139.18
February	Feb-22	200,321.50	2	200,321.50
March	Mar-22	237,333.16	3	237,333.16
April	Apr-22	226,295.40	4	254,388.39
May	May-22	247,640.97	5	248,888.06
June	Jun-22	247,202.49	6	251,986.50
July	Jul-22	244,880.53	7	248,965.29
August	Aug-22	223,329.56	8	222,835.26
September	Sep-22	268,431.12	9	260,751.54
October	Oct-22	292,171.58	10	279,065.16
November	Nov-22	281,877.99	11	277,181.78
December	Dec-22	304,753.66	12	292,521.33

Next, I made a column called "forecasting error," and in order to detect them, I subtracted the projected sales from the actual sales. I made an additional column and gave it the name "percentage forecasting error." I calculated the absolute average and discovered MAE from the forecasting error and the absolute average and MAPE from the percentage forecasting error. I arrived at the table that follows:

Table 13: MAE-MAPE from regression model for Stationery

Month	Date	Stationery Sales	Month Index	Estimated Sales	Forecasting Error
May	May-21	186,166.58	5	184,919.49	-1,247.09
June	Jun-21	192,801.94	6	188,017.93	-4,784.01
July	Jul-21	189,081.48	7	184,996.72	-4,084.76
August	Aug-21	158,372.38	8	158,866.68	494.30
September	Sep-21	189,103.38	9	196,782.96	7,679.58
October	Oct-21	201,990.17	10	215,096.59	13,106.42
November	Nov-21	208,517.00	11	213,213.21	4,696.21
December	Dec-21	216,320.43	12	228,552.76	12,232.33
January	Jan-22	220,139.18	1	220,139.18	0.00
February	Feb-22	200,321.50	2	200,321.50	0.00
March	Mar-22	237,333.16	3	237,333.16	0.00
April	Apr-22	226,295.40	4	254,388.39	28,092.99
May	May-22	247,640.97	5	248,888.06	1,247.09
June	Jun-22	247,202.49	6	251,986.50	4,784.01
July	Jul-22	244,880.53	7	248,965.29	4,084.76
August	Aug-22	223,329.56	8	222,835.26	-494.30
September	Sep-22	268,431.12	9	260,751.54	-7,679.58
October	Oct-22	292,171.58	10	279,065.16	-13,106.42
November	Nov-22	281,877.99	11	277,181.78	-4,696.21
December	Dec-22	304,753.66	12	292,521.33	-12,232.33

MAE	6237.12	
	MAPE	2.69%

Εδώ άλλαξα το πινακάκι MAE-MAPE γιατί είχε λάθος υπολογισμό.

Using the Excel application, I selected the "actual sales" and "forecasted sales" from the table above to create the following graph:

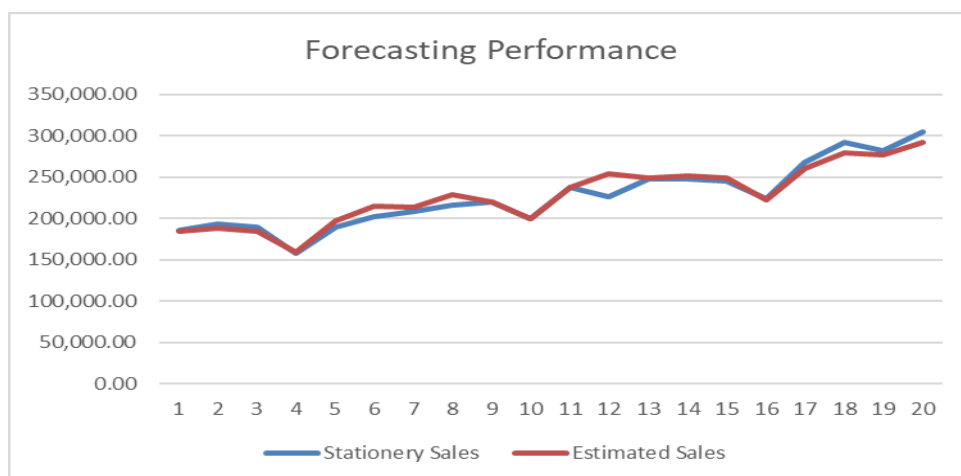


Figure 6: Forecasting Performance for Stationery for 2021-2022

We can infer from the preceding chart that this is an assignment with promise. Next, in order to compare and determine which of the two models I ran was the most effective, I compiled all of the MAPE from each, and the result is the table below:

Table 14: MAE-MAPE results from EWMA &Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	167.16	0.07%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	202.98	0.11%
REGRESSION MODEL			
MAE	MAPE		
6237.12	2.69%		

Και [το](#) [πινακάκι](#) [εδώ](#) [άλλαξε](#)

From the table above we can see that MAPE from EWMA Model of forecasted Sales and forecast from Regression Model of Forecasted Sales we come to the decision that the forecasted model has the smallest MAPE so its regard as better forecasting device.

CHAPTER 5: MAIN FINDINGS AND ITS IMPACT ON THEORY

In this analysis EWMA and Regression models are developed and used to forecast the sales of some categories of products before, during and after the Covid-19 period. Both of these models generate good forecasts from the available sample data. However, the Regression model has other advantages over the EWMA model as results:

- (i) It helps to forecast immediately and decided intuitions,
- (ii) The results of the Regression analysis offer thorough and centered perceptions,
- (iii) When the concentrate of interest is on the higher or lower sales the model can assist the management to make precise and suitable decisions and
- (iv) Finally, we can identify trends and seasonality.

From our investigation we determined that some products had been affected from the Covid-19 in a greater scale than others, because people were afraid for the unknown situation of this period.

5.1 Antiseptics

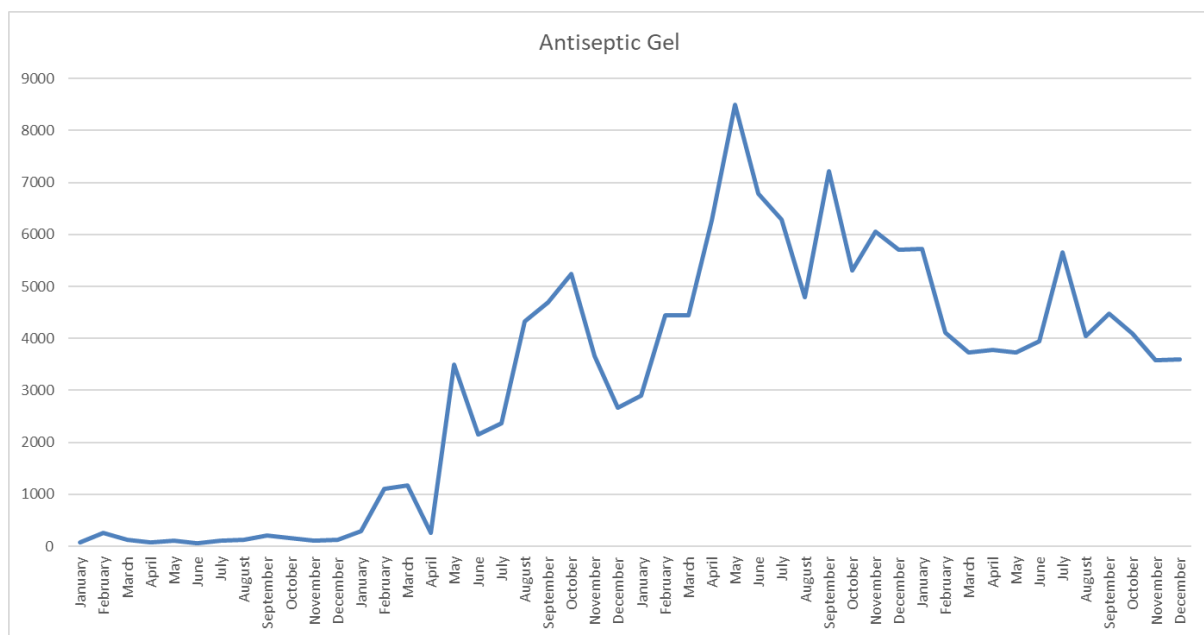


Figure 9: Antiseptics Sales 2019-2022

We can see that there is no seasonality in this category because the sales depend on the needs of the people to protect themselves from COVID-19 advance. We can see that sales increased at the beginning of COVID-19 March of 2020 and keep an uptrend with some downs during the period until May 2021 and then we have a down trend.

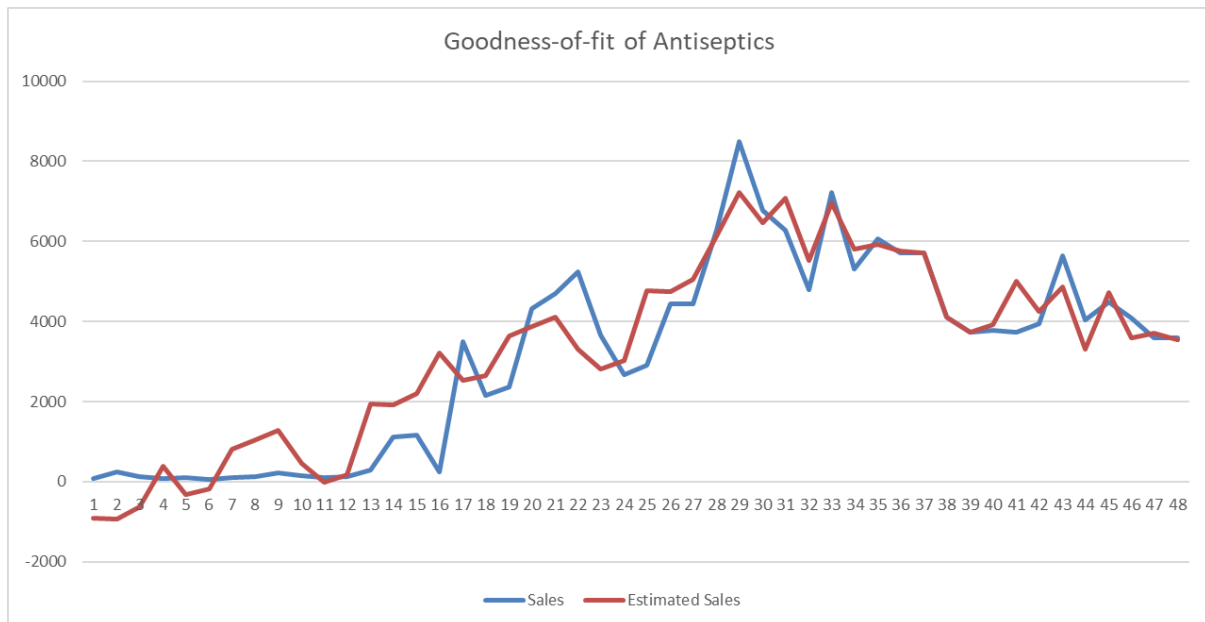


Figure 10: Goodness of fit for Antiseptics

We identify that estimated sales are close Sales for Antiseptics which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

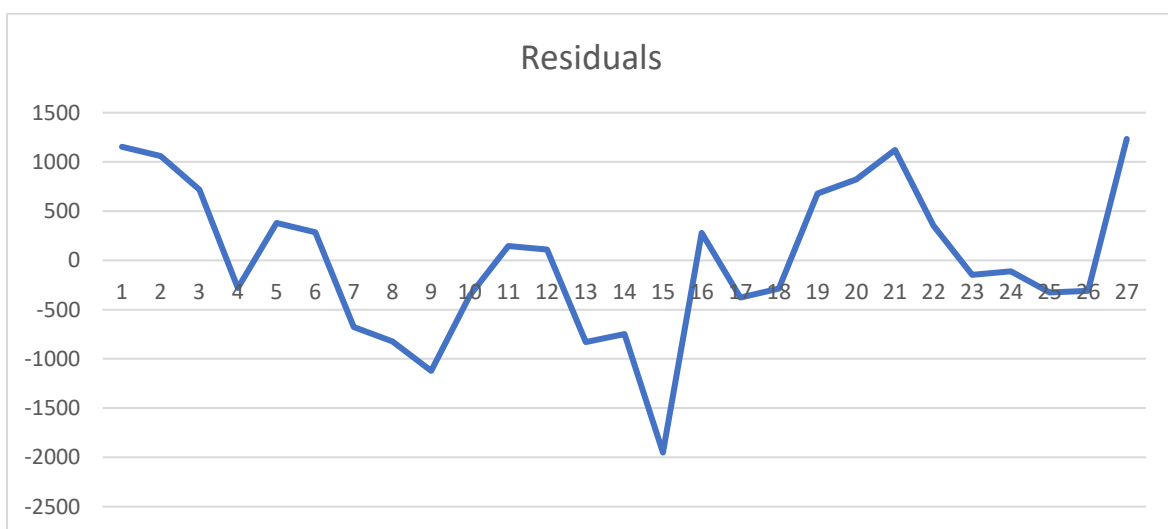


Figure 11: Residuals for Antiseptics

The residual diagram shows a moderately random form. We can see that first residuals are positive, then negative, then zero, then positive, then negative and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

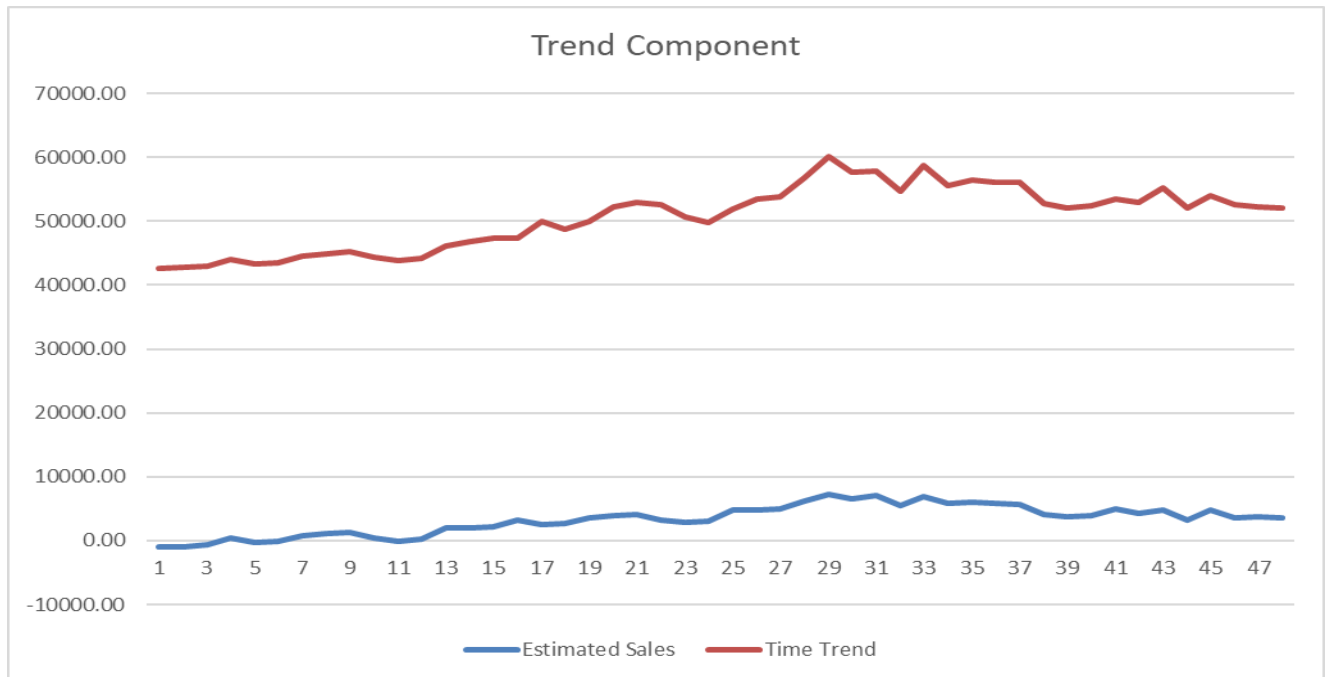


Figure 12: Time Trend Component for Antiseptics 2019-2022

We can imply that there is a correlation for sales in this graph. According to 'time trend' as times by we can see that's sales increases also in a small amount.

