CONSTRUCTING A FORECASTING MODEL FOR DECREASING DEMAND DEVIATION EFFECTS OF PRODUCTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

CANSU DEMİR KARTBOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF DATA INFORMATICS

MAY 2025

CONSTRUCTING A FORECASTING MODEL FOR DECREASING DEMAND DEVIATION EFFECTS OF PRODUCTS

Submitted by Cansu Demir Kartbol in partial fulfillment of the requirements for the degree of **Master of Science in Data Informatics Department, Middle East Technical University** by,

Date: 26.05.2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Cansu Demir Kartbol

Signature : _____

ABSTRACT

CONSTRUCTING A FORECASTING MODEL FOR DECREASING DEMAND DEVIATION EFFECTS OF PRODUCTS

Demir Kartbol, Cansu

MSc., Department of Data Informatics

Supervisor: Prof.Dr. Pekin Erhan Eren

Co-supervisor: Prof. Dr. Altan Koçyiğit

May 2025, 78 pages

Economic changes, global cases, technological advancements, market competition, and the nature of product demand increase demand variability. From a supply chain management perspective, developing a method that can effectively forecast this fluctuating demand is crucial for increasing profitability and optimizing inventory management. In this thesis, real-world data from a home appliance company is utilized to improve forecasting accuracy for cooling and freezer products, aiming to optimize supply chain strategies. A data-driven approach is adopted by utilizing various forecasting models such as SARIMA, LSTM, and Prophet. Additionally, external factors such as the impact of COVID-19, economic indicators, and inventory levels are considered. Model performance is improved through hyperparameter optimization and evaluated using metrics such as MAE, RMSE, and SMAPE.

The study aims to enhance demand forecasting accuracy and optimize supply chain strategies by exploring various clustering techniques to refine product and country groupings. The models are applied to both product-level and product-region-level datasets. The best result for product-level forecasting is obtained using the Prophet model without exogenous variables, which exceeds the base model by 11%. For product-region-level forecasting, the SARIMA model using exogenous variables achieves the best performance, exceeding the base model by 24%. Furthermore, for both breakdowns (product; product-region), the evaluation results indicate that clustering products and countries with similar sales dynamics enhances forecasting performance. Consequently, this thesis contributes to the literature by focusing on the underexplored home appliance sector, particularly refrigerators and freezers, and providing actionable forecasting approaches for supply chain optimization.

Keywords: Time Series, Clustering Methods, Home Appliance Sector, Sales Forecasting Models, Exogenous Variables

ÖZ

ÜRÜNLERDEKİ TALEP DEĞİŞKENLİĞİNİN ETKİLERİNİ AZALTMAYA YÖNELİK BİR TAHMİN MODELİ OLUŞTURMA

Demir Kartbol, Cansu

Yüksek Lisans, Veri Bilişimi Bölümü

Tez Yöneticisi: Prof.Dr. Pekin Erhan Eren

Ortak Tez Yöneticisi: Prof. Dr. Altan Koçyiğit

Mayıs 2025, 78 sayfa

Ekonomik değişiklikler, küresel olaylar, teknolojik gelişmeler, pazar rekabeti ve ürün talebinin doğal yapısı gibi faktörler, talepteki değişkenliği artırmaktadır. Tedarik zinciri yönetimi açısından, bu değişken talebi etkili bir şekilde tahmin edebilecek bir yöntem geliştirmek, karlılığı artırmak ve envanter yönetimini verimli hale getirmek adına son derece önemlidir. Bu tezde, dayanıklı tüketim ürünleri üreten bir şirketten alınan veriler kullanılarak, soğutma ve dondurucu ürünleri için tahmin doğruluğunu artırmaya yönelik çalışmalar yapılmış ve tedarik zinciri stratejilerinin optimize edilmesi hedeflenmiştir. SARIMA, LSTM ve Prophet gibi çeşitli tahmin modellerini kullanarak veri odaklı bir yaklaşım benimsenmektedir. Ayrıca, Covid-19 etkisi, ekonomik göstergeler ve envanter seviyeleri gibi dışsal faktörler de dikkate alınmıştır. Modellerin performansı, hiper parametre optimizasyonu yapılarak iyileştirilmiş ve MAE, RMSE ve SMAPE gibi değerlendirme metrikleri kullanılarak karşılaştırılmıştır. Çalışma, ürün ve ülke gruplamalarını iyileştirmek için çeşitli kümeleme tekniklerini inceleyerek, talep tahminlerinin doğruluğunu artırmayı ve tedarik zinciri stratejilerinin iyileştirilmesi amaçlamaktadır. Modeller, ürün düzeyindeki ve ürün-bölge düzeyindeki veri setlerine uygulanmıştır. Ürün düzeyindeki tahmin için en iyi sonuç temel modelden %11 daha iyi performans gösteren Prophet modeli ile elde edilirken; ürün-bölge düzeyindeki tahmin için en iyi sonuç temel modelden %24 daha iyi performans gösteren dışsal değişkenlerin dahil edildiği SARIMA modeli ile elde edilmiştir. Ayrıca, her iki kırınım (ürün; ürün-bölge) için de değerlendirme sonuçları, benzer satış dinamiklerine sahip ürün ve ülkelerin gruplandırılmasının tahminleme performansını artırdığını göstermektedir. Sonuç olarak, bu tez, yeterince araştırılmamış beyaz eşya sektörü, özellikle buzdolapları ve derin dondurucular üzerine odaklanarak ve tedarik zinciri optimizasyonu için uygulanabilir tahminleme yaklaşımları sunarak literatüre katkıda bulunmaktadır.

Anahtar Sözcükler: Zaman Serisi, Kümeleme Yöntemleri, Beyaz Eşya ve Ev Aletleri Sektörü, Satış Tahmini Modelleri, Dışsal Değişkenler

To My Dear Family

ACKNOWLEDGMENTS

First and foremost, I would like to sincerely thank my supervisors, Prof.Dr. Pekin Erhan Eren and Prof. Dr. Altan Koçyiğit for their constant support and guidance throughout this process. Their valuable advice and encouragement helped me at every stage of this study.

I am also grateful to my colleagues, who have kindly shared their ideas and assistance with me during this process. Their contributions and collaboration are immensely valuable to me.

I am deeply grateful to my dear friend Merve Pehlivan İlerisoy who stood by me in moments of exhaustion and discouragement, offering motivation and comfort when I needed it most.

I am deeply thankful to my mother, Aynur Güranlı, and my brother, Çağrı Demir, for always supporting and believing in me throughout my life. This study is a product of their years of dedication and belief in me.

I would also like to thank my father, Ali Osman Demir, for always supporting me, I can feel that he is watching over me somewhere and feeling proud.

Finally, I am profoundly grateful to my dear life partner, Buğra Kartbol, for his patience, constant motivation, and everlasting support. When we have handle with engagement, wedding planning, moving and starting new journey during the thesis process, he willingly took on extra responsibilities and always made me feel at ease. I am truly fortunate to have him by my side.

TABLE OF CONTENTS

ABSTRACT ..... **iv**

ÖZ ..... **v**

DEDICATION ..... **vi**

ACKNOWLEDGMENTS ..... **vii**

LIST OF NOTATIONS ..... **xvi**

CHAPTER

1. INTRODUCTION **1**

1.1. Motivation 1

1.2. Scope 2

1.3. Objective 2

1.4. Organization 2

2. BACKGROUND AND RELATED WORK **5**

2.1 Home Appliances Sector 5

2.2 Time Series Forecasting 6

2.2.1 Naive Model 8

2.2.2 ARIMA&SARIMA 8

2.2.3 LSTM 9

2.2.4 Prophet 10

2.3 Clustering Approaches 11

2.4 Evaluation Metrics 13

2.5 Related Work 15

3. METHODOLOGY & APPLICATION **21**

3.1 Problem Definition 22

3.2 Model Experiments with Similar Research 22

3.3 Data Collection and Preprocessing 23

3.4 Clustering Phase 26

3.4.1. Clustering Products 26

3.4.2 Clustering Countries 31
3.5 Feature Engineering 37
3.6 Modeling 42
3.6.1 Model Construction 42
3.6.2 Hyperparameter Optimization 45
3.7 Evaluation 47
4. RESULTS & DISCUSSION 49
4.1 Results and Comparison 49
4.2 Discussion 55
5. CONCLUSION 57
5.1 Summary and Findings 57
5.2 Theoretical Contributions 58
5.3 Practical Contributions 58
5.4 Limitations 59
5.5 Future Work 59
REFERENCES 61
APPENDICES 67
APPENDIX A THE EVALUATION RESULTS BASED ON PRODUCT AND PRODUCT-REGION GROUP 67
A.1 Evaluation Results Based on Product Group used by the company 67
A.2 Evaluation Results Based on Product-Region Group used by the company .. 68
A.3 Evaluation Results Based on Product Group Analysis Using Only Historical Sales 70
A.4 Evaluation Results Based on Product Group Analysis Using Exogenous Variables 71
A.5 Evaluation Results Based on Product-Region Group Analysis Using Only Historical Sales 72
A.6 Evaluation Results Based on Product-Region Group Analysis Using Exogenous Variables 73
APPENDIX B THE EVALUATION RESULTS ON THE TRAINING SETS . 75
B.1 The Evaluation Results on The Product Sales Dataset 75
B.2 The Evaluation Results on the Product-Region Sales Dataset 76
APPENDIX C THE STL DECOMPOSITION RESULTS 77

C.1 The Seasonal Decomposition Graphs for Product Category Cluster 77
C.2 The Seasonal Decomposition Graphs for Product Category Cluster-Sales Region Cluster 77