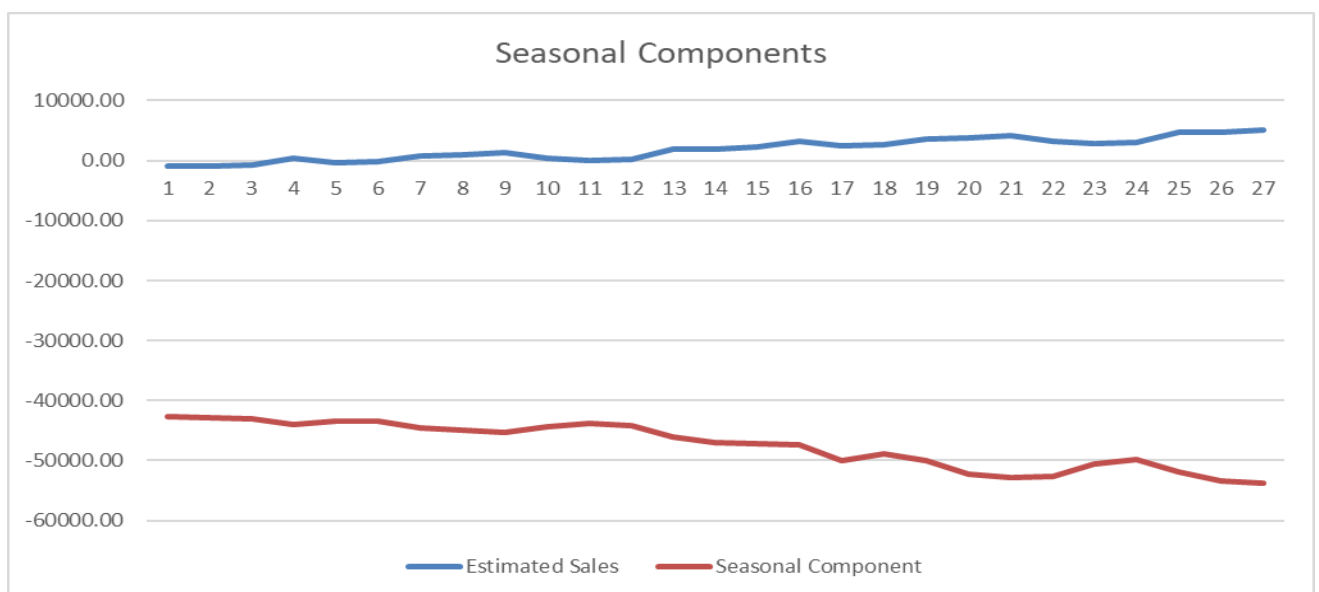


Figure 13: Seasonal Component for Antiseptics 2019-2022

We can imply that there is partial seasonality from this graph. The term "seasonality" refers to a time series' seasonal component, which is one of the variations that exhibits largely constant changes in timing, direction, and size throughout time (Moon et al., 2018).

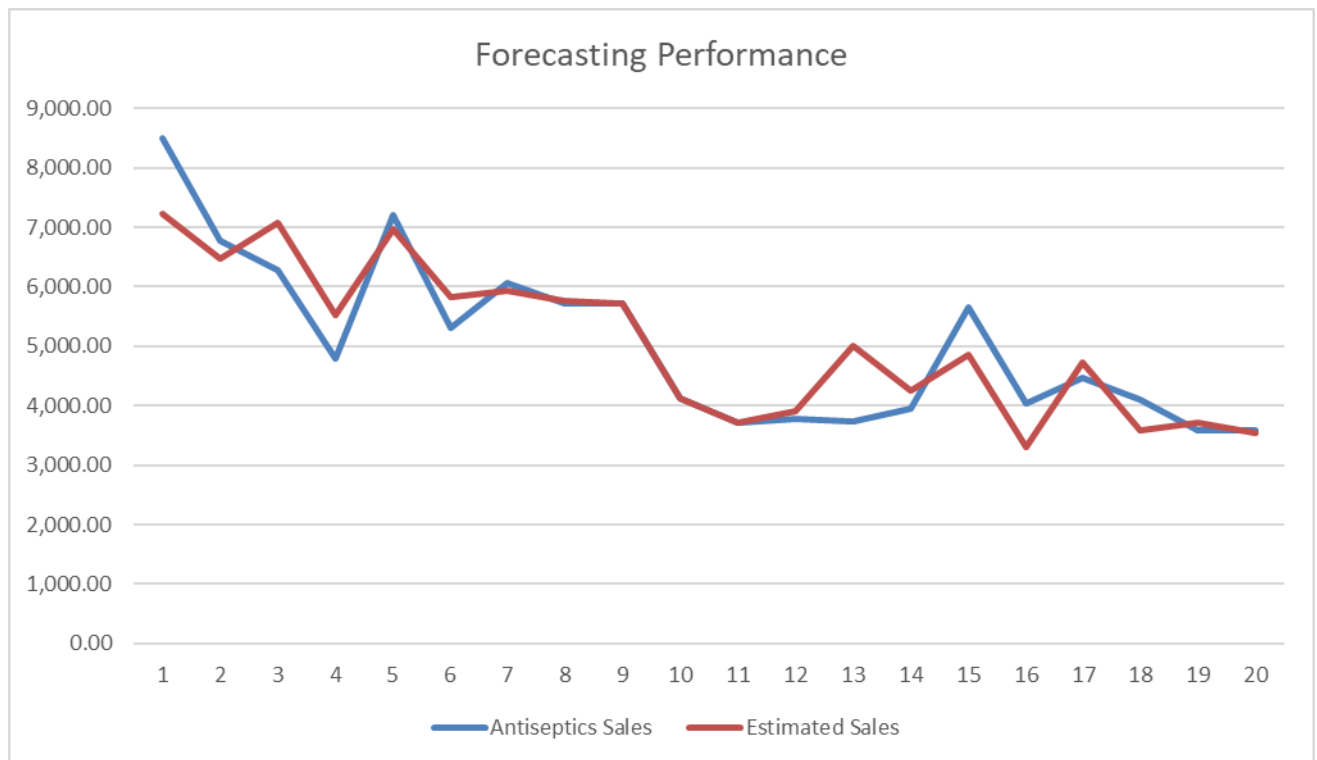


Figure 14: Forecasting Performance for Antiseptics for 2019-2022

We can reckon that the prior shows us chart that the actual sales are close to the estimated sales we calculated and present a down trend as Covid-19 outbreak passes.

In order to compare and determine which model is more effective than the other we performed all of the MAPE from each, and the results are contained in Table 15.

Table 15: MAE-MAPE results from EWMA & Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	8.45	0.16%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	6.11	0.51%
REGRESSION MODEL			
MAE	MAPE		
410.96	8.20%		

Taking into regard the MAPE from the EWMA models, and the MAPE from the regression model, we come to the assumption that the EWMA model can be considered as a better forecasting device, as it has the smaller MAPE, concerning the Antiseptics products category.

5.2: Chlorinated Cleaners

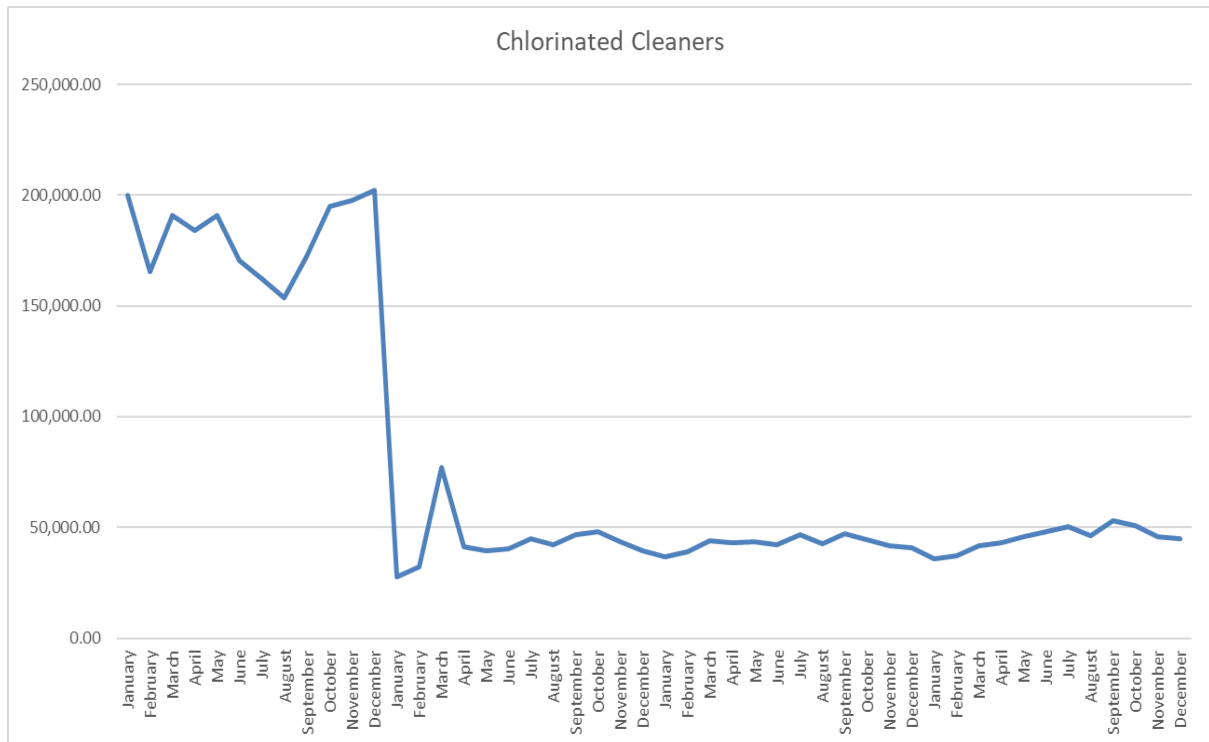


Figure 15: Chlorinated Cleaners sales 2019-2022

As we can tell from the graph there is no seasonality for these products as they consist of products that during the Covid-19 outbreak had an enormous demand cause helped people clean their houses and other surfaces in order to protect themselves from diseases. When people got vaccinated or felt more protected from diseases the sales dropped dramatically and remained in a somehow flat trend.

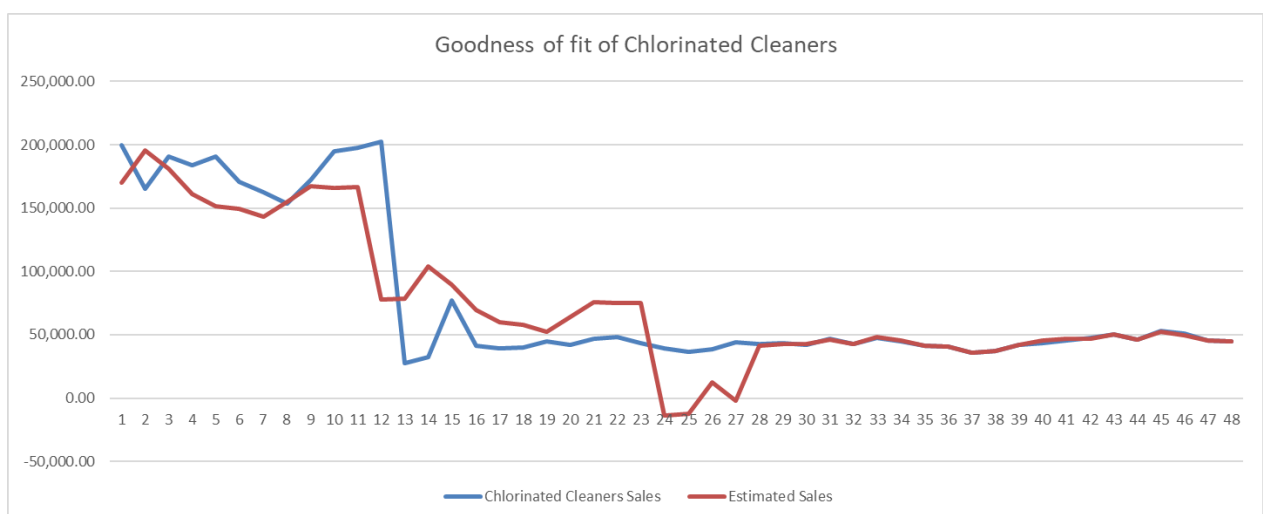


Figure 16: Goodness of fit for Chlorinated Cleaners

We identify that estimated sales are close enough to Sales for Chlorinated Cleaners, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

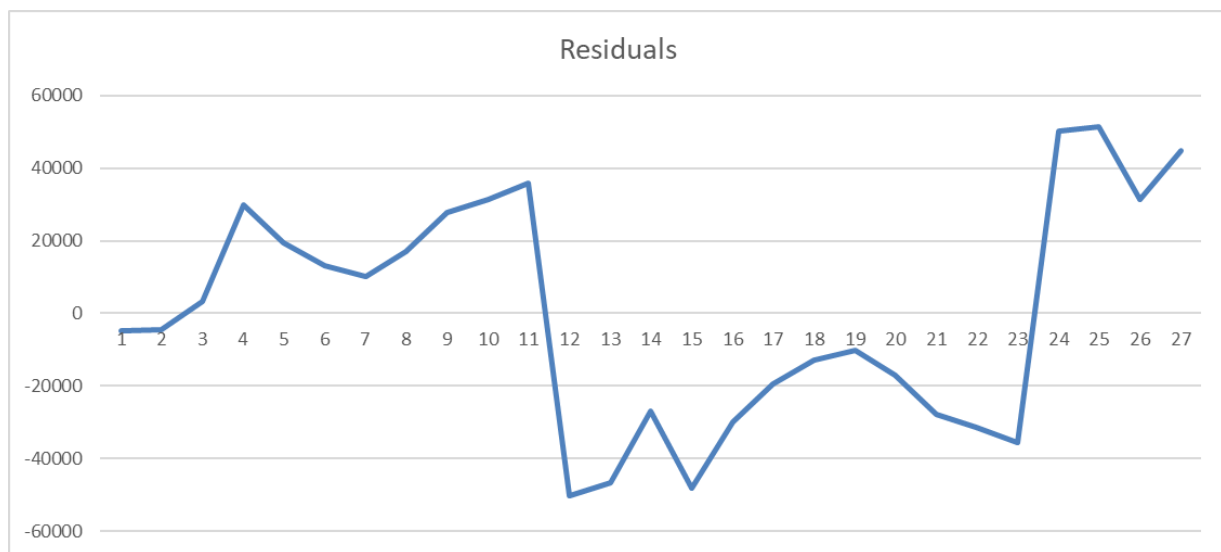


Figure 17: Residuals for Chlorinated Cleaners

The residual diagram shows a reasonably random structure. We can see that first residuals are zero, then positive, then negative, then positive, and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

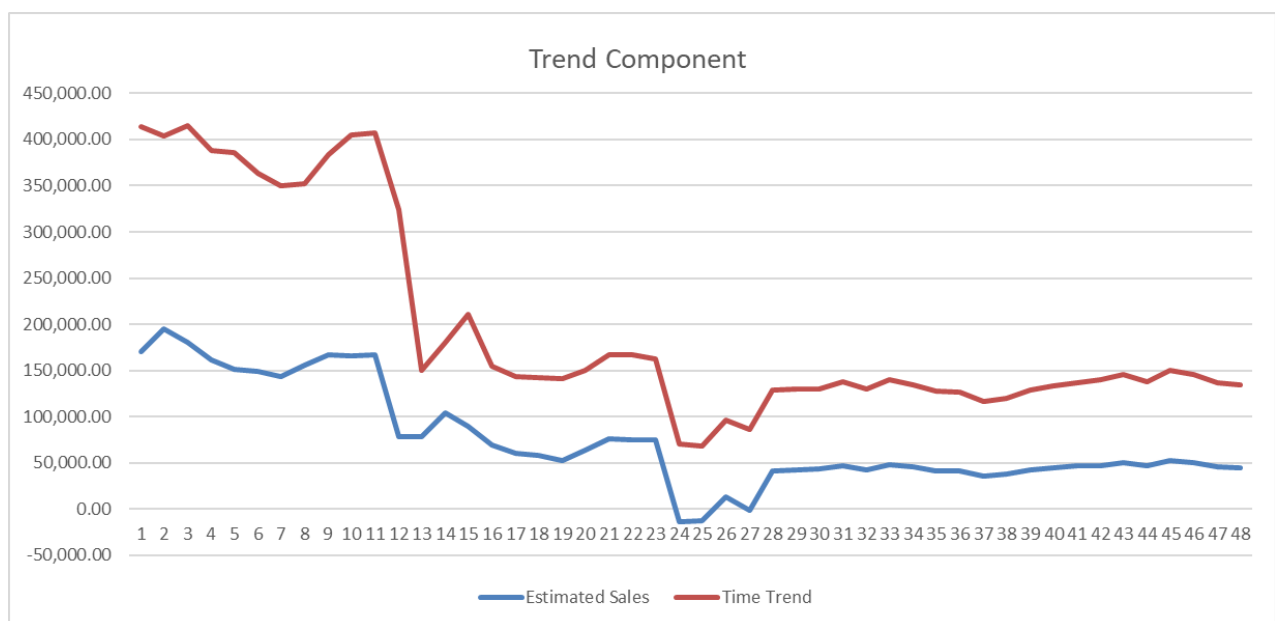


Figure 18: Time Trend Component for Chlorinated Cleaners 2019-2021

We can imply that there is a major correlation for sales in this graph. According to 'time trend' as time by we can see that's sales decreasing until a certain time and then stays in a steady number of sales.

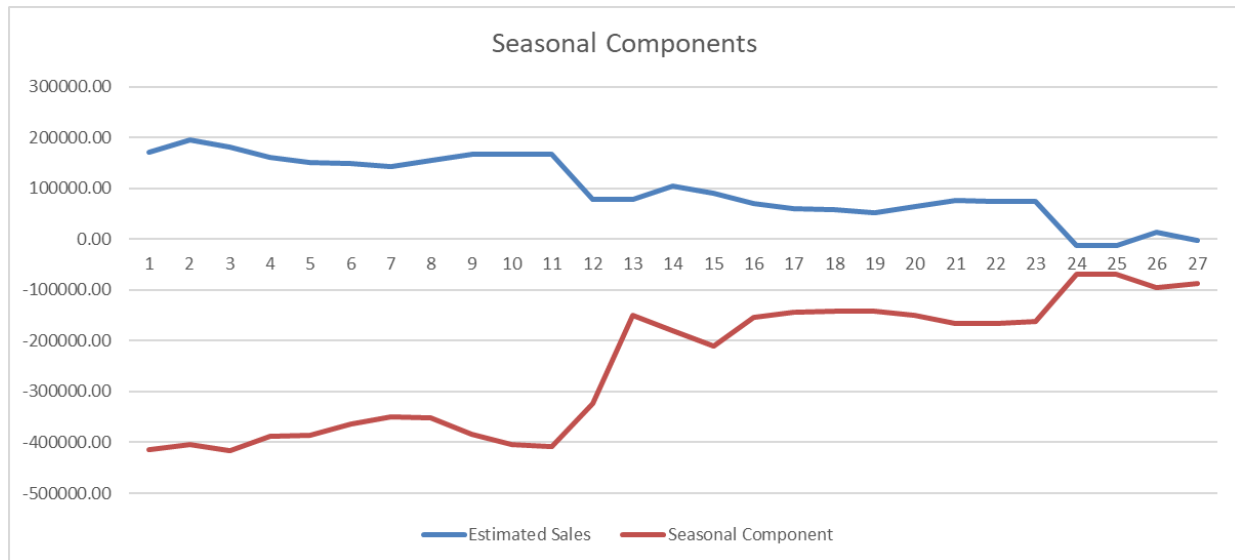


Figure 19: Seasonal Component of Chlorinated Cleaners 2019-2021

We can indicate that there is seasonality from this graph. As we can see, as time goes by the sales drop until they reach to a certain quantity.

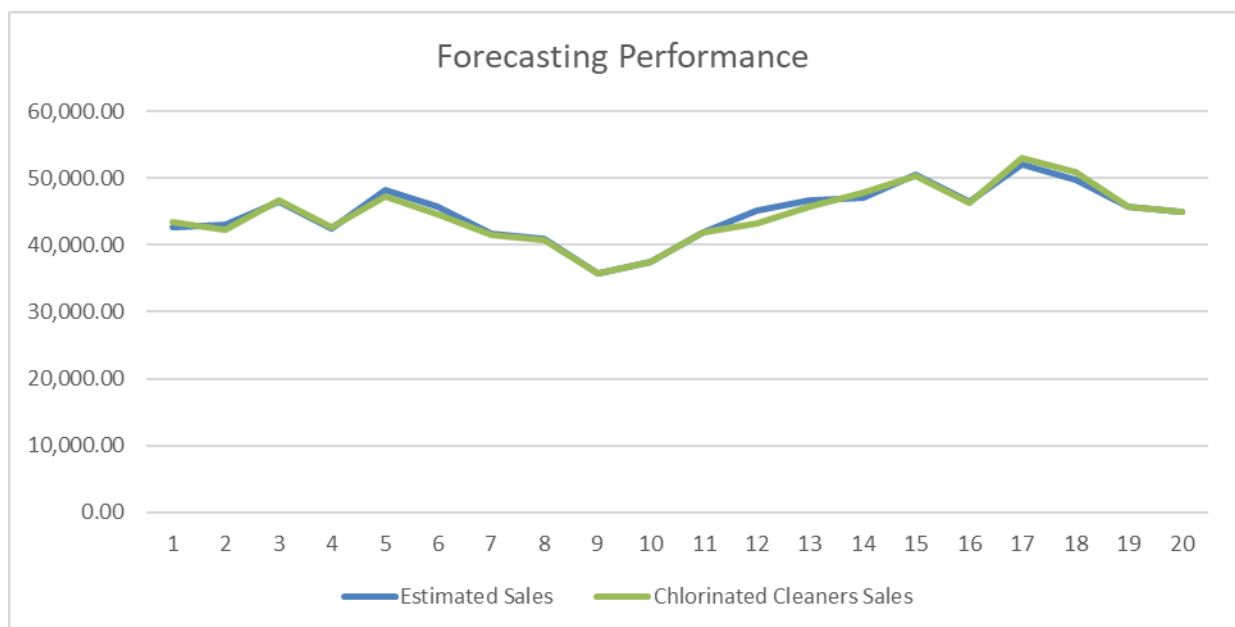


Figure 10: Forecasting Performance for Chlorinated Cleaners for 2021-2022

We can conclude from the above chart that this is quite a promising discovery. To compare and decide which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 16.

Table 16: MAE-MAPE results from EWMA & Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	11.86	0.07%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	167.95	0.36%
REGRESSION MODEL			
MAE	MAPE		
527.71	1.15%		

As we can notice, when taking into regard the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Chlorinated Cleaners products category.

5.3: Alcohol & Alcoholic Lotion

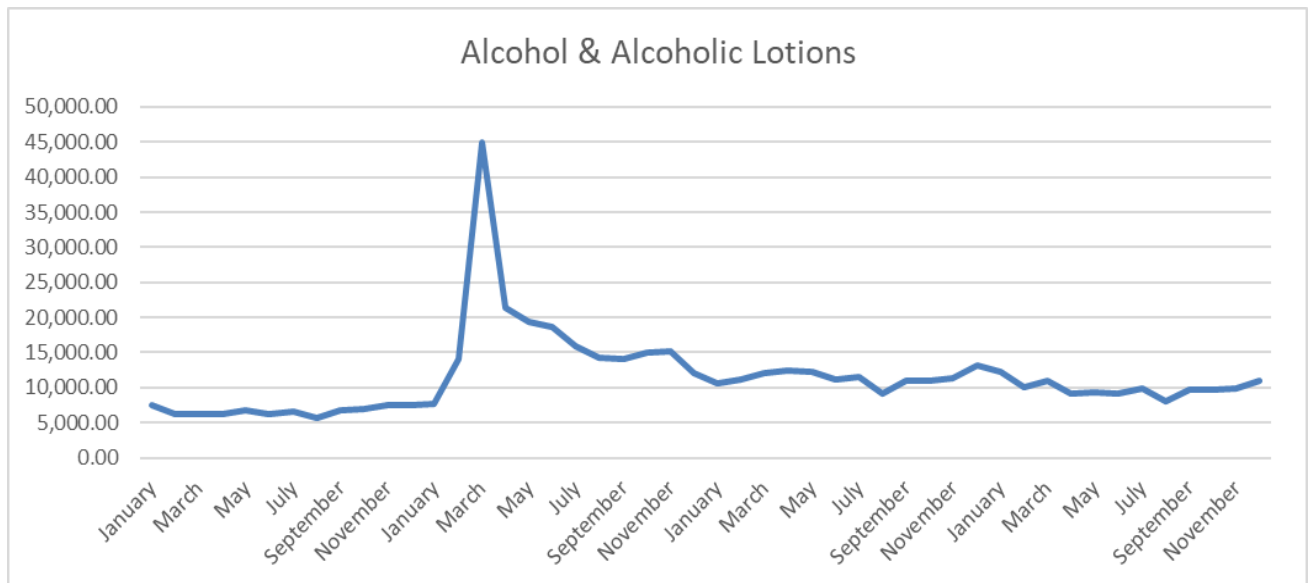


Figure 21: Alcohol & Alcoholic Lotion sales 2019-2022

As we can tell from the graph there is no seasonality for these products as they contain products that during the Covid-19 outbreak had an enormous demand cause helped people clean their hands and other surfaces in order to protect themselves from germs that may another people convey to them from coming into contact with them. When people get vaccinated or felt more protected from diseases the sales dropped dramatically and remained in a somehow flat trend.



Figure 22: Goodness of fit for Alcohol & Alcoholic Lotions

We detect that estimated sales are close enough to Sales for Alcohol & Alcoholic Lotions, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

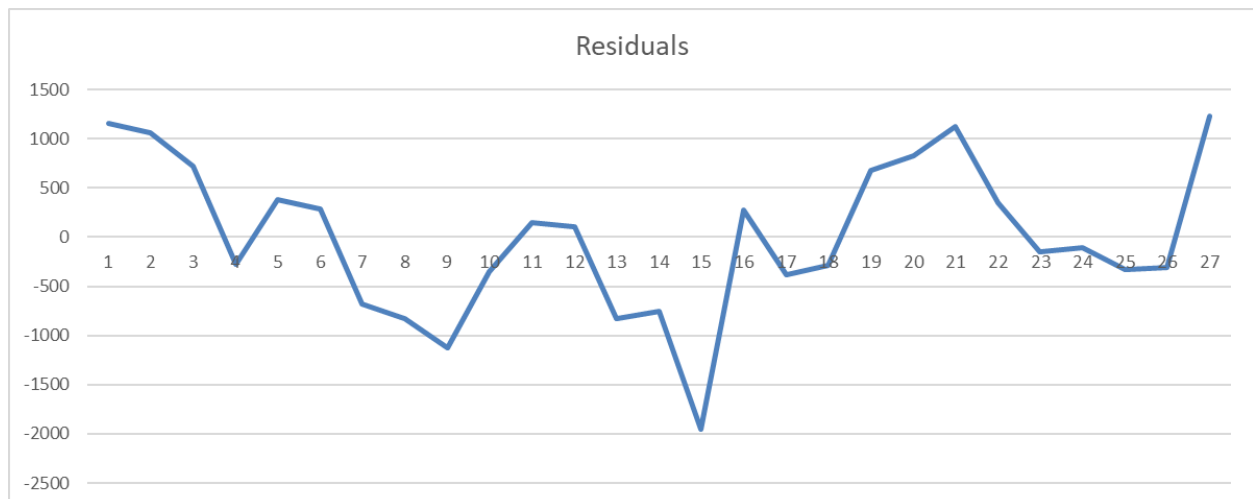


Figure 11: Residuals for Alcohol & Alcohol Lotions

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are positive, then negative, then positive, then negative, then positive again, then comes close to zero and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

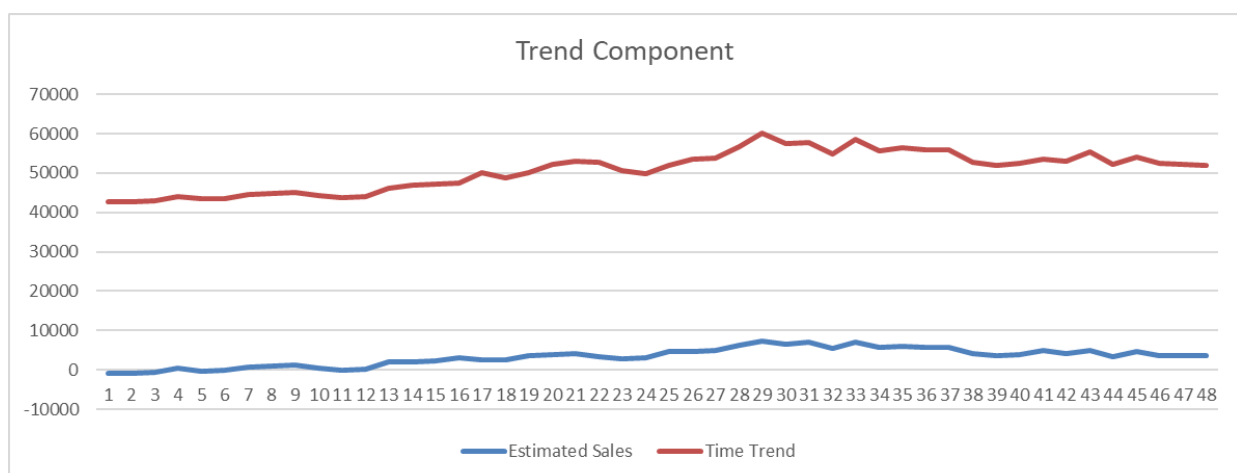


Figure 12: Time trend Component for Alcohol & Alcoholic Lotions 2019-2021

We can indicate that there is a correlation for sales in this graph. According to 'time trend' as times by we can see that's sales increases also in a small amount.