LIST OF TABLES

Table 1: Summary of Selected Sales Forecasting Studies 15
Table 2: Model Performance Comparison. Adapted from (Haque, Amin, & Miah, 2023) 19
Table 3: Utilized Python Libraries and Their Usage Purposes 21
Table 4: Sales Data Sample 23
Table 5: Descriptive Statistics of Negative Sales Records 24
Table 6: Country Basis Sales Data Sample 32
Table 7: Combined Dataset Descriptive Statistics 38
Table 8: The inflation rate calculation for the product dataset 39
Table 9: VAR Analysis for Product Clusters 40
Table 10: The hyperparameters range of the models 45
Table 11: Optimization Time of Forecasting Models 46
Table 12: Forecasting Approaches and Utilized Features 49
Table 13: The evaluation results on the product groups defined by the company 51
Table 14: The evaluation results on the product-region groups defined by the company 51
Table 15: The model results on the Product Sales Test Dataset by using only historical sales data 52
Table 16: The model results on the Product Sales Test Dataset by using different exogenous variables 52
Table 17: The model results on the Product-Region Sales Test Dataset by using only historical sales data 53
Table 18: The model results on the Product-Region Sales Test Dataset by using different exogenous variables 54
Table 19: Average sales in the test set for each product group 67
Table 20: The model results based on each product group used by the company 67
Table 21: Average sales in test set for each product-region group 68
Table 22: The model results based on each product-region group used by the company 69
Table 23: The model results based on product group by using historical sales data. 71
Table 24: The model results based on product group by using exogenous variables 71
Table 25: The model results based on the product-region group by using historical sales data 72
Table 26: The model results based on the product-region group by using exogenous variables 73
Table 27: The model results on the Product Sales Training Dataset by using only historical sales data 75
Table 28: The model results on the Product Sales Training Dataset by using different exogenous variables 75

Table 29: The model results on the training set of Product-Region Sales Dataset by using only historical sales data 76
Table 30: The model results on the training set of the Product-Region Sales Training Dataset by using different exogenous variables 76

LIST OF FIGURES

Figure 1: Major Home Appliances Production in 2021 by Country. Adapted from (Investment Office, 2024) 6
Figure 2: LSTM Cell Structure (R. Bhushan Gopaluni, 2023) 10
Figure 3: K-Means Clustering (Chawla, 2024) 12
Figure 4: Agglomerative Clustering Approach (Yadav, 2020) 13
Figure 5: Quarterly Amazon Sales Revenue. Adapted from (Singh, Kumar, Sharma, & Sharma, 2020) 16
Figure 6: Sales Trend for Article 2 (Zunic, Korjenic, Hodzic, & Donko, 2020) 17
Figure 7: The thesis methodology 22
Figure 8: Cooling/Freezer Sales Change (2015–2023) 25
Figure 9: Quarterly sales by years 25
Figure 10: Yearly sales by sales regions 26
Figure 11: Elbow Method and Average Silhouette Scores for Product Clustering 27
Figure 12: Average Yearly Sales by Product Clusters 27
Figure 13: Silhouette Plot for Product Clusters 27
Figure 14: PCA Graph for Product Clustering for First Approach 28
Figure 15: t-SNE Graph for Product Clustering for First Approach 28
Figure 16: Average Silhouette Scores for different cluster numbers and Silhouette Plot with 4-clusters 29
Figure 17: PCA Graph for Product Clustering for Second Approach 29
Figure 18: t-SNE Graph for Product Clustering for Second Approach 30
Figure 19: Sales Trend for Product Clustering 31
Figure 20: Elbow Method and Average Silhouette Scores for Country Clustering 32
Figure 21: Silhouette Plot for Country Clusters 33
Figure 22: Average Yearly Sales by Country Clusters 33
Figure 23: PCA Graph for Sales Region Clustering for First Approach 34
Figure 24: t-SNE Graph for Sales Region Clustering for First Approach 34
Figure 25: Average Silhouette Scores for different cluster numbers and Silhouette Plot with 4-clusters 34
Figure 26: PCA Graph for Sales Region Clustering for Second Approach 35
Figure 27: t-SNE Graph for Sales Region Clustering for Second Approach 35
Figure 28: Sales Trend for Sales Region Clustering for Second Approach 36
Figure 29: Correlation Analysis between Sales Amount and Exogenous Variables for Product Sales Data 41
Figure 30: Correlation Analysis between Sales Amount and Exogenous Variables for Product-Region Sales Data 41
Figure 31: Parallel Coordinates Graph for LSTM 46
Figure 32: Parallel Coordinates Graph for SARIMA 47
Figure 33: Parallel Coordinates Graph for Prophet 47
Figure 34: Average Predicted vs. Actual Sales Across Product Category Clusters 52

Figure 35: Average Predicted vs. Actual Sales Across Product Category-Sales Region Clusters 54
Figure 36: STL Decomposition for Product Category Clusters..... 77
Figure 37: STL Decomposition for Product Category-Sales Region Clusters..... 78

LIST OF ABBREVIATIONS

APAC	Asia-Pacific
ARIMA	Autoregressive Integrated Moving Average
ARIMAX	Autoregressive Integrated Moving Average with Exogenous Variables
CA	Central Asia
CPI	Consumer Price Index
ETS	Exponential Smoothing
EWMA	Exponentially Weighted Moving Average
GDP	Gross Domestic Product
ICS	Index of Consumer Sentiment
IQR	Interquartile Range
LightGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory Neural Network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEA	Middle East and Africa
mMAPE	Modified Mean Absolute Percentage Error
MSE	Mean Squared Error
NWRMSLE	Normalized Weighted Root Mean Squared Logarithmic Error
PCA	Principal Component Analysis
PE	Prediction Error
$\boldsymbol{R^2}$	Coefficient of Determination (Goodness of Fit)
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Variables
SKU	Stock Keeping Unit
SMAC	Sequential Model-Based Algorithm Configuration
SMAPE	Symmetric Mean Absolute Percentage Error
t-SNE	t-Distributed Stochastic Neighbor Embedding
TPE	Tree-Structured Parzen Estimator
WA_MAE	Weighted Average of Mean Absolute Error
WA_RMSE	Weighted Average of Root Mean Squared Error
WA_SMAPE	Weighted Average of Symmetric Mean Absolute Percentage Error
XGBM	Extreme Gradient Boosting Model

LIST OF NOTATIONS

AIC(k)	Akaike Information Criterion for a model with k-lags
Ct	The total number of reported COVID-19 deaths globally in month t
Cumulative Contribution	The cumulative contribution up to product p, expressed as a percentage.
Ec,t	Economic indicator values set for country c in month t
Er,t	Economic indicator values set for sales region cluster r in month t
ϵ_t	The error term, accounting for randomness or unexplained variations in the data.
g(t)	The trend function, capturing the long-term increase or decrease in the data.
GDPGrowth r,t	GDPGrowth rate for sales region cluster r in month t
h(t)	The holiday and special event effects function, representing the impact of specific events.
Inflation r,t	Inflation rate for sales region cluster r in month t,
Interest rate r,t	Interest rate for sales region cluster r in month t
IQR_w	The interquartile range (IQR) calculated for the data in the w-th window.
MAE_p	The Mean Absolute Error for the p-th product group.
p,d,q	Non-seasonal ARIMA parameters: autoregressive, differencing, and moving average orders.
P, D,Q,s	Seasonal ARIMA parameters: seasonal autoregressive, differencing, moving average, and seasonal period.
$Q1_w$	The first quartile (25th percentile) of the data in the w-th window
$Q3_w$	The third quartile (75th percentile) of the data in the w-th window.
$RMSE_p$	The Root Mean Squared Error for the p-th product group.
s(t)	The seasonality function, modeling periodic patterns in the time series.

$\mathbf{Sales_i}$	Total sales amount of i-th product cluster
$\mathbf{Sales_{r,t}}$	Total sales amount of sales region cluster in month t
Sales(t-lag)	The sales value at time t−lag
Sales(t+(12-lag))	The sales value shifted forward by (12−lag) periods from time t.
$\mathbf{Sales_Lag_{lag}(t)}$	The sales value at time t−lag, representing the lagged sales data by a specific number of periods (lag)
$\mathbf{SMAPE_p}$	The Symmetric Mean Absolute Percentage Error for the p-th product group.
Stock(t)	The stock value at time t
Stock p,r,t	The stock level for product p, sales region r in month t,
Stock(t-lag)	The stock value at time t−lag
$\mathbf{Stock_Lag_{lag}(t)}$	The stock value at time t−lag, representing the lagged stock data by a specific number of periods (lag)
Unemployment r,t	Unemployment rate for sales region cluster r in month t
$\mathbf{WAEI_{i,t}}$	Weighted Average of i-th economic indicator value in month t
$\mathbf{w_i}$	The weight assigned to the i-th observation
$\mathbf{w_p}$	The weight representing the average sales of the p-th product group.
$\mathbf{x_i}$	The i-th observation within the w-th window, representing an individual data point being evaluated for outlier detection.
$\mathbf{y_i}$	The actual value of i-th observation
$\mathbf{\bar{y}_i}$	The mean value of all actual values
$\mathbf{y(t),\ \hat{y}_i}$	The predicted value at time t or i-th observation

CHAPTER 1

INTRODUCTION

Supply chain is a complex system that includes many steps varying from the production of the raw materials of the products to the delivery of the product to the end customer (Power, 2005). Sales forecasting is crucial for business operations to deal with increasing inventory costs and managing deviations in demand and fluctuations in economic conditions (Gustriansyah, Ermatita, & Rini, 2022). Furthermore, the development of the technology has changed the supply chain mechanism, making data-driven decision-making more critical. Additionally, the volume of data has grown significantly, and this has led to the need for efficient data management and analysis within supply chains. Therefore, by analyzing sales patterns and accurately predicting the demand with a data-driven approach, businesses can determine their procurement, production, and inventory strategies to meet market needs without overstocking or understocking.

1.1. Motivation

Fluctuations in global economies and technological development have rapidly affected business operations and forced companies to change their business practices. Especially, the Covid-19 pandemic has had significant effects not only from a health perspective but also on the global economy and trade. During the pandemic, supply chain disruptions, interruptions in international trade, and sudden changes in consumer habits have led to major transformations in the retail sector (Dmitry, 2020). Many companies have started to adopt strategic innovations during this period, making digital transformation inevitable and forcing them to restructure their operational processes (McKinsey Global Institute, 2021).