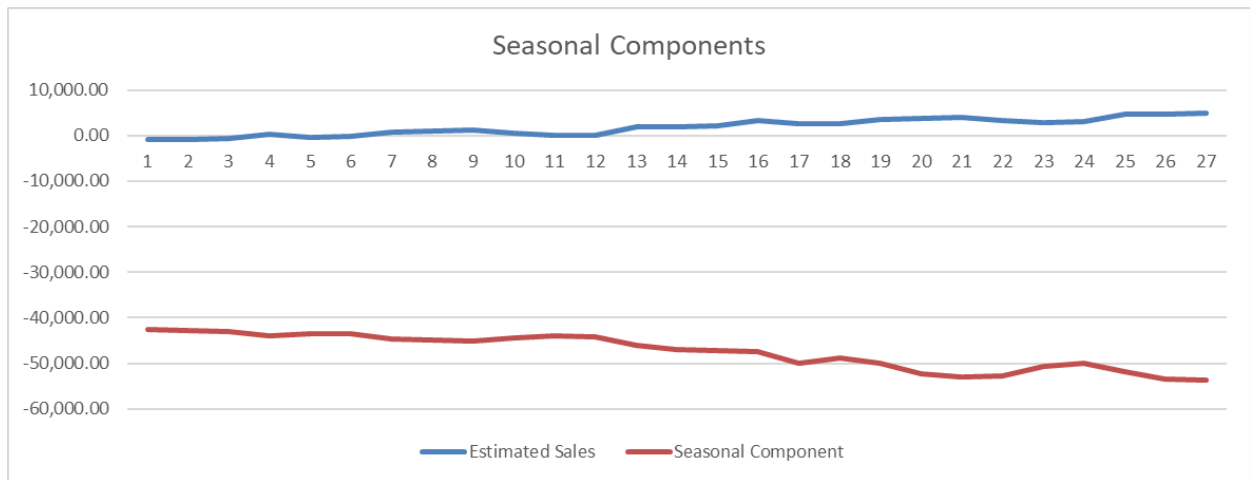


Figure 25: Seasonal Component for Alcohol & Alcoholic Lotions

We can imply that there is seasonality from this graph. Because we don't see a repeated pattern in sales. The sales have a slow increase pattern.

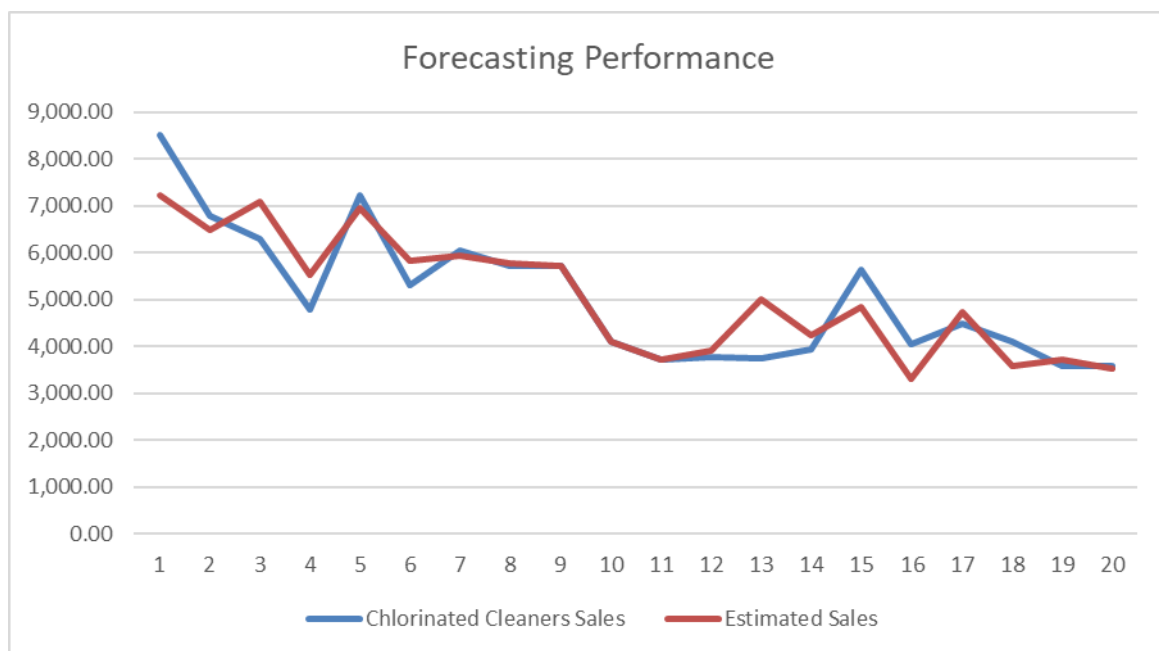


Figure 26: Forecasting Performance for Alcohol & Alcoholic Lotions

We can assume from the exceeding chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 17.

Table 17: MAE-MAPE results from EWMA & Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	8.45	0.16%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	30.16	0.15%
REGRESSION MODEL			
MAE		MAPE	
410.96		8.20%	

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Alcohol & Alcoholic Lotions products category.

5.4: Coffee

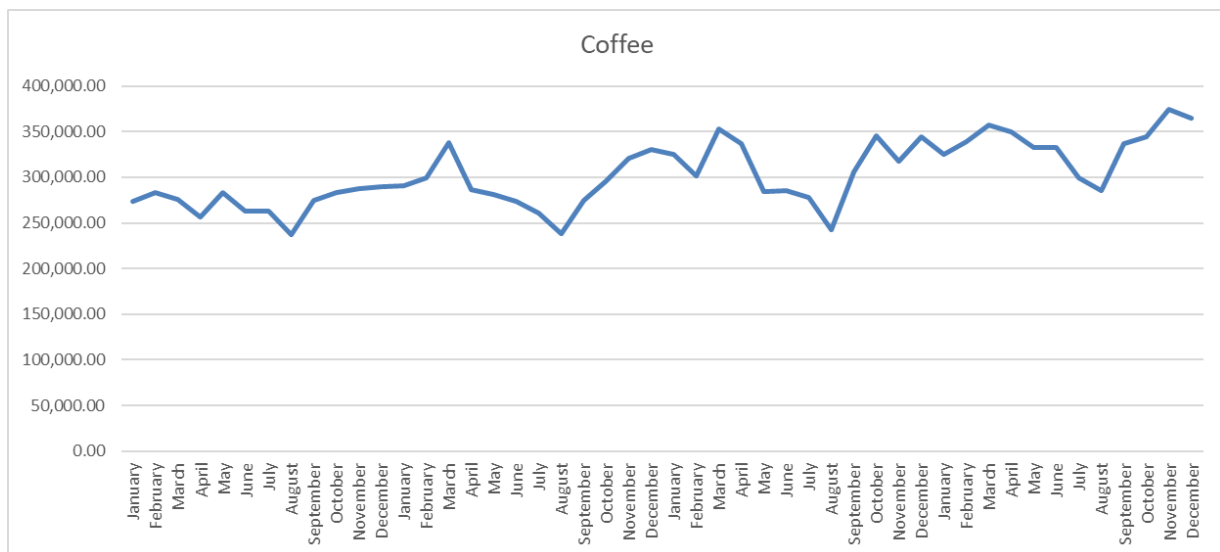


Figure 137: Coffee sales 2019-2022

We can detect that there is a seasonality in sales of Coffee and that's because the measures Government took for the operation of Coffee shop's during the Covid-19 pandemic. As we can see the spikes that sales show during the months that Government had the Coffee shop's close for hygienic reasons.

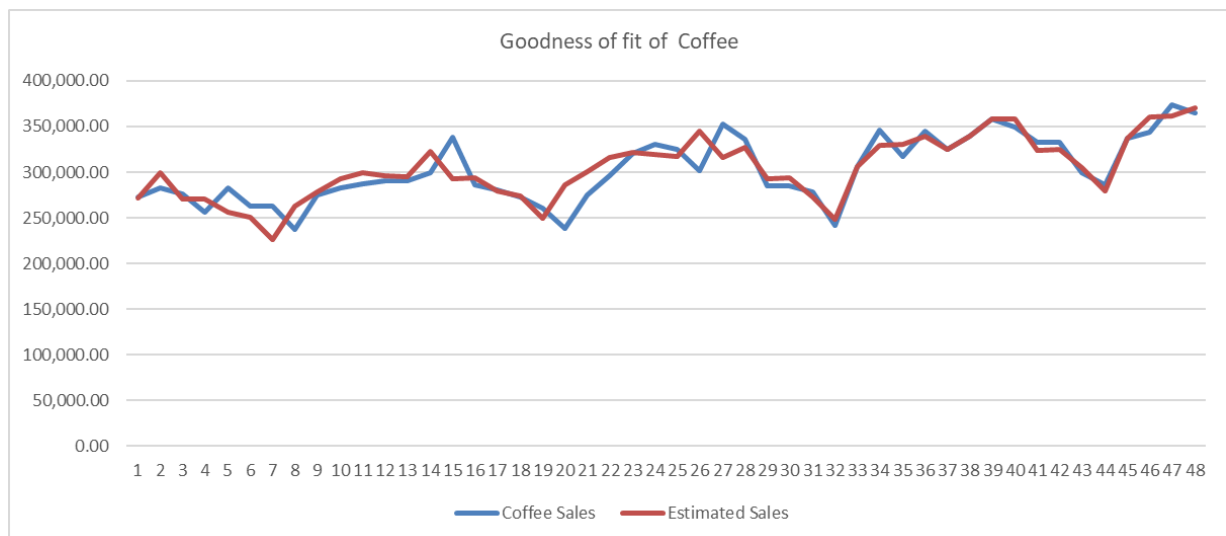


Figure28: Goodness of fit of Coffee

We detect that estimated sales are close enough to Sales for Coffee, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

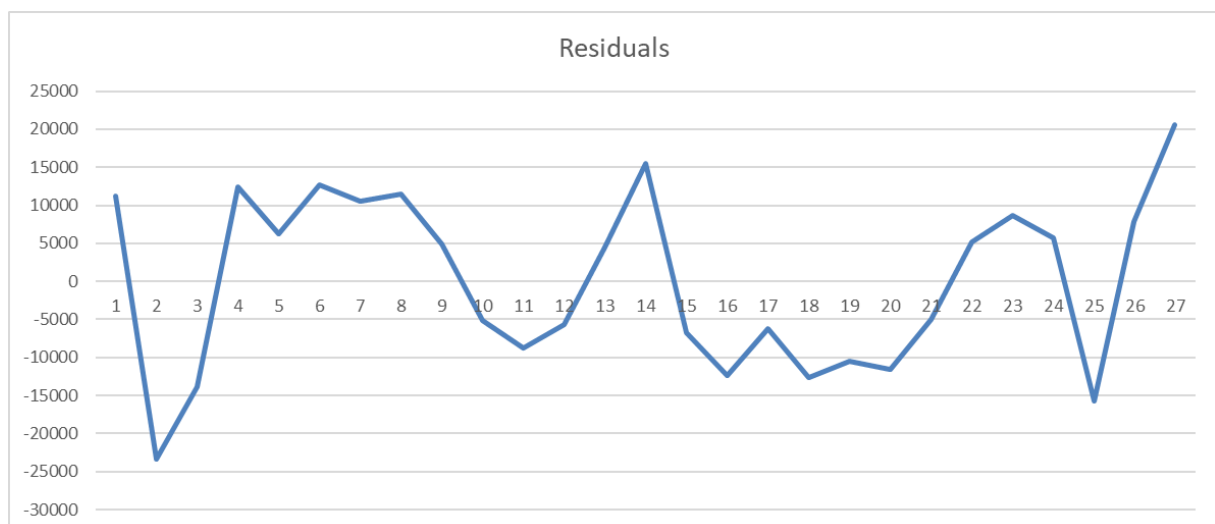


Figure29: Residuals For Coffee

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are positive, then negative, then positive, then negative, then positive, then negative, then positive, then negative and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

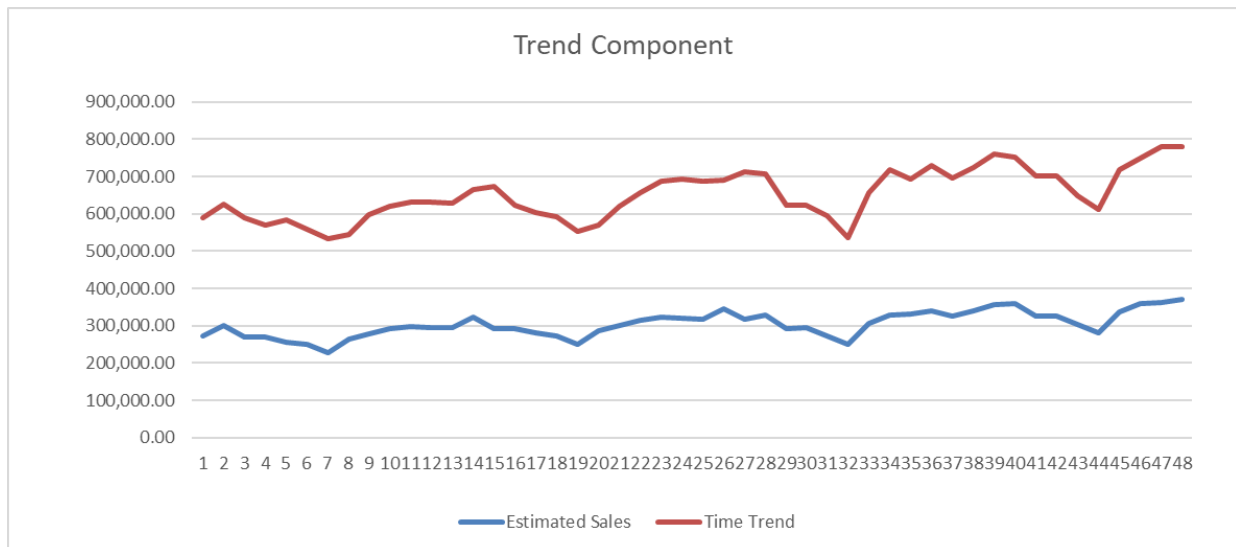


Figure 30: Time Trend Component for Coffee 2019-2021

We can indicate that there is a seasonality for sales in this graph. According to 'time trend' as times by we can see that's sales according to time frame increases or decreases.