Another remarkable development during this period has been the increase in the competitive advantage of Chinese companies. China is one of the countries that quickly alleviated the effects of the pandemic and gained a strategic advantage in the global market by reducing product costs. Chinese manufacturers leveraged their low-cost advantage to expand market share in various sectors and have become key players in global competition (World Economic Forum, 2022). China's vast production capacity and low-cost labor have increased competitive pressure on other countries during this time. In this competitive environment, it has become more crucial than ever for companies to remain sustainable in the market, direct their strategic investments effectively, and manage their budgets accurately (Sun, Lee, & Yang, 2021).

In complex and competitive environments, companies need to anticipate future demand and develop strategies based on accurate sales forecasts. However, in the home appliance sector, data-driven forecasting methods are still underutilized, and there is limited research specifically focused on demand forecasting for this market. My motivation for this thesis is to employ a data-driven approach to enhance forecasting accuracy and supply chain strategies for the mentioned company and to enrich the existing literature for this market.

1.2. Scope

This study utilizes real-world data from a global home appliance company, which is Europe's largest home appliance manufacturer, selling to more than 100 countries with more than 50,000 employees. Cooling and freezer products have been selected as the primary focus of this research. These products are categorized based on their product type, such as single-door, double-door, multi-door, etc., and their widths by the marketing and sales team. The monthly sales data have been obtained with this breakdown, and it includes detailed information on the countries where the products were sold. The analysis encompasses sales across all markets rather than focusing on specific countries or regions. Additionally, the study did not delve into the SKU-level details within each product category.

1.3. Objective

This thesis aims to construct a demand forecasting model for a home appliances company to deal with supply chain complexity, demand variation and decrease the risks of poor supply chain management. Furthermore, it aims to improve the company's clustering approach by regrouping products and countries based on their historical sales patterns and develop forecasting models at the product-country cluster level. This will also support the implementation of cluster-specific supply chain strategies, such as common promotional campaigns for similar product groups, standardized sales strategies across countries having similar sales trends, and product-country-specific inventory planning strategies to ensure product availability for expected demand.

Accurate and specialized demand forecasting allows companies to allocate their resources, reduce costs, optimize budget planning, and improve overall supply chain efficiency. Moreover, it leads to an increase in customer satisfaction and gaining competitive advantage (Hamiche, Abouaïssa, Goncalves, & Hsu, 2018). Developing this model is key to the company's long-term success, helping it adapt to changing market conditions and supporting sustainable growth.

1.4. Organization

This thesis consists of four chapters and sections under them. The first chapter is the introduction which gives brief information about the problem, the motivation, scope, and objective of the research. The second chapter focuses on background and related work; it consists of different sections related to the home appliance sector, time series

forecasting methods, and approaches used in the study and similar studies in the literature. The third chapter focuses on the followed methodology and application of the models. It also includes a comparative analysis of the models and their respective results. The final chapter concludes the thesis and summarizes the findings. Moreover, it presents the limitations of the study and aspects that are open to improvement and proposes recommendations for future research.

CHAPTER 2

BACKGROUND AND RELATED WORK

This chapter provides an overview of the study through three main sections. The first section introduces the home appliance sector. The second section introduces time series forecasting and details the specific models used in the thesis. Additionally, the clustering techniques and evaluation metrics that are utilized for assessing the model performance are explained. The final section provides an overview of studies in the literature related to the topic, presenting their approaches and the methods they used.

2.1 Home Appliances Sector

The home appliance sector stands out as a rapidly growing and changing industry worldwide. This sector includes major -white goods- and minor appliances, major appliances include refrigerators, washing machines, dishwashers, and ovens, which fall under the category of durable consumer goods. The sector's growth is driven by factors such as increasing urban populations, rising household incomes, technological advancements, and improved living standards. Technological innovations in the white goods sector have promoted the widespread adoption of smart devices and energy-efficient products, becoming one of the key factors shaping consumer demand. The proliferation of white goods with IoT features has been one of the factors transforming the market's growth dynamics. As of 2023, the global white goods market has experienced an average annual growth rate of 4-5%, reaching a market size of approximately $700 billion (Statista, 2024).

The Asia-Pacific region emerges as the largest market in the global white goods sector. Countries such as China, India, and South Korea make significant contributions to global growth with their large populations and rapidly increasing household incomes. China is the world's largest producer and exporter of white goods sector. The country's high production capacity, low labor costs, and strong supply chains enable it to play an effective role in the global market. India is boosting demand in the sector with its rapidly growing middle class and increasing urbanization rates. In developed markets such as Europe and North America, demand for white goods is largely driven by the need for replacements and energy-saving expectations. In these regions, environmentally friendly and energy-efficient products are preferred as part of efforts to reduce energy consumption. Government-implemented energy efficiency incentives also contribute to consumer preferences for such products (Market Research Future, 2024).

Turkey has a significant position in the global home appliance sector, ranking as the second-largest producer in the world, as shown in Figure 1. The country benefits from its geographic proximity to Europe, low labor costs, and strong industrial infrastructure, which provide a competitive advantage. Home appliance manufacturers in Turkey are developing energy-efficient and environmentally friendly products, thanks to their high production capacity and advanced R&D capabilities. The Turkish home appliance sector has an annual average production capacity of 33 million units and generates revenue of approximately $20 billion (Investment Office, 2024).

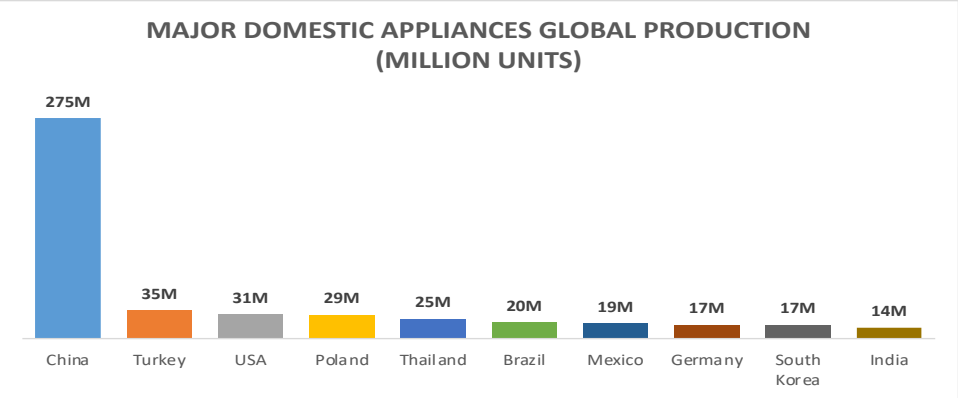


Figure 1: Major Home Appliances Production in 2021 by Country. Adapted from (Investment Office, 2024)

2.2 Time Series Forecasting

Time series forecasting is a part of predictive analytics focused on estimating future values of a variable by considering and interpreting the patterns in the historical data. A time series defines a time-ordered sequence of data collected or recorded at regular intervals such as hourly, daily, monthly, annually, etc. This technique is applied in many different industries, such as finance, economics, weather prediction, stock market analysis, energy consumption forecasting, etc. (Kolambe, 2024; Qiwei, Boylan, Huijing, & Labib, 2018)

There are key concepts related to time series data, and before predicting future values, these concepts should be clarified. One of them is the trend, which reflects the long-term movement in the data, which may either rise, fall, or be steady over the period. For instance, the gradual increase in sales of a product over several years indicates an upward trend. In contrast, seasonality refers to short-term, recurring patterns in the data that are related to specific calendar events, such as increased retail activity during the holiday season or an increase in electricity consumption in the summer months. Another vital element is cyclicality, which captures non-calendar-based fluctuations, such as economic cycles of growth and decline. Additionally, another term is noise, meaning the random, unpredictable variations in the data which doesn't align with patterns of trends or seasonality (Panwar, 2021).

Time series forecasting can be categorized into two main types: univariate and multivariate forecasting. In univariate forecasting, predictions are made using only a single time series. The model learns from the historical values of a single variable to forecast future points. On the other hand, multivariate forecasting involves predicting a target variable using not only its historical values but also exogenous variables that can influence the target. These models are generally more complex but can potentially provide better accuracy, especially in real-world applications where multiple factors influence the outcome (Wang & Aviles, 2023).

Time series forecasting is utilized in numerous disciplines. In the financial domain, it assists in predicting variables like stock prices, interest rates, and exchange rates. Economists leverage forecasting to anticipate key indicators such as GDP growth, inflation, and unemployment trends. Retailers apply time series analysis to forecast consumer demand and manage their inventory efficiently, while energy providers perform it to predict electricity or gas consumption. Even in healthcare, time series models are used to forecast patient numbers and predict the spread of diseases. Therefore, a time-series forecasting strategy can be adapted to many varying fields using compatible data and variables (Kolambe, 2024; Zonglei, Ma, Li, Wang, & Li, 2023).

Time series forecasting models can broadly be classified into three categories: Statistical Models, Machine Learning Models, and Hybrid Models. Statistical models represent the traditional, mathematically grounded approach to time series forecasting. These models are based on classical statistical techniques, relying on rigid assumptions about the data, such as stationarity, linearity, and the presence of trends. When these assumptions are satisfied, statistical models tend to perform well and provide reliable forecasts. However, their effectiveness diminishes when faced with more complex or non-linear patterns within the data, where their rigid assumptions may no longer be applicable (Rao, 2015).

Machine learning models adopt a more flexible approach compared to traditional statistical methods when analyzing the structure of data. This approach allows them to capture non-linear relationships, account for outliers, and model more complex temporal patterns. These models rely on data-driven approaches, which enable the incorporation of a broader set of features or external variables, enhancing their capacity to handle the complexities of real-world time series data (Parmezan, Souza, & Batista, 2019).