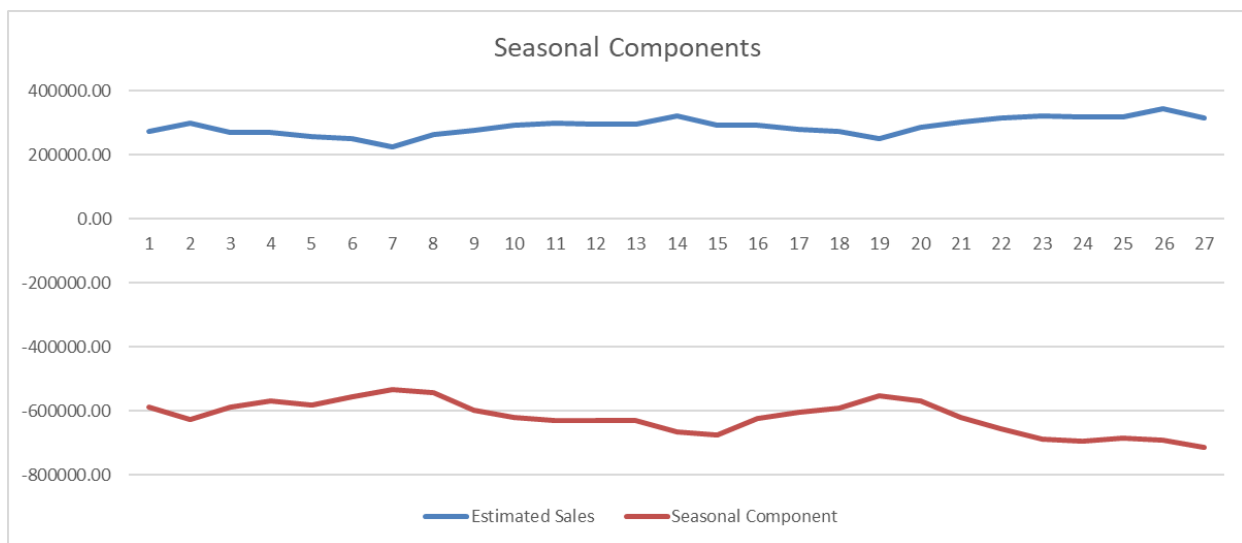


Figure 31: Seasonal Component for Coffee 2019-2021

We can imply that there is some partial seasonality from this graph. Because as we can see the line of estimated sales and Seasonal component have a same pattern.

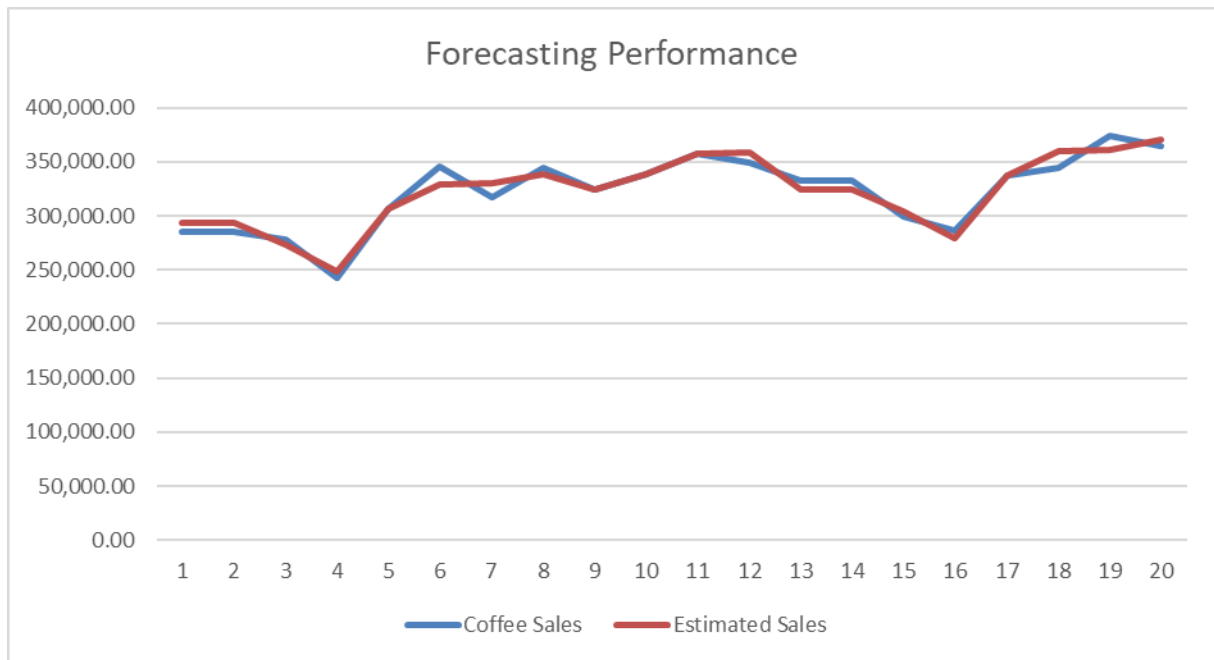


Figure 32: Forecasting Performance for Coffee 2021-2022

We can assume from the surpassing chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 18.

Table 18: MAE-MAPE results from EWMA & Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	223.86	0.07%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	184.23	0.06%
REGRESSION MODEL			
MAE		MAPE	
6711.06		2.09%	

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Alcohol & Alcoholic Lotions products category.

5.5 Snacks

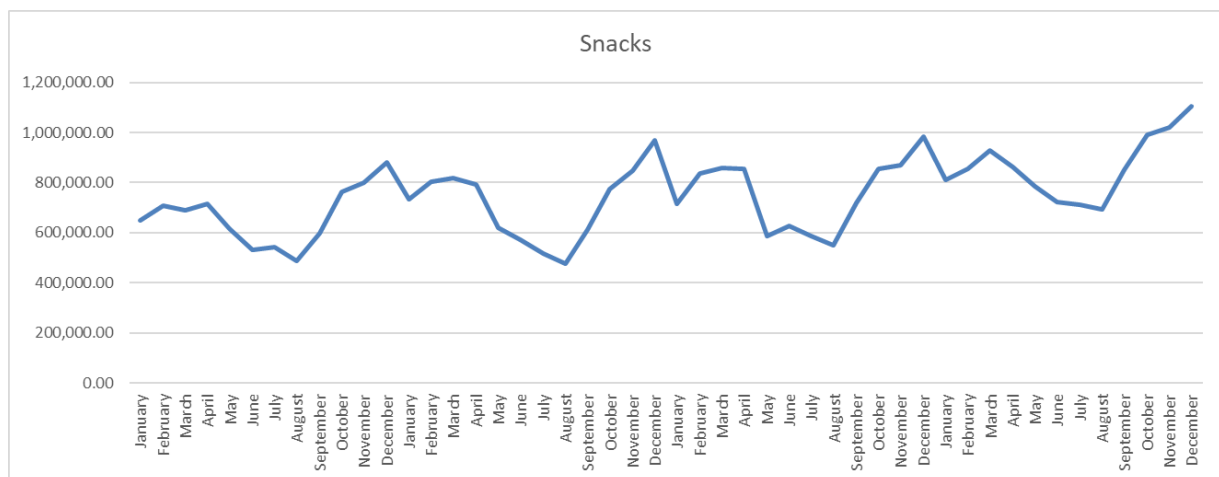


Figure 33: Snacks sales 2019-2022

We can detect that there is a seasonality in sales of Snacks. As we can see the spikes that sales show increase are the same for each year.

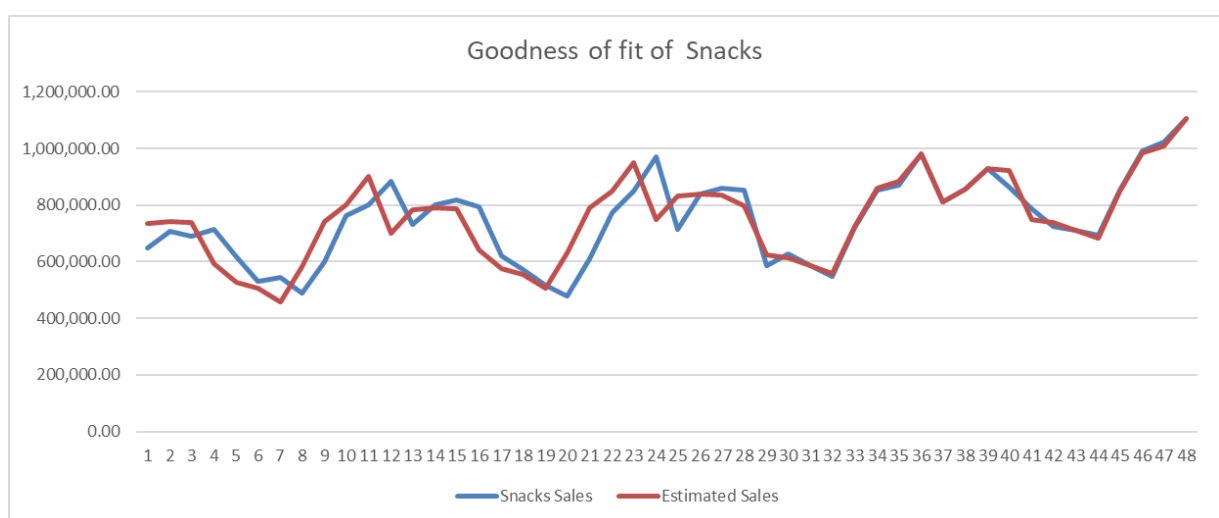


Figure 34: Goodness of fit for Snacks

We detect that estimated sales are close enough to Sales for Snacks, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

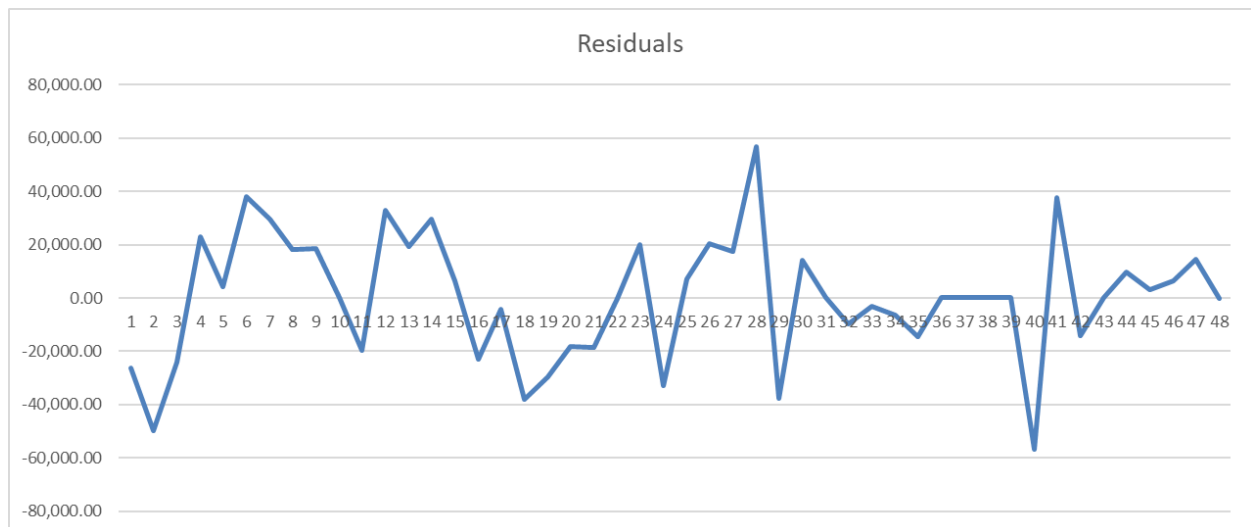


Figure 35: Residuals for Snacks

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are negative, then positive, then negative, then positive, then negative, then positive, and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

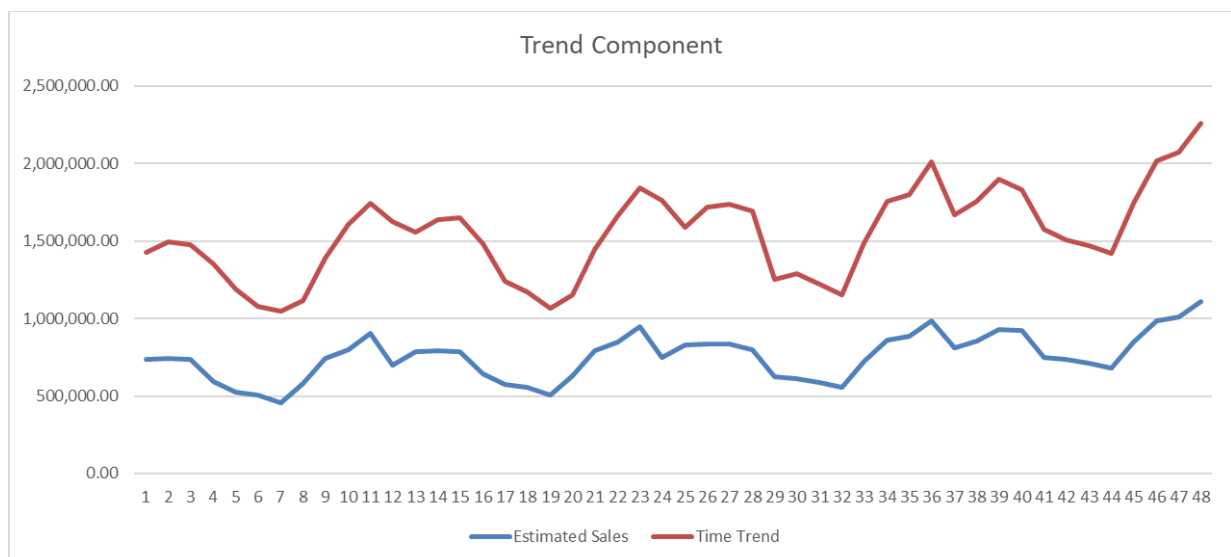


Figure 36: Time Trend Component for Snacks 2019-2021

We can indicate that there is a seasonality for sales in this graph. According to 'time trend' as times by we can see that's sales according to time frame increases or decreases.

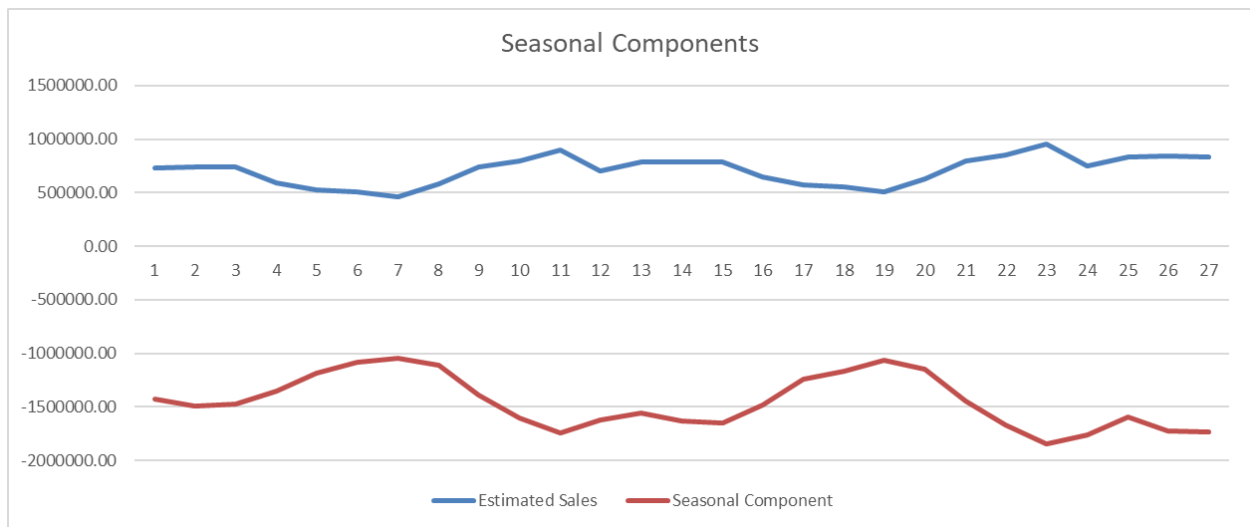


Figure 37: Seasonal Component for Snacks 2019-2021

We can imply that there is some partial seasonality from this graph. Because as we can see the line of estimated sales and Seasonal component have a same pattern.