Hybrid models integrate the strengths of both statistical and machine learning methodologies. By combining the interpretability and time-efficiency of statistical models with the ability to capture complexity and work with big data of machine learning techniques, hybrid models aim to decrease the limitations in either approach when used in isolation. The main goal is to use the strengths of both to improve forecasting performance in various scenarios (Khashei & Bijari, 2012).
Based on this categorization, various time series forecasting models have been employed in this study. In the following section, these models -Naive, SARIMA,

LSTM, and Prophet- are introduced by explaining each model's structure, key parameters, and their applications.

2.2.1 Naive Model

A naive model is a basic forecasting technique that uses simple assumptions to predict future values based on historical data. The key feature of a naive model is that it doesn't seek to capture complex patterns, trends, or seasonality. Instead, it serves as a baseline to compare other models. In this thesis, the Simple Naive Model is used as the baseline model, and this approach assumes that the next value in the time series will be the same as the most recent observed value (Svetunkov, 2023).

2.2.2 ARIMA&SARIMA

ARIMA and SARIMA are widely considered powerful tools in time series analysis. Each model has its unique features and applications, but to use them effectively, it's significant to first understand the concept of stationarity. A time series is considered stationary when its statistical properties, such as the mean and variance, remain constant over time. This is a crucial assumption for many statistical time series models, especially ARIMA-based ones. If the data is non-stationary, it must be transformed before applying these models. There are common methods for providing stationarity, including differencing, logarithmic transformations, and detrending.

The ARIMA model is one of the most used approaches in time series forecasting. ARIMA consists of three main components: Autoregressive (AR), Integrated (I), and Moving Average (MA). The AR part models the time series based on its past values, assuming that previous observations can help predict future values. The Integrated component involves differencing the data to make it stationary, which is necessary for the model to function properly. Finally, the Moving Average component captures the error terms or residuals, modeling them as a linear combination of past errors. ARIMA is particularly well-suited for univariate time series without seasonal patterns, where past values and residuals are key factors in forecasting future values (Ariyo, Adewumi, & Ayo, 2014).

ARIMAX, or ARIMA with Exogenous Variables, builds on the ARIMA model by including external factors that can influence the target variable. ARIMAX maintains the same AR, I, and MA components as ARIMA, but by considering external factors, it can enhance the accuracy of forecasts when external factors play a significant role in the dynamics of the data (Djimasbe, Gyamfi, & Donatus, 2024).

The SARIMA model, or Seasonal ARIMA, adds another layer of complexity by incorporating a seasonal component. This model is particularly useful for time series data that exhibit recurring patterns over fixed periods, such as sales or temperature data following seasonal trends. SARIMA extends ARIMA by adding seasonal autoregressive, moving average, and differencing terms, which allow the model to capture both short-term fluctuations and seasonal cycles and these seasonal terms work with the standard ARIMA components.

SARIMAX combines the features of both SARIMA and ARIMAX, making it the most flexible. It can handle time series that exhibit both seasonal patterns and are influenced by exogenous variables. This makes SARIMAX especially useful for complex forecasting situations where, for example, a time series might follow a seasonal pattern while also being impacted by external factors like economic indicators or weather conditions. By incorporating both seasonal dynamics and external influences, SARIMAX offers a comprehensive solution for challenging time series forecasting tasks. (Alharbi & Csala, 2022).

2.2.3 LSTM

Long Short-Term Memory (LSTM) networks are a specific type of recurrent neural network (RNN) designed to handle sequences of data. LSTMs have an inherent memory structure, and it helps maintain important information over Extended time periods in contrast to standard RNNs failing in problems such as vanishing or exploding gradients on larger sequences. This property makes LSTMs well-suited for time-series forecasting, natural language processing (NLP), and speech recognition, where maintaining context over long sequences is crucial. This property allows them to identify long-term dependencies, and as a result, LSTMs are widely used in many fields, such as finance, healthcare, and autonomous systems. Hence, various LSTM model variants have been developed to solve specific challenges in sequential data processing and improve performance across different tasks (Houdt, Mosquera, & Napoles, 2020).

The Vanilla LSTM, considered the simplest form, consists of a single LSTM layer followed by a dense output layer. It effectively captures basic sequence dependencies, so it is suitable for uncomplicated tasks. In contrast, the Stacked LSTM model adds multiple LSTM layers, increasing the model's capacity to learn more complex patterns and interactions over longer sequences. There is also the Bidirectional LSTM, which processes input sequences both forward and backward, capturing dependencies from past and future contexts. Additionally, the Encoder-Decoder LSTM architecture, generally used in sequence-to-sequence tasks like neural machine translation, has separate encoder and decoder components to process input and output sequences of varying lengths. Finally, the LSTM with an Attention Mechanism selectively weighs different input timesteps, making it useful in tasks like text summarization and image captioning (Dudukcu, Taskiran, & Taskiran, 2021).

The core of LSTMs lies in their three main gates: the forget gate, input gate, and output gate, each designed to handle information flow in a different way. The forget gate determines which past information should be discarded, using a sigmoid activation function to output values between 0 and 1 for controlled retention. Meanwhile, the input gate decides what new information should be added to the memory by using a combination of sigmoid and tanh activations to update the cell state. Finally, the output gate determines which information from the current memory state should be passed to the next hidden state. As demonstrated in Figure 2, this structure allows LSTMs to selectively filter information, enabling them to manage both short-term and long-term dependencies more effectively than standard RNNs (Gopaluni, 2023).

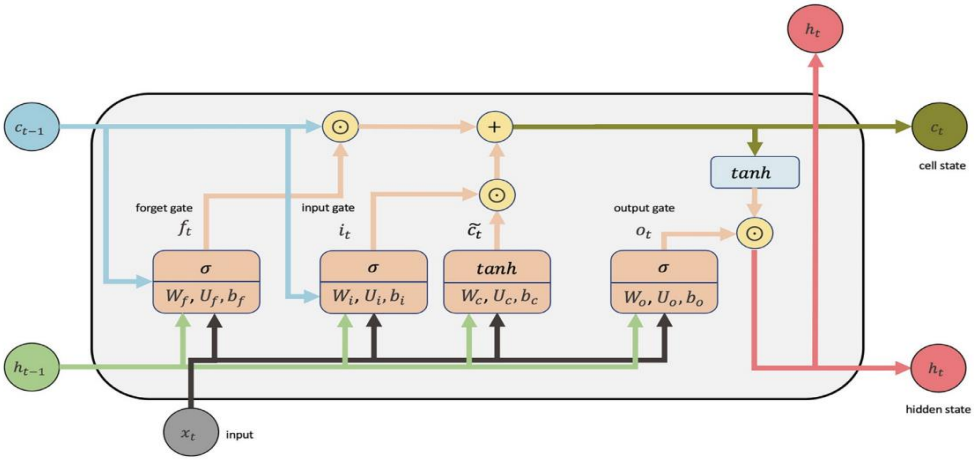


Figure 2: LSTM Cell Structure (R. Bhushan Gopaluni, 2023)

Some preprocessing steps are required before passing your dataset to an LSTM model. The data should be structured into sequences; there are different methods, such as sliding window, sequence-to-sequence, etc. For time series forecasting problems, the sliding window method is the most common approach. In this method, each window contains past observations used to predict the next value. Sequences are created by moving the window across the data, where each sequence includes selected timesteps as input features (X), and the next time step is determined as the target (y). Then, the sequences are divided into training and testing that respect order in time (Brownlee, 2019).

The performance of an LSTM model largely depends on selecting suitable hyperparameters, such as the number of layers, number of neurons, learning rate, batch size, sequence length, and regularization methods. Increasing the number of layers can help the model learn more complex patterns, but it also raises the risk of overfitting. Similarly, adding more neurons per layer can improve the model's capacity to represent data but requires more computational resources. Regularization techniques like dropout and recurrent dropout help handle overfitting by randomly deactivating some neurons during training. Additionally, parameters like sequence length and batch size play crucial roles in capturing relevant temporal patterns while reducing noise and computational inefficiency (Nakisa, Rastgoo, Rakotonirainy, Maire, & Chandran, 2018).

In conclusion, LSTM networks are an effective approach for analyzing sequential data due to their ability to retain information over extended sequences and effectively extract long-term dependencies. By implementing proper data preprocessing steps and conducting careful hyperparameter tuning, it is possible to achieve optimal predictive performance.

2.2.4 Prophet

Prophet is an open-source forecasting tool that was developed by Facebook's Core Data Science team in 2017 to meet the need for an interpretable and scalable

forecasting framework that could adapt to time series with strong seasonality and missing data. The main motivation behind Prophet's development is to provide a robust model that non-experts could use, making accurate forecasting processes accessible for different industries. Unlike many other time series models, Prophet adopts an additive approach, which allows it to efficiently capture various seasonal patterns, holidays, and trend shifts, making it particularly well-suited for forecasting tasks in business, finance, and other sectors. The model can be mathematically expressed as follows (Hyndman & Athanasopoulos, 2021).

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

Prophet can be used for different purposes, varying from forecasting demand to planning capacity, predicting revenues, and managing resources. It is especially popular in industries such as retail, logistics, and finance, where understanding the impact of factors like seasonality and holidays is critical for planning and decision-making. The model can effectively handle missingness and irregularity in time series data; moreover, it is possible to reflect the impact of external variables on the model and be able to generate more interpretable and accurate forecasts (Jha & Pande, 2021).

The Prophet model is built on an additive decomposition framework by separating a time series into three primary components: trend, seasonality, and holidays. The trend component, g(t) represents the overall upward or downward trajectory of the series and can be modeled by using a linear or logistic growth model. The seasonality component, s(t) captures regular periodic fluctuations within the data, such as weekly or yearly patterns, often represented using the Fourier series. The holidays component, h(t) which is optional, adjusts for special events that cause temporary shifts in the series. This structure, combined with automatic changepoint detection, improves the model's ability to manage sudden changes and complex patterns in the data. (Daraghmeh, Agarwal, Manzano, & Zaman, 2021).

Prophet offers a wide range of hyperparameters to improve model performance and adjust the forecasts to specific situations. The most common parameters include changepoint_prior_scale, which controls the sensitivity to changepoints in the trend, and seasonality_prior_scale, which adjusts the contribution of seasonal patterns. The holidays_prior_scale parameter allows users to control the influence of holiday effects, while the seasonality_mode can be set to either 'additive' or 'multiplicative', depending on whether the seasonal patterns increase linearly or exponentially with the trend. Additionally, users can determine forecast horizons, specify custom seasonality, and adjust uncertainty intervals; it provides extensive flexibility to adjust the model's behavior based on the data's characteristics (Meta Open Source, 2024).

2.3 Clustering Approaches

Clustering is an unsupervised learning method used to group similar data points based on their features without labels. It seeks to uncover meaningful structures by identifying the underlying processes, characteristics, and groupings within the data.

There are two main types of clustering algorithms: partitional and hierarchical clustering. Partitional methods divide data into non-overlapping groups, while hierarchical methods build a tree-like structure of nested clusters. K-means clustering is one of the most popular partitional algorithms. It aims to divide the data into K clusters, where each point belongs to the cluster with the closest mean, called the centroid.

The algorithm follows an iterative process. First, K centroids are randomly initialized; second, each point is assigned to the nearest centroid based on Euclidean distance; third, centroids are recalculated as the mean of points in each cluster; and finally, steps two and three are repeated until convergence. K-means is computationally efficient and works well with large datasets, making it a favorable choice for especially simple clustering tasks. However, it assumes that clusters are spherical, have similar sizes, and are evenly spaced, so it is less effective when dealing with overlapping or irregularly shaped clusters (B., George, Manikandan, & Thomas, 2020).

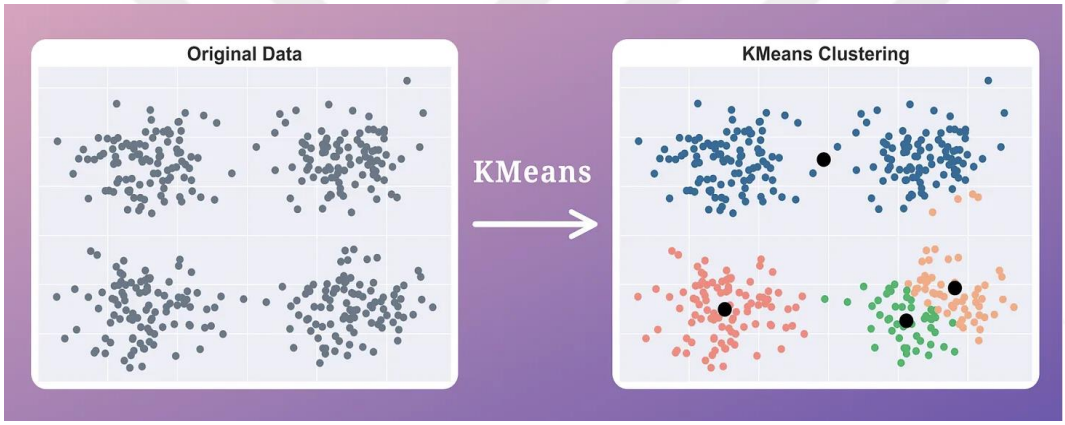


Figure 3: K-Means Clustering (Chawla, 2024)

Agglomerative Clustering, a type of hierarchical clustering, operates differently by starting with each data point as separate clusters. It then iteratively merges the closest clusters based on a specified linkage criterion. There are four different linkage methods such as single linkage which considers the minimum distance between members of the two clusters; complete linkage which relies on the maximum distance between members of the two clusters; average linkage which is based on the average of all distances between members of the two clusters and lastly ward linkage which considering minimizing the total within-cluster variance. This process continues until a predefined number of clusters is reached or all points are merged into a single cluster. Agglomerative Clustering can work well with data with hierarchical relationships or clusters of different sizes and densities (B., George, Manikandan, & Thomas, 2020; Tokuda, Comin, & Costa, 2022).

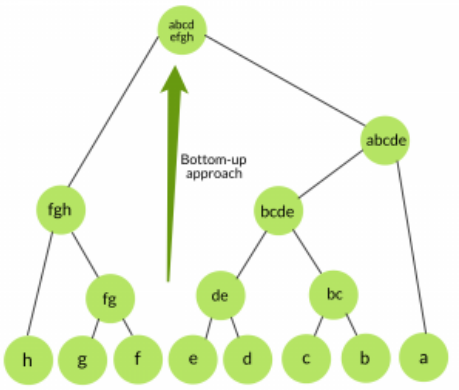


Figure 4: Agglomerative Clustering Approach (Yadav, 2020)

Determining the optimal cluster number is significant, and there are different ways to do it; two common techniques are used in this thesis. The Elbow Method and the Average Silhouette Score. The Elbow Method plots the sum of squared distances, called inertia, between data points and their assigned centroids for different values of K. As K increases, the total inertia decreases, but there is a point referred to as the elbow where the rate of decrease slows significantly. This elbow point suggests the ideal number of clusters, achieving a balance between cluster compactness and complexity. On the other hand, the Silhouette Score measures how similar a point is to its own cluster compared to others, ranging from -1 to 1. A higher score indicates that data points are well-matched to their clusters and don't fit well with neighboring clusters. By calculating the average silhouette score for various values of K, the number of clusters that maximize this score can be identified as the optimal number of clusters (Kodinariya & Makwana, 2013).

Another important point is the choice of distance metric in clustering. Euclidean distance calculates the straight-line distance between points, and it is effective for detecting geometric relationships within clusters. On the other hand, cosine distance measures similarity by focusing on the angle between vectors rather than their magnitude. This orientation-based approach is practical when the relative direction of data points are more significant than their absolute values (Korenius, Laurikkala, & Juhola, 2007).

In summary, K-Means is effective for quick partitioning tasks, especially with compact and well-separated clusters, while Agglomerative Clustering offers greater flexibility for identifying complex structures and hierarchical relationships. Both algorithms rely on feature vectors to represent data and use distance metrics that are compatible with the data's nature and clustering aim. The choice of algorithm, linkage criteria, cluster number decision, and distance metric should be compatible with the dataset's characteristics and the specific clustering objectives.

2.4 Evaluation Metrics

Evaluation metrics are used to assess the performance of predictive models by measuring the accuracy and reliability of forecasts. They offer a quantitative way to

compare different models and identify which model yields the best performance. In time series forecasting, evaluation metrics determine how closely the model's predictions align with the actual values. There are distinct metrics in the literature, and some variations of these metrics are specified to the problem. The metrics used in this thesis are RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and SMAPE (Symmetric Mean Absolute Percentage Error); each has its own strengths and specific scenarios where it is more suitable.

RMSE calculates the square root of the average squared differences between predicted and actual values. It penalizes larger errors more than smaller ones, making it sensitive to outliers. RMSE is especially useful when larger errors are more problematic, as it emphasizes significant deviations. However, its sensitivity can be a disadvantage when evaluating models on datasets with varied scales because it can be disproportionately influenced by large errors. Despite this, RMSE is widely used due to its straightforward interpretation and compatibility with models optimized using least squares methods (Chai & Draxler, 2014). RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \quad (2)$$

where $\hat{y}_i$ denotes the prediction of i-th observation and y_i denotes the actual value of i-th observation

Furthermore, MAE is the average of the absolute differences between predicted and actual values. Unlike RMSE, it does not square the errors, making it less sensitive to outliers. MAE provides a more linear representation of errors, so it is more suitable when all errors have equal importance, regardless of their magnitude. MAE is easy to interpret and provides an intuitive measure of model accuracy, reflecting the average size of the prediction error and it is calculated as follows:

$$MAE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} |\hat{y}_i - y_i|} \quad (3)$$

where $\hat{y}_i$ denotes the prediction of i-th observation and y_i denotes the actual value of i-th observation

The final metric mentioned is SMAPE, which is a percentage-based metric that calculates the average of absolute percentage errors between predicted and actual values. It is adjusted symmetrically to prevent extreme errors from dominating the metric. It's beneficial when evaluating models across different scales, as it provides a relative measure of accuracy. SMAPE is well-suited for datasets where relative errors matter more than absolute ones, such as in sales forecasting, where percentage differences can be more meaningful. However, SMAPE can be sensitive to values near zero, making it less reliable when actual values are small (Menna, Traini, & Cortellessa, 2024). This metric is formulated as follows:

$$SMAPE = \frac{100}{N} \times \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \quad (4)$$

where $\hat{y}_i$ denotes the prediction of i-th observation and y_i denotes the actual value of i-th observation

2.5 Related Work

In this section, similar works about sales forecasting are elaborated. Although there is extensive research on time series forecasting in the literature, sales forecasting papers are analyzed to align with the purpose of the thesis, and a comprehensive review of these is conducted. Table 1 summarizes the selected studies on sales forecasting, including variable types, domain of the study, evaluation metrics, and models used.

Table 1: Summary of Selected Sales Forecasting Studies

	Article	Feature	Domain	Evaluation Metrics	Models
1	Singh, Kumar, Sharma, & Sharma, 2020	Time series feature: Quarterly sales (Amazon/2005-2017)	E-commerce	MAPE	Exponential Smoothing SARIMA Neural Network Autoregression Model
2	Zunic, Korjenic, Hodzic, & Donko,2020	Time series feature: Monthly and quarterly sales (A retail company/ 2014-2018)	Retail	PE MAPE	Prophet
3	Weng, Liu, & Xiao, 2020	Time series feature: Daily sales (Rossman/2013-2015) Weekly sales (Jollychic e-commerce platform/2017-2018) Daily sales (Corporacion Favorita Grocery/2013-2017) Statistical features Categorical features	Retail/E-commerce	NWRMSLE	LSTM LightGBM Combined Model
4	Haque, Amin, & Miah, 2023	Time series feature: Daily sales (Walmart/2011-2016) Statistical features Categorical features External features	Retail	RMSE MAE	Lasso regression Ridge regression XGBM LightGBM Decision Tree

Table 1 (cont.)