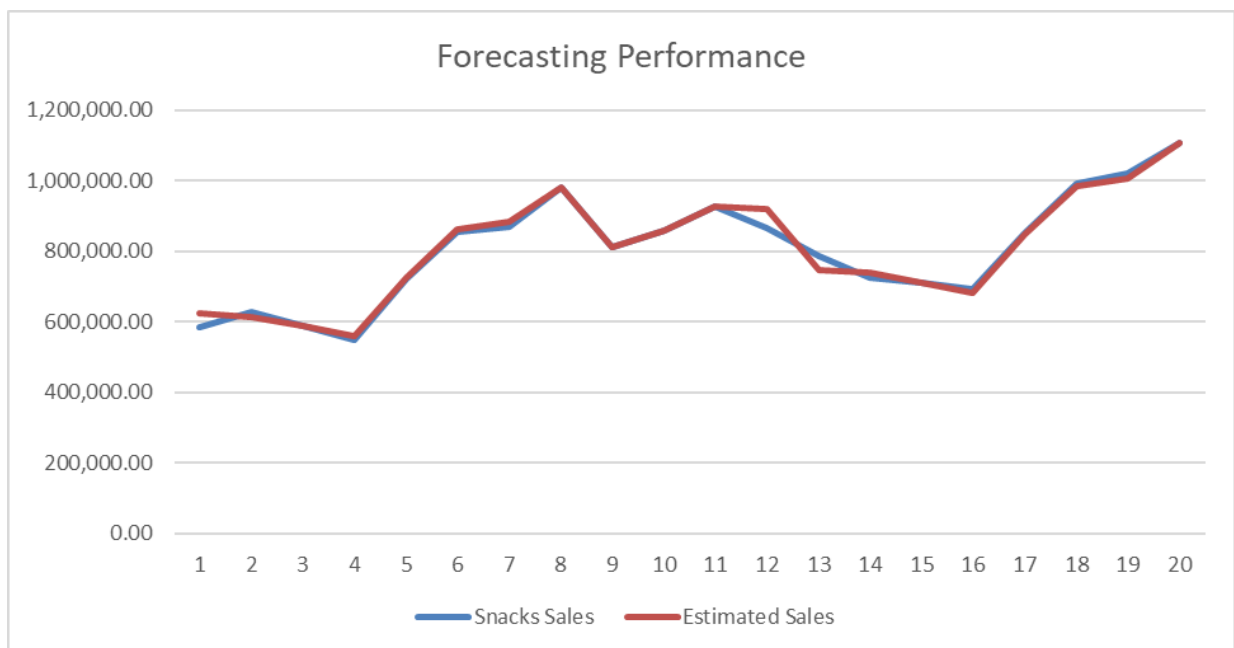


Figure 38: Forecasting Performance for Snacks 2021-2022

We can assume from the surpassing chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 18.

Table 19: MAE-MAPE results from EWMA & Regression Model

EWMA MODELS			
ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	789.15	0.10%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	824.22	0.12%
REGRESSION MODEL			
MAE	MAPE		
11415.09	1.53%		

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Snacks products category.

5.6: Flour

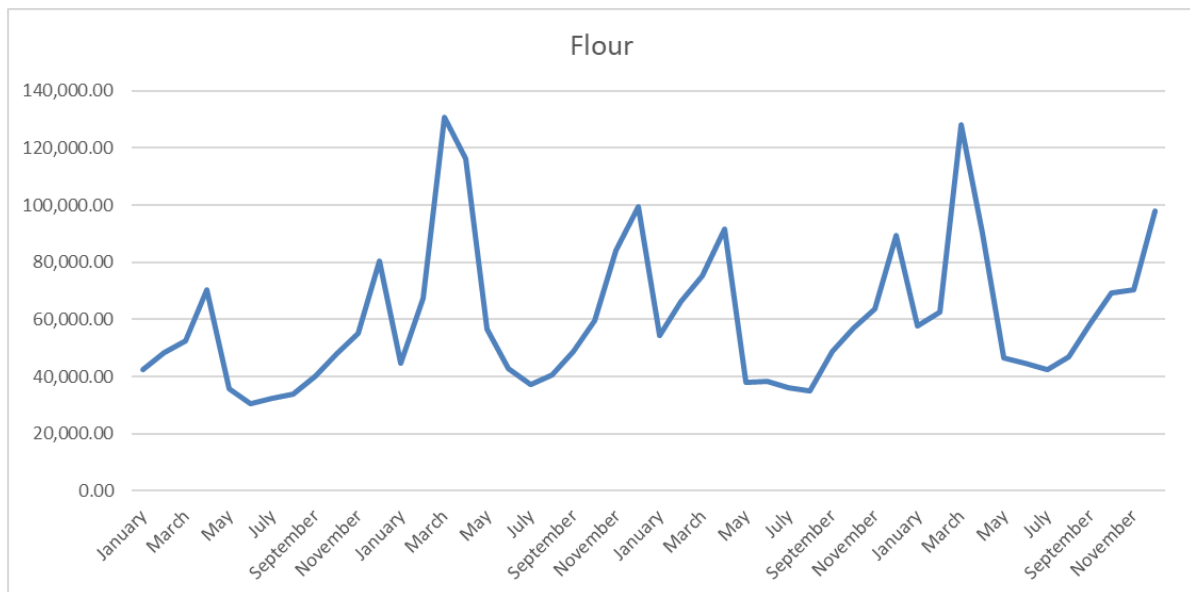


Figure 39: Flour sales 23019-2022

We can detect that there is a seasonality in sales of Flour. As we can see the spikes that sales show increase are the same for each year.

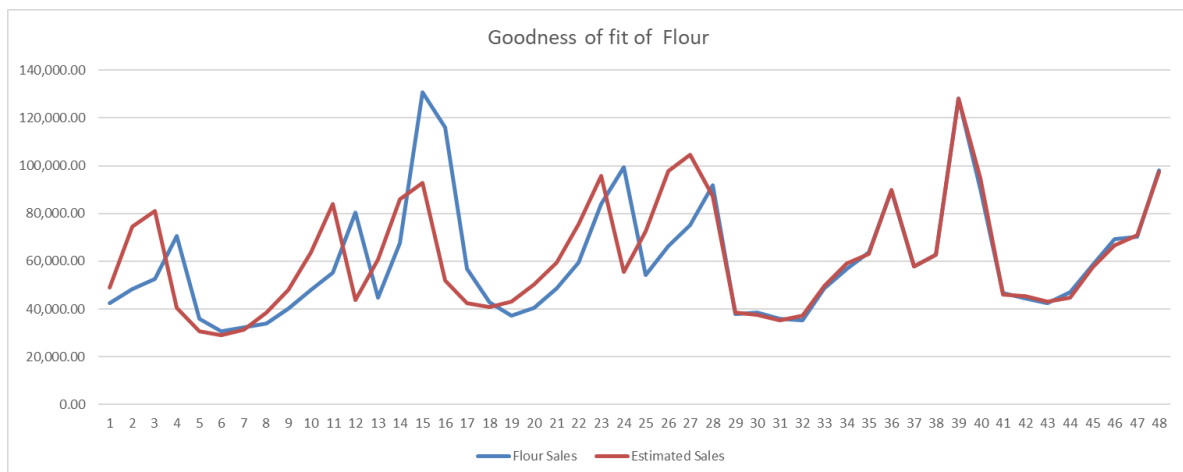


Figure 40: Goodness of fit of Flour

We detect that estimated sales are close enough to Sales for Flour, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

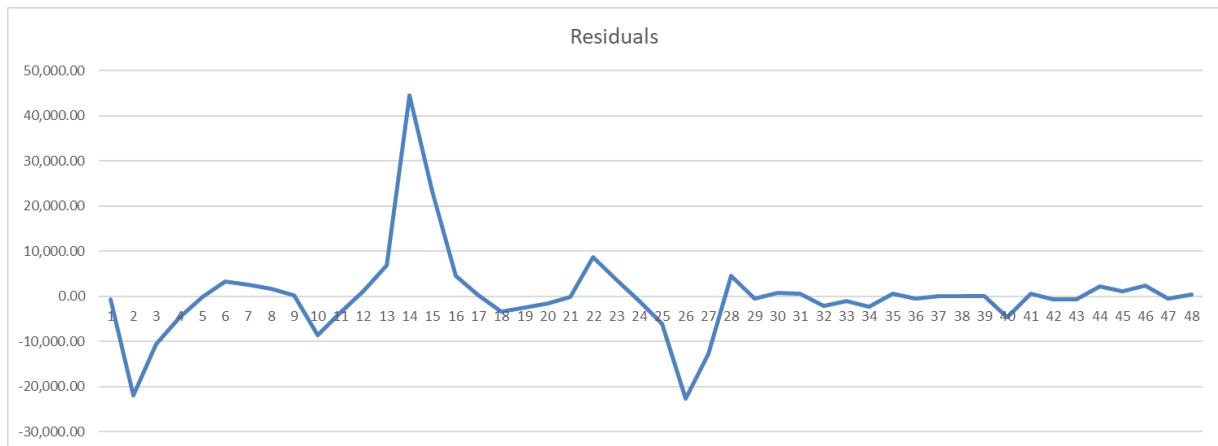


Figure 41: Residuals for Flour

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are negative, then positive, then negative, then positive, then negative, then positive, then negative and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

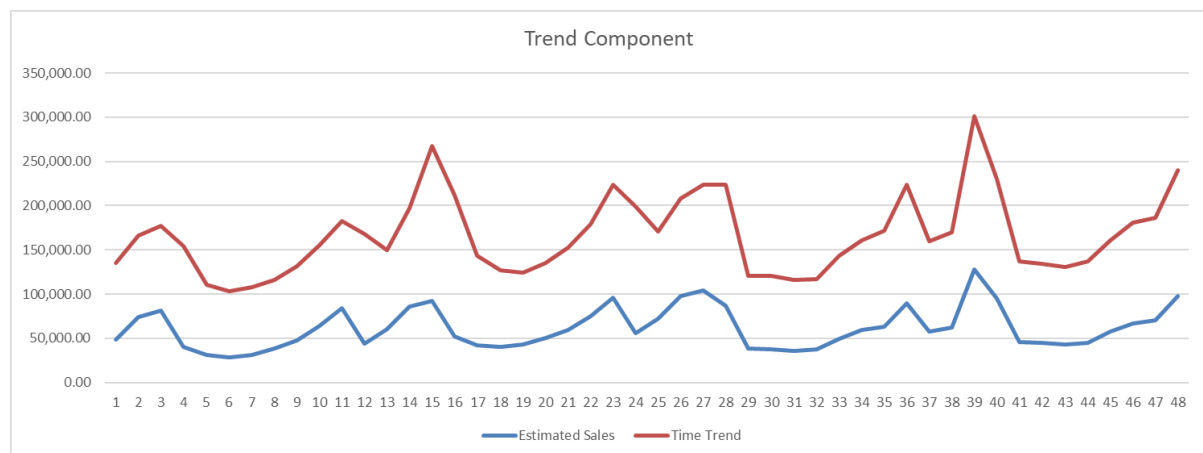


Figure 42: Time Trend Component for Flour 2019-2021

We can indicate that there is a seasonality for sales in this graph. According to 'time trend' as times by we can see that's sales according to time frame increases or decreases at the same time frame.

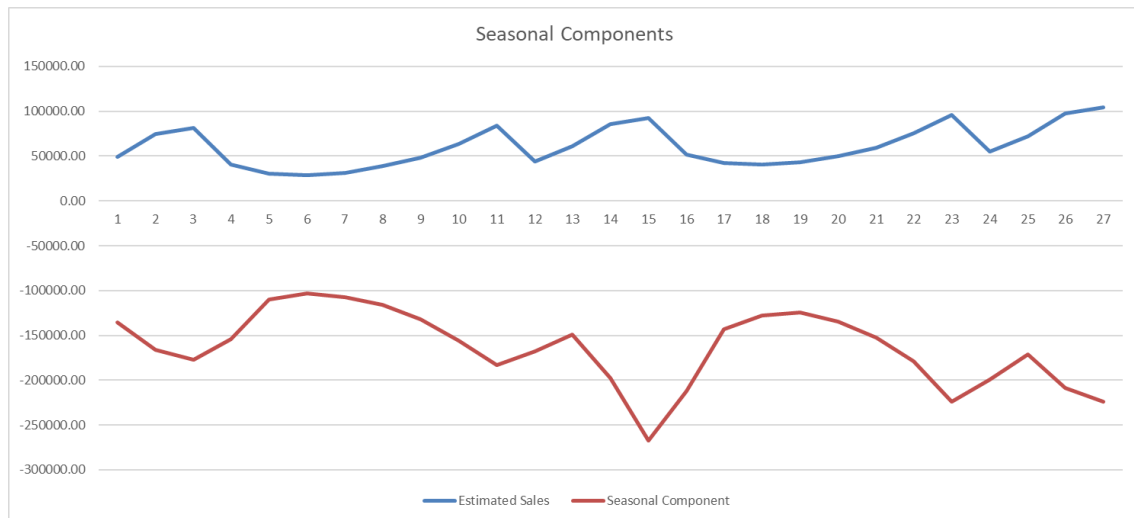


Figure 43: Seasonal Component for Flour 2019-2021

We can imply that there is a seasonality from this graph. Because as we can see the line of estimated sales and Seasonal component have a same pattern

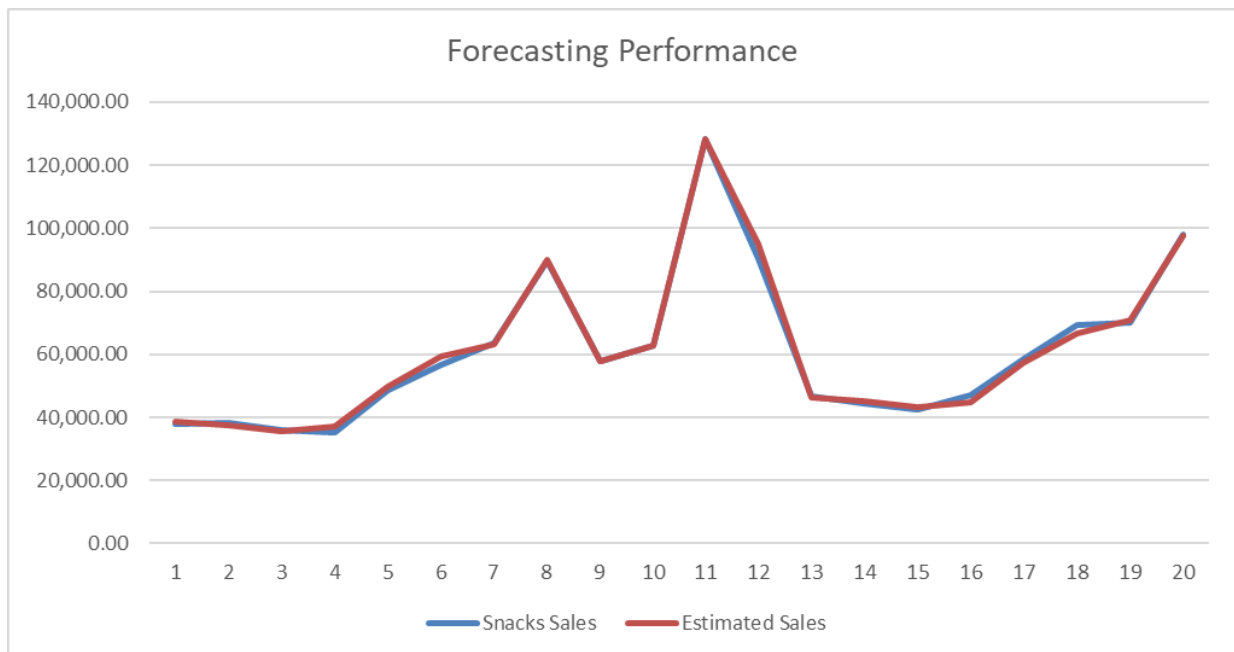


Figure 44: Forecasting Performance for Flour 2021-2022

We can assume from the surpassing chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 20.