5	Duan & Dong, 2024	Time series feature: Monthly sales (Air conditioner company/2018-2021) Statistical features Categorical features External features	Home Appliance Manufacturing (Air Conditioners)	MAE RMSE MSE R2	LSTM RF XGBM
6	Bandara, Shi, Bergmeir, & Hewamalage, 2019	Time series feature: Daily sales (Walmart/January 2018-June 2018) Categorical features Calendric features	E-commerce	mMAPE	LSTM ARIMA EWMA

As mentioned before, there are different approaches and different model options in the time series forecasting field. In most of the papers, traditional and deep learning models' performance are compared, and the ensemble models can be chosen also. Moreover, one of the time series forecasting approaches that exists in the literature is univariate forecasting which only uses historical sales data in the forecasting process. Singh et al. (2020) utilize Amazon's historical sales data and conduct a comparison of traditional and machine learning models. The study aims to forecast Amazon's quarterly sales, and the 2005-2018 quarterly net sales were used. The sales data has seasonality, which can be seen below; the revenue generated on Black Friday is generally more than 10 times the median sales of the year. Also, the Christmas period's sales are higher, so sales peak in Q4 (Singh, Kumar, Sharma, & Sharma, 2020). Seasonality and its quarterly pattern are similar to the sales behavior of refrigerators.

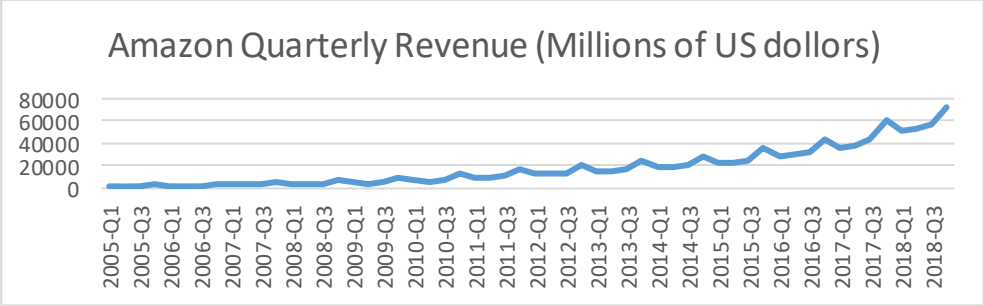


Figure 5: Quarterly Amazon Sales Revenue. Adapted from (Singh, Kumar, Sharma, & Sharma, 2020)

For sales forecasting, Singh et al. (2020) applied three different models: Holt-Winters Exponential Smoothing, ARIMA, and the Neural Network Autoregression Model. The first model considers trend and seasonality, using multiplicative level, additive trend, and additive seasonality by ETS (M, A, A). Because the data isn't stationary, differentiating is performed by using the d parameter in ARIMA. Additionally, the SARIMA model is used to take into account seasonality, and one seasonal differencing is applied. Finally, a Neural Network Autoregression model is conducted, and 20 feed-

forward networks are used, with each network featuring one input layer, one hidden layer, and one output layer. For the comparison of the model performance, the MAPE metric is used. The most accurate results are achieved using the Seasonal ARIMA (SARIMA) model, likely due to the seasonality present in the data.

Another example of the univariate forecasting model is related to forecasting the sales of the retail company. Zunic et al. (2020) use the data collected from the production environment in one of the biggest retail companies in Bosnia and Herzegovina. Daily sales data are aggregated into monthly sales. The products are classified into four groups based on the length of the sales data availability, such as inactive products or best-selling products. As shown in Figure 6, sales data contain seasonality; therefore, Facebook's Prophet algorithm- known for handling time series with strong seasonality and non-linear trends- is used.

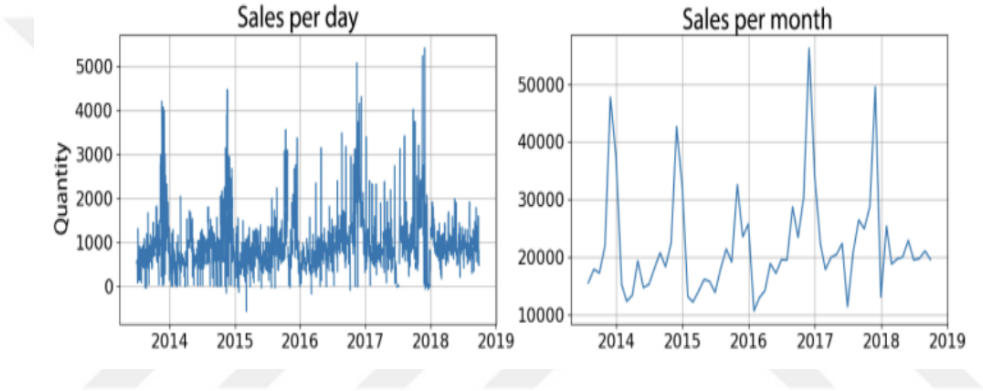


Figure 6: Sales Trend for Article 2 (Zunic, Korjenic, Hodzic, & Donko, 2020)

The study employs PE and MAPE as evaluation metrics. The evaluation results indicate that the framework can achieve reasonably accurate forecasts, with approximately 50% of products showing a MAPE below 30% for monthly forecasts and about 70% achieving the same accuracy for quarterly forecasts. Approximately 40% of the product portfolio can be forecasted with MAPE < 15% on a quarterly basis, and these are mostly the best-selling items, making up over 80% of the total annual sales. Based on those facts, the obtained results are significant for the company's profitability (Zunic, Korjenic, Hodzic, & Donko, 2020). This article directs this thesis to adopt a similar approach by focusing on the products that account for the majority of sales while excluding products that represent the minority of sales and increasing variety.

Furthermore, a multivariate approach can be chosen; in this approach, different types of exogenous variables can be utilized. These exogenous variables can be analyzed in four subgroups. The first one is statistical features extracted from the time series data, like the total sales volume every seven days, a median of the sales, etc. The second one is the external factors affecting the target factor, like economic indicators, inventory level, etc. The third one is categorical variables related to the nature of the data, like product type, sales country, etc., and the final one is calendric features like the day of the week, the month of the year, etc.

Weng et al. (2020) adopt multivariate time series forecasting when constructing the forecasting model. Their study utilizes three different store sales data to evaluate the performance of sales forecasting models. Weng et al. (2020) state that traditional linear and statistical models' performance is worse than deep-learning models. Therefore, they have decided to use deep-learning models. Because RNN causes long-term dependence problems in long-period data sets, LSTM is chosen. Moreover, LightGBM is a Gradient Boosting Machine framework, and it results faster than XGBM and yields higher accuracy, so it is used for forecasting sales. These time series forecasting models are applied both individually and in combination. The exogenous variables used can be categorized into two groups: statistical features (mean, median, standard deviation, etc.) and categorical features (state, type, etc.). To evaluate the performance of the models, Weng et al. (2020) use NWRMSLE (Normalized Weighted Root Mean Squared Logarithmic Error) as the evaluation metric. This metric avoids penalizing large differences in prediction when both the predicted and the actual values are large, and it is formulated as follows:

$$\text{NWRMSLE} = \sqrt{\left(\sum_{i=1}^{N} w_i(\ln(\hat{y}_i + 1) - \ln(y_i + 1))^2\right) / \sum_{i=1}^{N} w_i} \quad (5)$$

where w_i represents the weight assigned to the i-th observation $\hat{y}_i$ denotes the prediction of the i-th observation and y_i denotes the actual value of the i-th observation

According to the results, the proposed hybrid model outperforms separate model usage scenarios. In terms of running time, LightGBM is faster than other models, and LSTM and the combination of the two models' running times are nearly equivalent. Therefore, in terms of both accuracy and efficiency, combining these two models is more advantageous because the model combines the interpretability of LightGBM with the LSTM's ability to capture time series patterns (Weng, Liu, & Xiao, 2020).

Haque et al. (2023) also adopt the multivariate forecasting approach, and they explore the integration of macroeconomic variables into retail demand forecasting models to enhance prediction accuracy. Consumer Price Index (CPI), Index of Consumer Sentiment (ICS), and unemployment rates are chosen as economic variables to understand complex consumer behavior more effectively. Furthermore, categorical features related to products like the product category, statistical features like rolling average sales, and calendric features like a day of the week are used. Five machine learning models -Lasso regression, Ridge regression, XGBM, LightGBM, and Decision Tree regression- are performed and the models are trained on Walmart sales data (Haque, Amin, & Miah, 2023).

For all models, the important features are determined, and 15 features are selected between 131 possible features set. RMSE and MAE are chosen as model evaluation metrics. The models are performed using selected economic variables and without using economic variables. According to the results in Table 2 using economic factors improves the model performance, and economic conditions are also considered

significant for refrigerator sales. Therefore, this article guides this thesis in terms of the use of evaluation metrics and the determination of a method for selecting exogenous variables.

Table 2: Model Performance Comparison. Adapted from (Haque, Amin, & Miah, 2023)

	Performance of Models			
	RMSE		MAE	
Model	Without Economic Variables	With Economic Variables	Without Economic Variables	With Economic Variables
Lasso	1.80	1.79	0.89	0.88
Ridge	1.73	1.73	0.85	0.85
LGBM	1.71	1.72	0.85	0.85
XGBM	1.72	1.72	0.84	0.84
Decision Tree	2.35	2.36	1.00	1.00

Duan and Dong (2024) utilize the multivariate approach to forecast air conditioner sales in China and use over 100,000 data points from January 2018 to December 2022. Four feature categories are utilized: time features like sales month, product attributes like weight, color, etc., target volume trend features like inventory level, and environmental factors referring to market environment like competitor market share. The mutual information method is applied to determine which interacting features have the most significant impact on sales, thereby allowing the model to focus on the most significant predictors.

Three different models are performed: LSTM, RF, and XGBM. Because LSTM can capture temporal dependencies, it is effective for identifying long-term trends and seasonality in time series data. RF handles complex non-linear relationships, and XGBM manages high demand fluctuations. The authors use these models separately and in combination with these models. MAE, MSE, RMSE, and R^2 are chosen as evaluation metrics. The results indicate that the combined model outperforms single models, demonstrating improved accuracy and stability (Duan & Dong, 2024).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\bar{y}_i - y_i)^2} \quad (6)$$

$\hat{y}_i$ denotes the prediction of i-th observation, y_i denotes the actual value of i-th observation and $\bar{y}_i$ denotes the mean value of all actual values

Bandara et al. (2019) also prefer the multivariate forecasting approach and utilize two distinct sales datasets from Walmart, consisting of e-commerce sales data collected in real time. Due to the real-time nature of the data, missing values are present, and a forward-filling strategy is employed for imputation. Products are categorized using two different methods: the first method relies on sales volume, creating three categories such as high, medium, and low sales volume. The second approach involved K-means clustering based on time series and sales-related features. Two different types

of exogenous variables are used: features related to products, like product type and product category, and calendric features, like weekday/weekend.

LSTM model is used, and different versions of LSTM are generated, such as multivariate LSTM, univariate LSTM, and LSTM trained on first grouping strategy, and second strategy, etc. Bayesian optimization and SMAC (Sequential Model-Based Algorithm Configuration) are adopted for the models' hyperparameter optimization. After the hyperparameter optimization, models' performances are compared with ARIMA, EWMA, and the naive model using the mMAPE evaluation metric (Equation 7). Results demonstrate that the LSTM model outperformed the other models in terms of accuracy (Bandara, Shi, Bergmeir, & Hewamalage, 2019). This article provides a good example of using clustering methods for forecasting sales, and a similar approach is applied in the thesis.

$$\text{mMAPE} = \frac{1}{m} \times \sum_{i=1}^{m} \left(\frac{|\hat{y}_i - y_i|}{1 + y_i} \right) \tag{7}$$

$\hat{y}_i$ denotes the prediction of i-th observation and y_i denotes the actual value of i-th observation

In summary, there are various examples of univariate and multivariate forecasting in the literature. Even if seasonality exists in the data, univariate forecasting can perform well if there is a stable trend. However, if the trend itself is variable, incorporating external factors can improve the results. It is crucial to select these factors accurately, and different methods, such as mutual information, are proposed in the literature for factor selection. In terms of the forecasting models, the LSTM model is the most used in the literature, while SARIMA and Prophet models also succeed in capturing seasonality.

The contributions of this thesis to the literature are as follows. It focuses on the home appliance sector, specifically refrigerators and freezers, addressing a domain that remains underexplored in the literature. Rather than limiting the analysis to a single product, country, or aggregated monthly sales, the study incorporates a diverse range of products and sales regions, enriched by the integration of economic indicators from all countries where the products are sold, providing a more comprehensive perspective. While many studies conducted during the pandemic analyzed the impact of COVID-19 on sales, there is a significant gap in research examining the combined periods of pre-pandemic, pandemic, and post-pandemic. This thesis fills this gap by exploring how sales dynamics evolved across these periods, and the inclusion of external factors such as inventory levels further enhances the analysis, offering a detailed understanding of the influences on sales dynamics. Finally, although clustering techniques are widely used in the literature, this research introduces a different approach by proposing alternative clustering methods for a company with an existing grouping and actionable and data-driven recommendations for organizational strategies.

CHAPTER 3

METHODOLOGY & APPLICATION

The methodology, which is demonstrated in Figure 7, has been systematically designed to address the thesis objectives. It contains seven key stages. The first part includes the problem definition, which is crucial for accurately addressing the business needs. The second stage involves model experiments based on similar research, aiming to establish an initial framework for time series models. The third part focuses on data collecting and preprocessing. The next phase is clustering and feature engineering for the data analysis. This is followed by Modeling which involves constructing models and optimizing hyperparameters. Then, in the evaluation part, the performance of the models is compared based on the determined evaluation metrics and business objectives. Finally, in the deployment part, the best solution is analyzed and reported, and the thesis is documented. The process is iterative, allowing for continuous feedback and improvements. During this process, the Python environment has been used, and different libraries have been utilized; Table 3 explains the purposes for which these libraries are used. Documentation links are provided for some libraries used in the models and hyperparameter optimization, with footnotes below indicating the relevant library.

Table 3: Utilized Python Libraries and Their Usage Purposes

Library	Purpose
pandas	Used for data processing and manipulation
NumPy	Used for numerical computing and array operations
matplotlib and seaborn	Utilized for data visualization
optuna[1]	Applied for hyperparameter optimization
TensorFlow[2]	Employed for building the LSTM model
statsmodels[3]	Applied for statistical analysis and building the SARIMA model
sklearn	Used for data preprocessing and evaluation metrics
prophet[4]	Utilized for building Prophet model

[1] https://optuna.org/
[2] https://www.tensorflow.org/guide
[3] https://www.statsmodels.org/stable/user-guide.html
[4] https://facebook.github.io/prophet/docs/quick_start.html#python-api

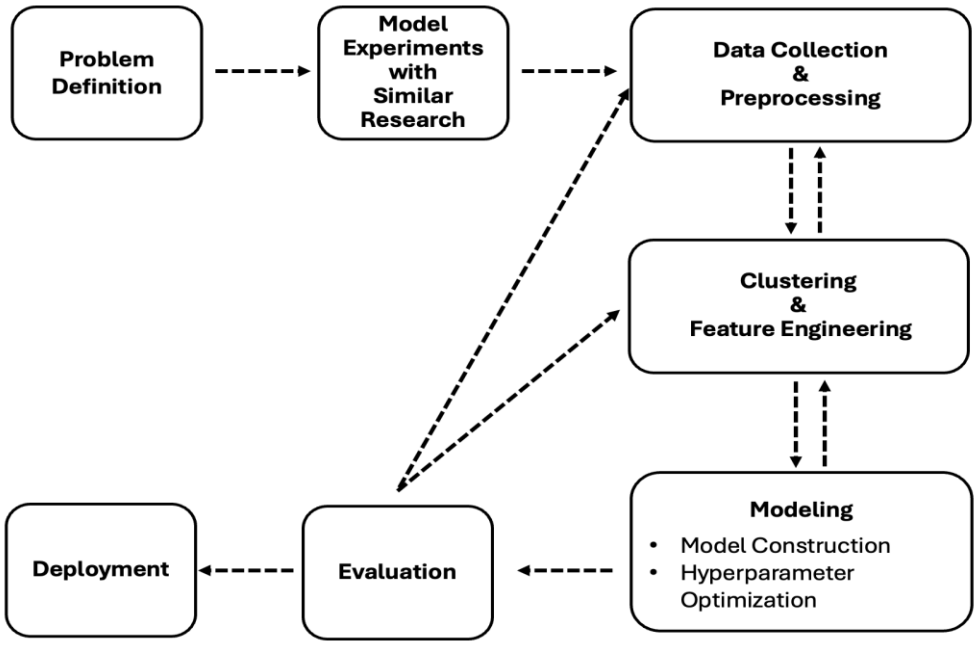


Figure 7: The thesis methodology

3.1 Problem Definition

As mentioned earlier, the home appliances industry, especially high-demanded products like refrigerators, faces significant challenges due to fluctuating consumer demand, market uncertainties, and the increasing complexity of supply chain operations. These challenges lead to problems such as inaccurate inventory planning, overproduction, stockouts, and wasted resources. These problems not only increase operational costs but also disrupt supply chain performance and affect customer satisfaction and market competitiveness. Therefore, for overcoming these challenges, constructing an accurate demand forecasting model and supply chain strategies have become a crucial need.

3.2 Model Experiments with Similar Research

Before analyzing the company's data, literature review has been conducted, and five related articles have been selected. These articles share two common features: they focus on sales forecasting and provide publicly available datasets. Additionally, they employ different time series models, which makes them useful resources for conducting initial model construction trials and obtaining insights.

The five datasets of the selected articles have been accessed through online resources, and an initial analysis has been conducted through data visualization to understand their underlying patterns, trends, and characteristics. Then, the preprocessing steps mentioned in the respective articles have been applied to datasets. Following this, various model trials have been performed on these datasets, leveraging the methodologies described in the articles. Model parameters have been adjusted and

fine-tuned to optimize performance. Additionally, appropriate evaluation metrics have been utilized for model comparisons. The models' performance has been assessed both within each dataset and across different datasets to evaluate their reliability and generalizability. This process has resulted in the creation of model structures that can be utilized effectively for the thesis problem.

3.3 Data Collection and Preprocessing

After the model trials with similar works in the literature, the mentioned home appliance company's data has been analyzed. The data used are categorized into two main types: internal company data and external factors data. The internal data obtained from the company includes sales data and stock levels, which are directly related to the forecasting problem. This data serves as the foundation for understanding historical sales demand and inventory change patterns.

The refrigerator sales dataset covers the period from 2015 to 2023, and it is the monthly breakdown. A sample of this dataset is demonstrated in Table 4.