Table 20: MAE-MAPE results from EWMA & Regression Model

ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	158.76	0.22%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	176.87	0.29%
REGRESSION MODEL			
MAE	MAPE		
1062.55	1.94%		

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Flour products category.

5.7: Dairy

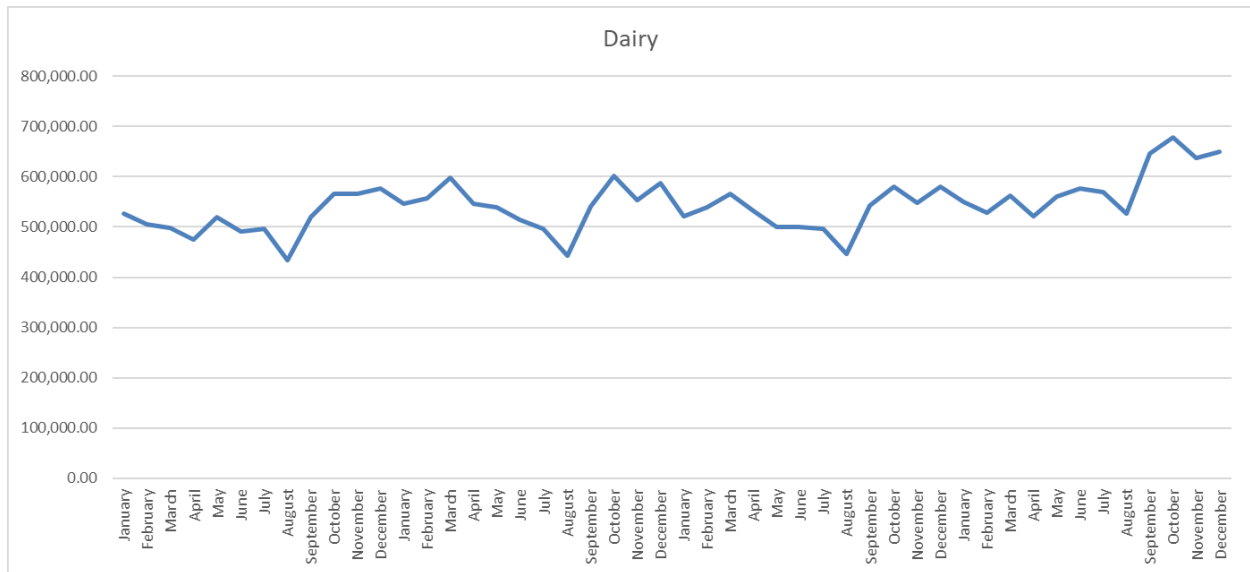


Figure 45: Dairy sales 2019-2022

We can detect some seasonality in sales of Dairy. As we can see the spikes that sales show increase or decrease are at the same timeframe for each year.

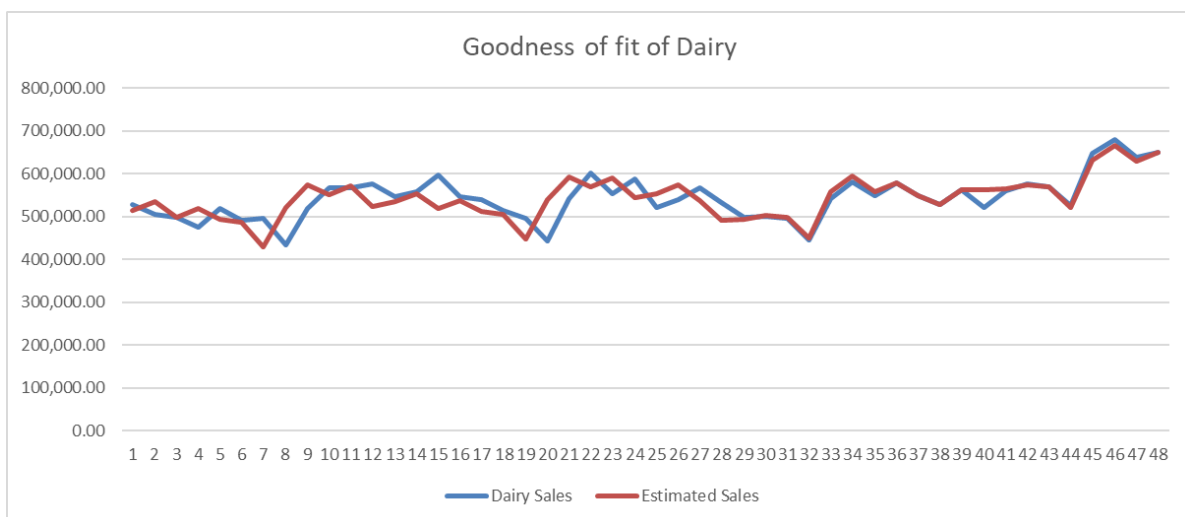


Figure 46: Goodness of fit Dairy

We detect that estimated sales are close enough to Sales for Dairy, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

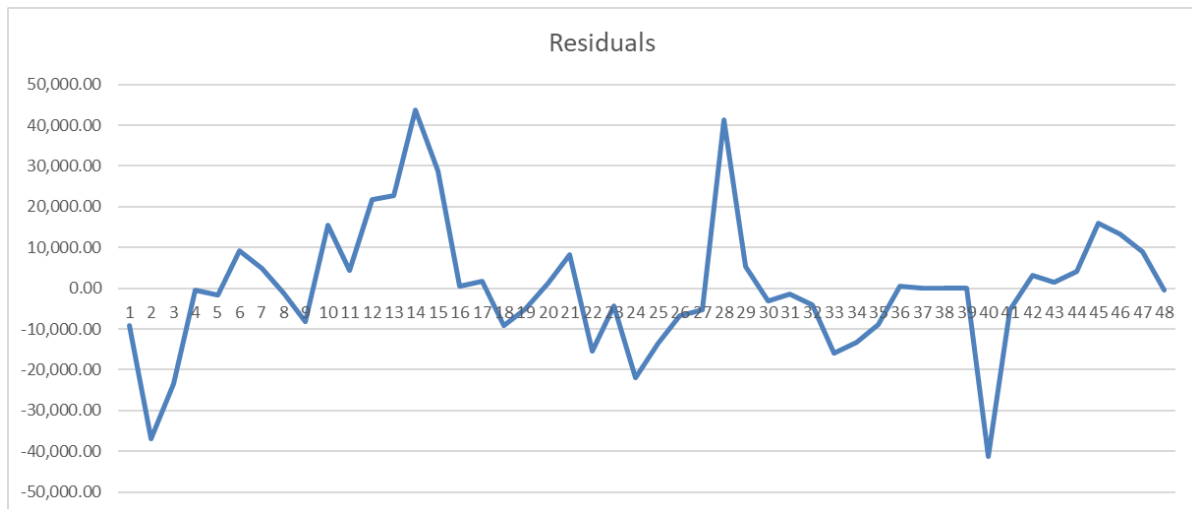


Figure 47: Residuals for Dairy

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are negative, then positive, then negative, then positive, then negative, then positive, then negative and last residuals are positive. This arbitrary pattern signifies that a linear model offers a good match to the data.

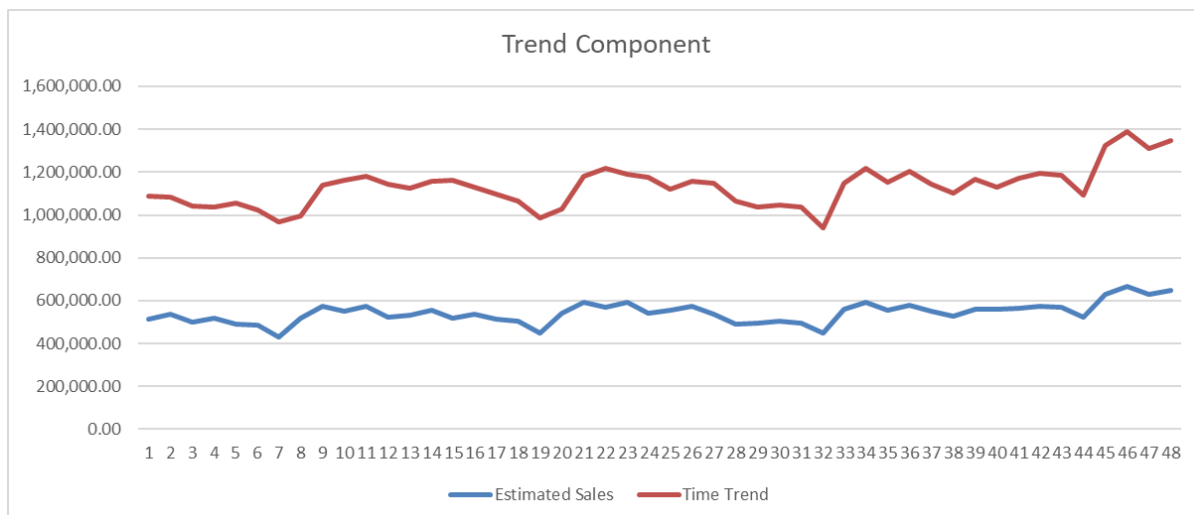


Figure 48: Time Trend Component for Dairy 2019-2021

We can indicate that there is a seasonality for sales in this graph. According to 'time trend' as time by we can see that's sales increases or decreases at the same time frame.

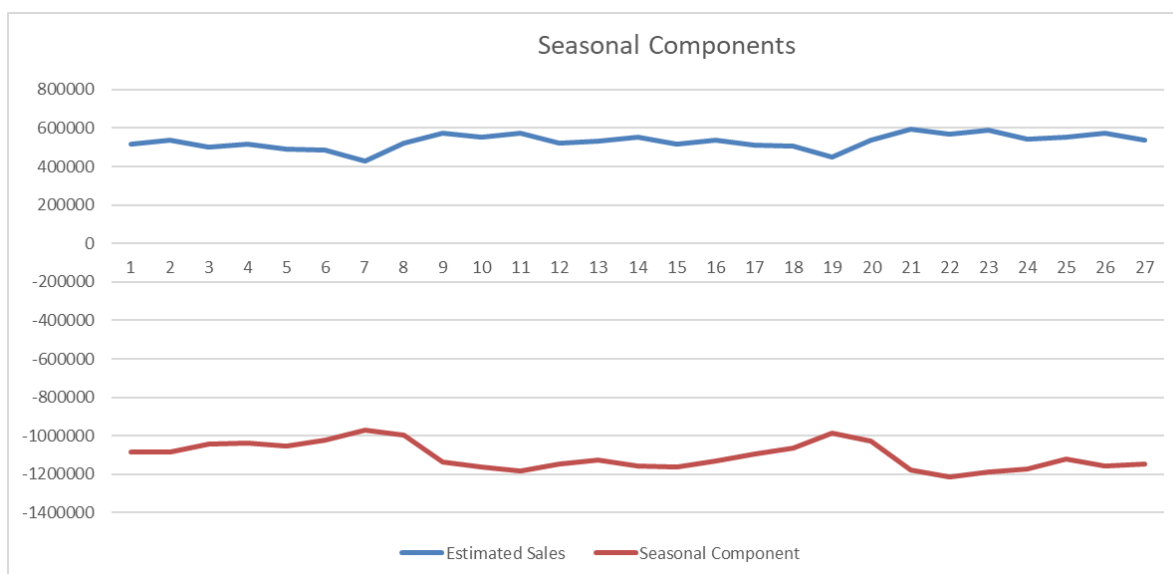


Figure 49: Seasonal Component for Dairy 2019-2021

We can imply that there is a seasonality from this graph. Because as we can see the line of estimated sales and Seasonal component have a same pattern

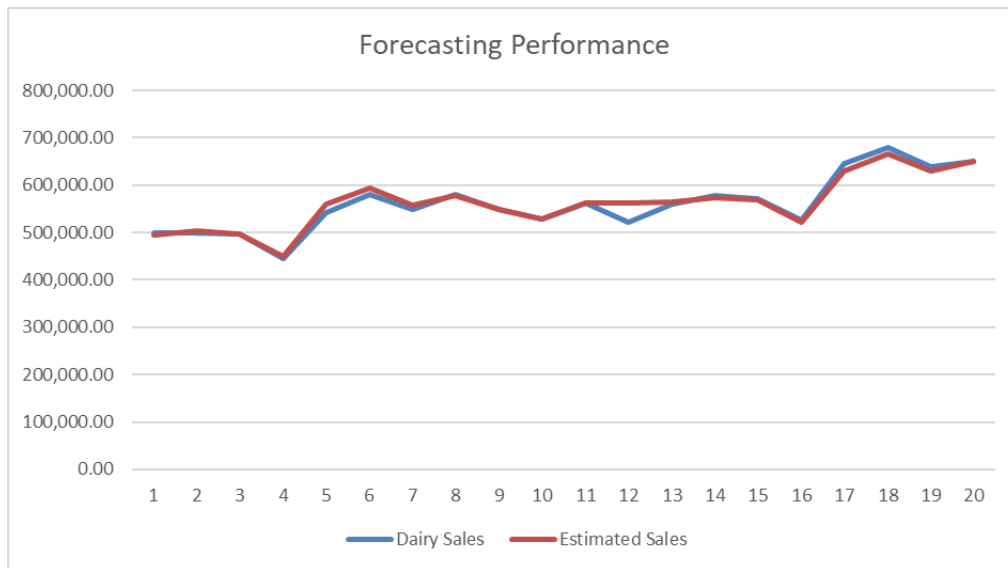


Figure 50: Forecasting Performance for Dairy 2021-2022

We can assume from the surpassing chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 21.

Table 21: MAE-MAPE results from EWMA & Regression Model

ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	362.81	0.06%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	353.78	0.07%
REGRESSION MODEL			
MAE	MAPE		
7322.91	1.31%		

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Dairy products category.

5.8: Alcoholic Beverages

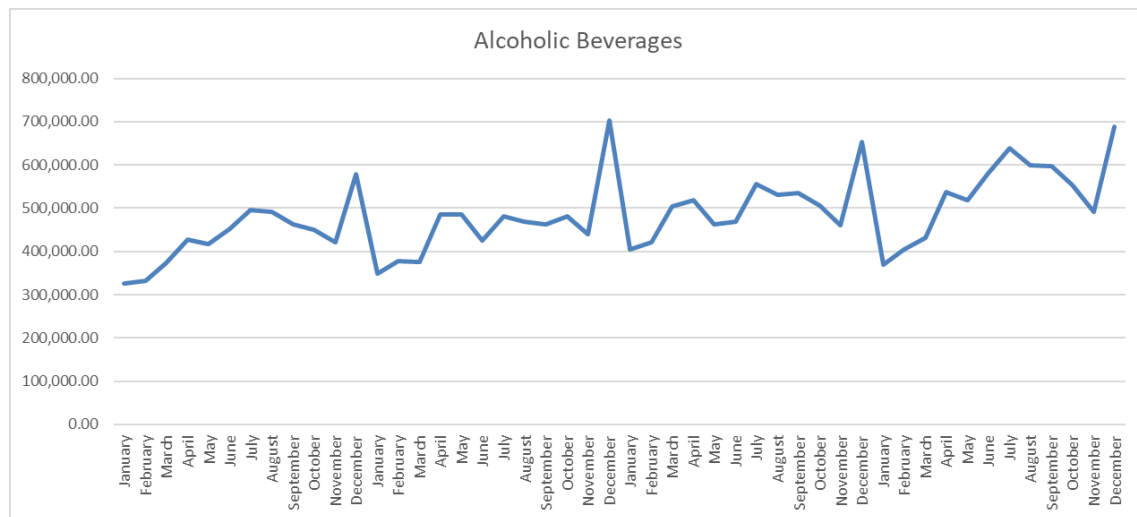


Figure 51: Alcohol Beverages sales 2019-2022

We can detect some seasonality in sales of Dairy. As we can see the spikes that sales show increase or decrease are at the same timeframe for each year.

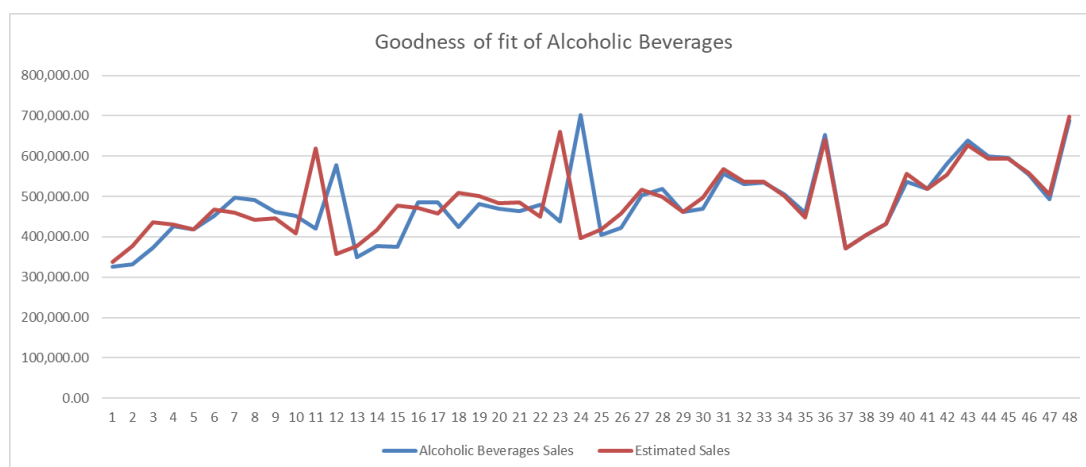


Figure 52: Goodness of fit Alcohol Beverages

We detect that estimated sales are close enough to Sales for Alcohol Beverages, which gives the company the advantage of predicting the necessary orders to fulfill the needs of consumers.

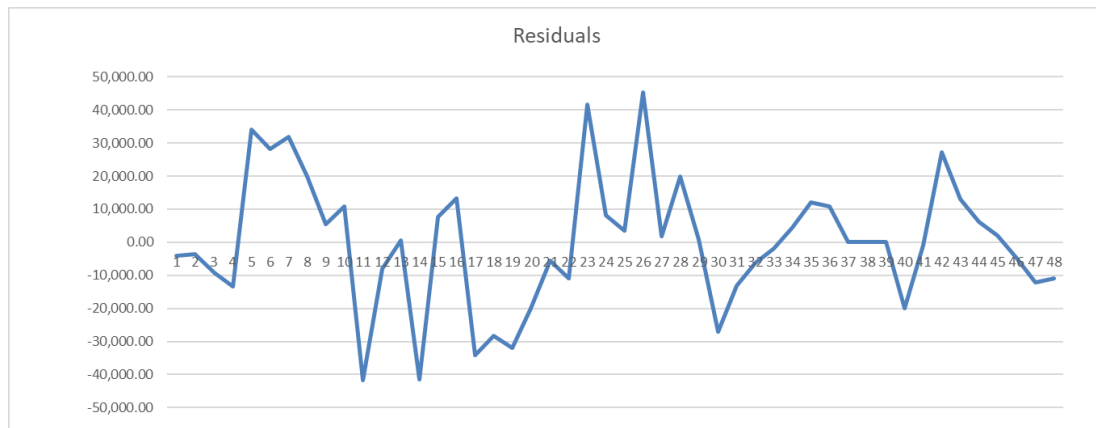


Figure 53: Residuals for Alcohol Beverages

The residual diagram shows a reasonably arbitrary formation. We can see that the first residuals are negative, then positive, then negative, then zero, then negative, then positive, then negative, then positive, negative, then positive, then negative, then positive and last residuals are negative. This arbitrary pattern signifies that a linear model offers a good match to the data.

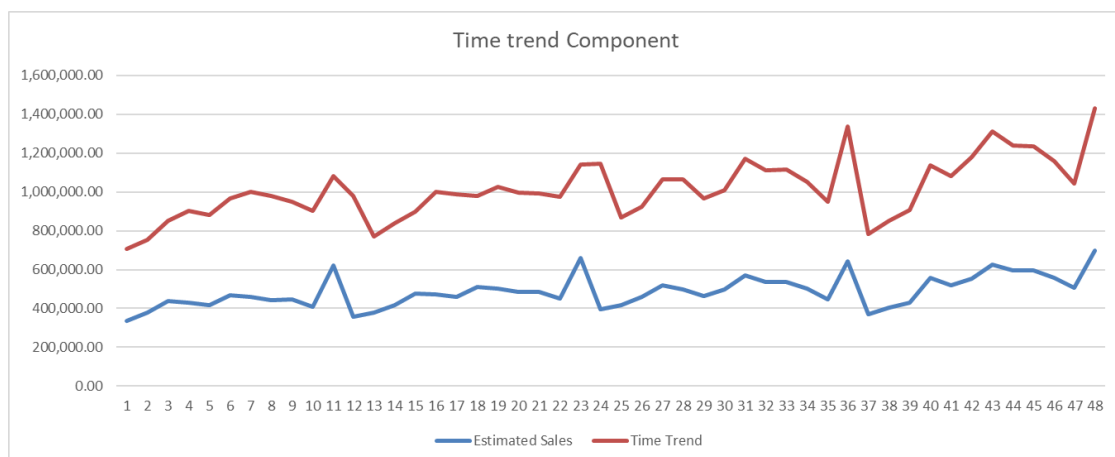


Figure 54: Time Trend Component for Alcohol Beverages 2019-2021

We can indicate that there is a seasonality for sales in this graph. According to 'time trend' as time by we can see that's sales increases or decreases at the same time frame.

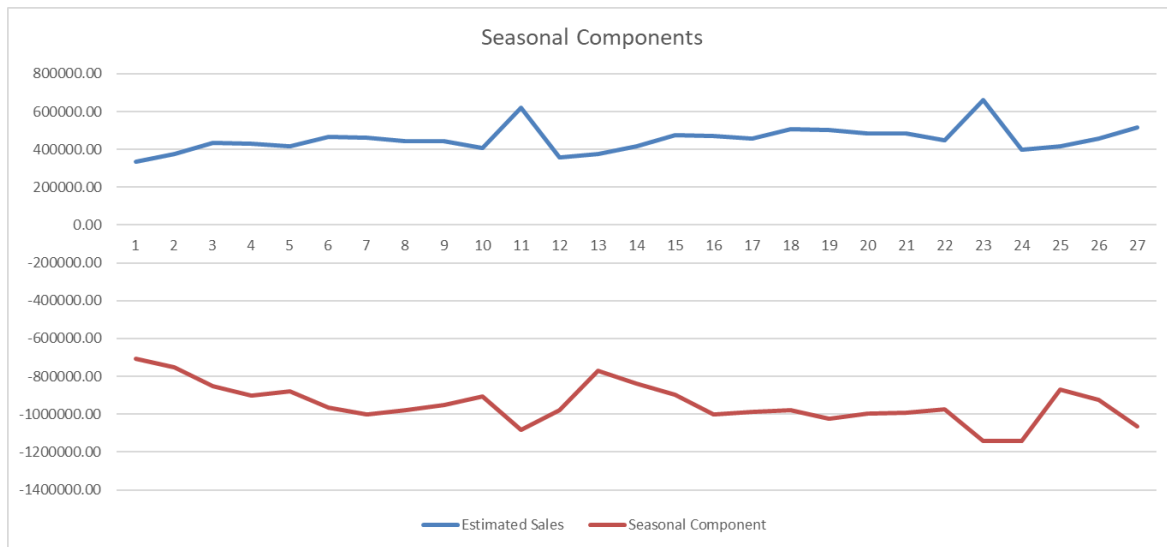


Figure 55: Seasonal Component for Alcohol Beverages 2019-2021

We can imply that there is a seasonality from this graph. Because as we can see the line of estimated sales and Seasonal component have a same pattern

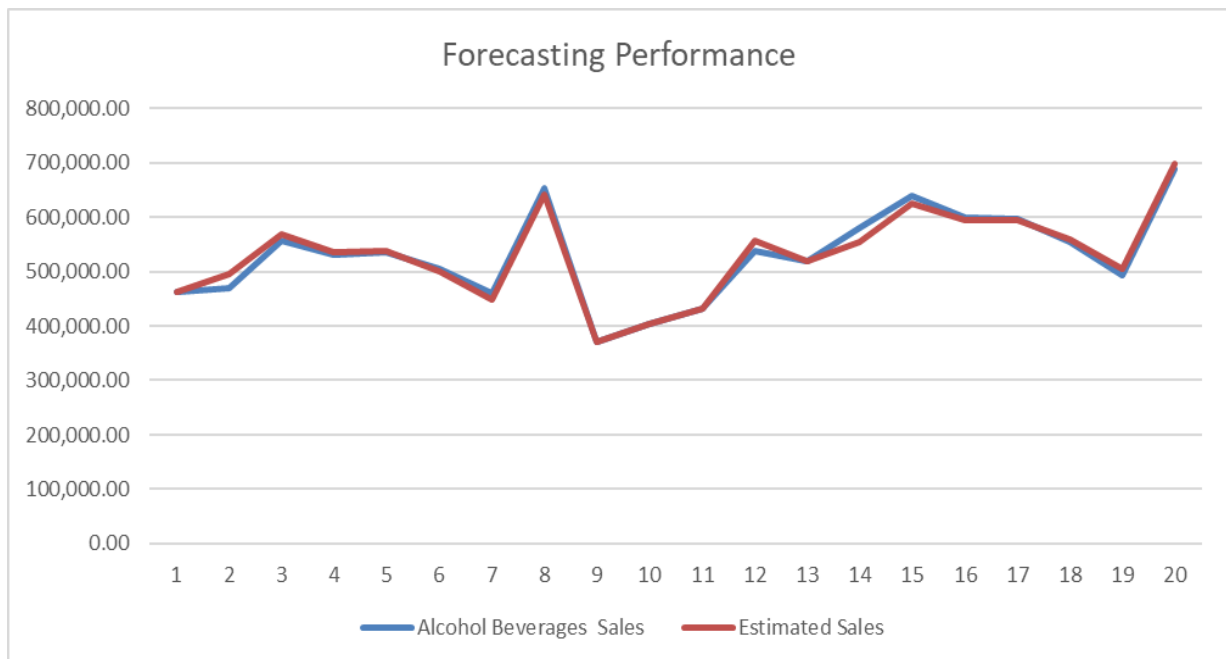


Figure 56: Forecasting Performance for Alcohol Beverages 2021-2022

We can assume from the surpassing chart that this is quite a promising discovery. To evaluate and choose which of the two models we ran is the most effective, we compiled all of the MAPE from each, and the results are included in Table 22.

Table 22: MAE-MAPE results from EWMA & Regression Model

ESTIMATED			
A/A	λ	MAE	MAPE
1	0.01	692.63	0.13%
FORECASTED			
A/A	λ	MAE	MAPE
1	0.01	611.48	0.13%
REGRESSION MODEL			
MAE	MAPE		
8677.64	1.60%		

As we can notice from Table below, when taking into attention the MAPE from the EWMA models, and the MAPE from the regression model, we draw the conclusion that the EWMA model shall be considered a better forecasting device, as it has the smaller MAPE, as refer to this point to the Dairy products category.

CHAPTER 6: CONCLUTIONS AND PROPOSAL FOR FURTHER RESEARCH

6.1 Conclusions

Fitting forecasting control methods and time series estimations can be done in a variety of ways. We have described this research. executed and compiled two distinct forecasting techniques. Using time series models based on trend and seasonality records, we have examined and forecasted. Regression analysis, as opposed to EWMA, showed to be a more appropriate good fit model, and gave us more insightful results, according to our study.

6.2 Proposal for further research

Our research on sales before, during, and after the COVID-19 period has been completed. It provides significant evidence of how products respond to the COVID-19 pandemic in the Greek market. However, the study has certain limitations that may leave room for additional research on the topic.

Although the frequency of the data are static (period of a month), shows us how people react in a situation like Covid-19 pandemic or another crisis. The statistical significance test of individual coefficients may have led to the selection of other models in the absence of thorough testing. Moreover, we discovered that not all variables are equally significant, so we didn't incorporate them in order to make our assumptions.

The following are some topics that might be included in a follow-up survey: First, the researcher thinks it's a good idea to look at how sales were affected by the Covid-19 outbreak and restrictions by using a different technique, such a logarithm regression for the t-1, t-2, and t-3 months prior.

An equally fascinating method would be for a study to look at weekly sales. This would bring it more in line with reality and uncover other factors that should be considered. The ability to compare the Super Market industry to others that are similar would also be of interest. In this manner, we may gain crucial insights and a clearer understanding of consumer behavior, which would help the management team make critical decisions.

References

http://www.h2oalliance.org/images/Organizations_at_Risk.pdf (1.1)

http://ethesisarchive.library.tu.ac.th/thesis/2019/TU_2019_6122040063_12585_12330.pdf (1.2)

Beverage Market Analysis and Forecast. The case of Three Cents,
TZIORA AIKATERINI, July 2021 (1.3)

<https://ageconsearch.umn.edu/record/333527/> (1.4)

<http://www.ielka.gr/wp-content/uploads/2013/07/An-overview-of-the-Greek-grocery-retail-sector-white-paper.pdf> (2.1)

<https://kpmg.com/gr/el/home/insights/2023/04/the-future-of-retail.html> (2.2)

<https://www.mdpi.com/2075-5309/13/3/627> (2.3,2.4)

Mark A. Moon, John T. Mentzer, Carlo D. Smith, and Michel S Garver. (2018). *Demand and Supply Integration: The key to world Class Demand Forecasting*. DEG Press. (Research tools and 3.3,3.4,3.5)

<https://www.isixsigma.com/dictionary/exponentially-weighted-moving-average-ewma/> (3.2 and 4.3.1)

https://books.google.gr/books?hl=en&lr=&id=fW_9BV5Wpf8C&oi=fnd&pg=PR1&dq=Statistical+Models:+Theory+and+Practice&ots=2jHaYzFZTJ&sig=ZakJM14vok5wXILDp6as0sdnY70&redir_esc=y#v=onepage&q=Statistical%20Models%3A%20Theory%20and%20Practice&f=false David A. Freedman (27 April 2009). *Statistical Models: Theory and Practice*. Cambridge University Press. ISBN 978-1-139-47731-4. (4.3)

<https://www.ablebits.com/office-addins-blog/linear-regression-analysis-excel/> (4.3.2)