Table 4: Sales Data Sample

	Year	Month	Country	Sales Region	Cooling/Freezer	Width	Product Hierarchy 3	Sales Amount	Quarter
0	2015	1	Albania	Europe	Cooling	53-59 cm	53-59 cm-Cooling	174	Q1
1	2015	1	Albania	Europe	Cooling	60-63 cm	60-63 cm-Cooling	74	Q1
2	2015	1	Albania	Europe	Cooling	70-73 cm	70-73 cm-Cooling	29	Q1
3	2015	1	Albania	Europe	Cooling	70-73 cm	70-73 cm-Cooling	40	Q1
4	2015	1	Algeria	Russia&CA&Caucasus&MEA	Cooling	60-63 cm	60-63 cm-Cooling	1530	Q1
5	2015	1	Australia	APAC&AHHA	Cooling	53-59 cm	53-59 cm-Cooling	560	Q1
6	2015	1	Australia	APAC&AHHA	Cooling	53-59 cm	53-59 cm-Cooling	234	Q1
7	2015	1	Australia	APAC&AHHA	Cooling	60-63 cm	60-63 cm-Cooling	205	Q1
8	2015	1	Australia	APAC&AHHA	Cooling	60-63 cm	60-63 cm-Cooling	19	Q1
9	2015	1	Australia	APAC&AHHA	Cooling	70-73 cm	70-73 cm-Cooling	597	Q1

The raw sales dataset contains nine columns, each representing key attributes used in the analysis, and the detailed explanations of the columns are as follows:

- **Year**: Represents the year in which the product was sold. The dataset contains data for nine years (2015–2023).
- **Month**: Represents the month of sales for the product, and the dataset includes 108 months of data.
- **Quarter**: Represents the quarter (Q1, Q2, etc.) in which the product was sold; the dataset covers 36 quarters.
- **Country**: Indicates the country where the product was sold.
- **Sales Region**: The company operates in five major sales regions, as defined by its sales and marketing teams:
    - Turkey & South Asia
    - Europe
    - Russian & CA & Caucasus & MEA

    - APAC
    - America
- **Cooling/Freezer**: This column distinguishes between two types of refrigerators: cooling and freezers.
- **Width**: The width of the products is a key classification criterion, with eight distinct width ranges, from 0–52 cm to 91–92 cm.
- **Product Hierarchy 3**: This column combines product width and the cooling/freezer type, a breakdown commonly used by the company for sales analysis, with 12 different hierarchies.
- **Sales**: The column contains the number of units sold per product in the indicated country, month and year. The dataset includes 48,192 rows, with sales values ranging from - to 52.042.

In addition to internal data, external factors have been incorporated to enhance the accuracy of the demand forecasting model. These include economic indicators and COVID-19-related data. Economic indicators, such as interest rate, inflation, unemployment rates, and GDP growth, were gathered from sources like Trading View, Trading Economics, and the national statistical websites of countries without data available on these platforms. For COVID-19, monthly death data have been collected from the "Our World in Data.org" platform. These external variables have been integrated to understand the broader economic and social factors influencing demand fluctuations.

Initially, the datasets were imported, and then a thorough check for missing values was conducted, revealing no missing data in the dataset. Similarly, a duplicate row check has been applied to ensure that there are no repeated entries. Then, since sales values are expected to be zero or greater, the data have been inspected for any negative values in the “Sales” Column, and 18 negative sales records were found; descriptive statistics of these 18 records can be seen in Table 5. These negative values have been identified as product returns. Since the proportion of negative sales records to the total number of rows (48192) is quite low, and the average value of negative sales is -14.7, these sales were considered negligible and set to zero.

Table 5: Descriptive Statistics of Negative Sales Records

	Year	Month	Sales
count	18	18	18
mean	2020.9	5,5	-14.7
std	0.8	3.3	32.2
min	2020	1	-135
25%	2020	3	-13
50%	2021	5	-2
75%	2021	9.5	-1
max	2022	10	-1

To better understand the dataset and its characteristics, visualizations have been created. The sales change between 2015-2023 in the breakdown of cooling and freezer

is demonstrated in Figure 8. Based on the graph, sales reached their peak in 2021, followed by a decline for both cooling and freezer groups in the following years.

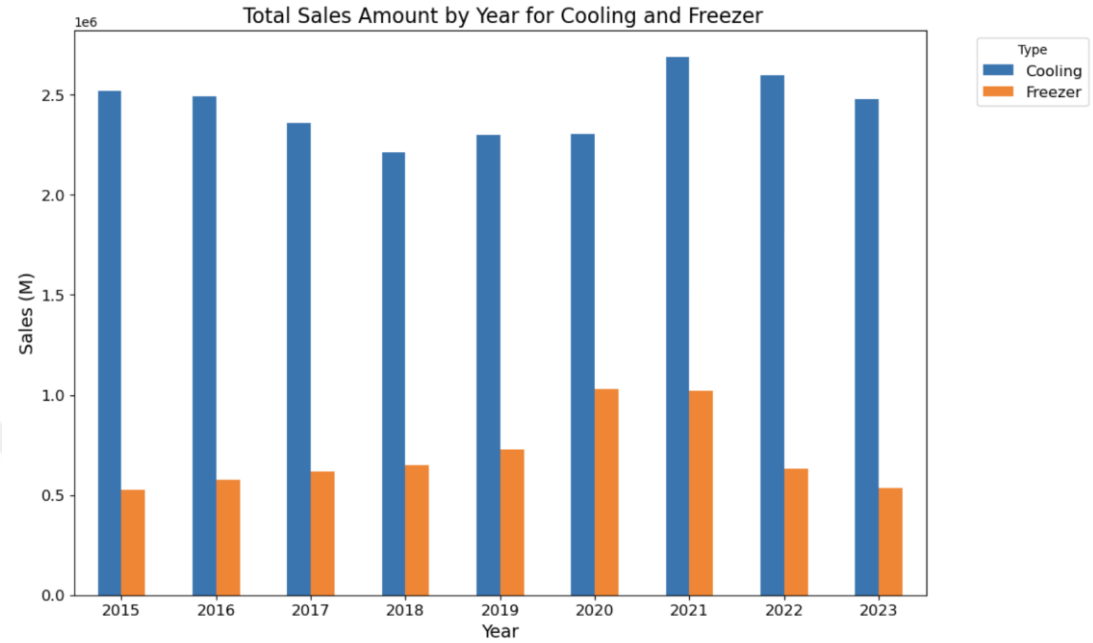


Figure 8: Cooling/Freezer Sales Change (2015–2023)

Figure 9 shows the quarterly sales trends, highlighting the seasonality in sales. Sales typically increase during Q2 and Q3, known as the high-demand periods, while sales in Q1 and Q4, referred to as the low-demand periods, tend to decrease.

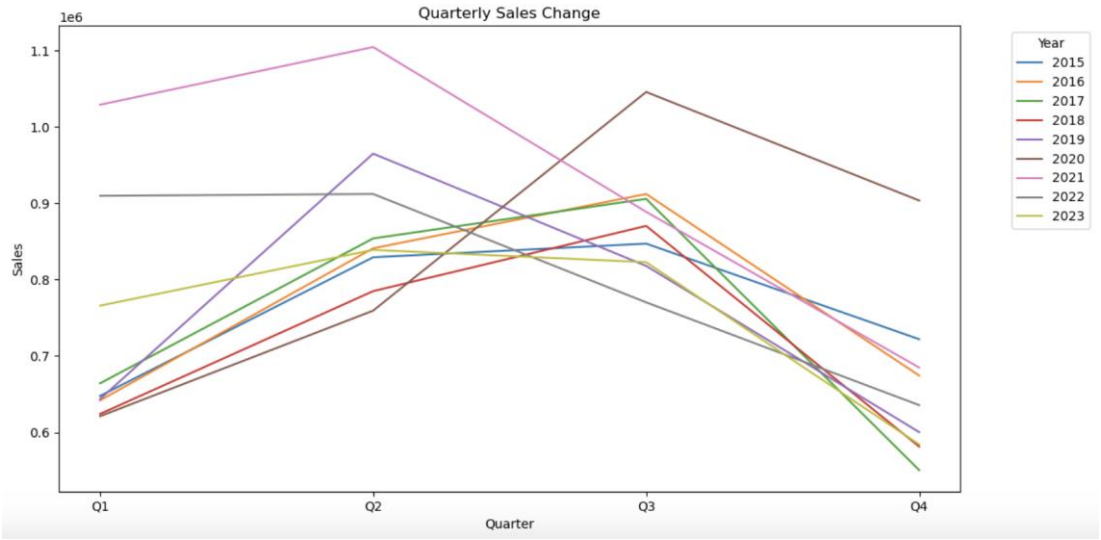


Figure 9: Quarterly sales by years

Also, as mentioned earlier, the dataset contains sales region information, and Figure 10 shows the yearly sales change across different sales regions. The company's key markets are Turkey & South Asia and Europe, which represent the majority of sales.

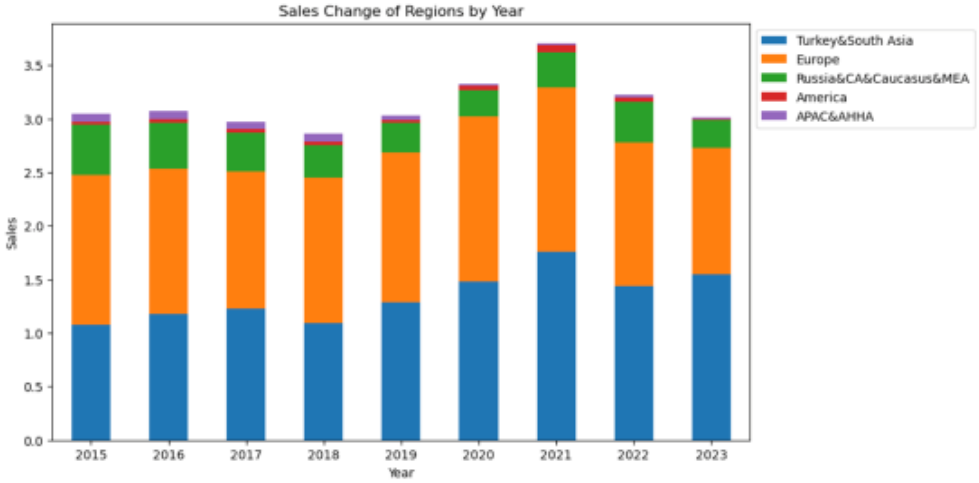


Figure 10: Yearly sales by sales regions

3.4 Clustering Phase

As mentioned earlier, the company analyzes sales data based on their classification criteria. It uses width range and sales organization categories. This section aims to explore whether different product or sales organization clusters can be used to improve forecasting accuracy and supply chain management.

3.4.1. Clustering Products

In this part, the sales dataset has been grouped by product category, and for each of the 12 product categories, a feature vector with 108 columns- each representing the sales for a specific month- has been created. Two product hierarchies have been identified as new product categories, one introduced in 2020, and the other in 2022, and their former sales records have been assigned a value of zero. Two clustering approaches have been applied for product categorization. The first approach categorizes products based on their monthly sales volume by using Euclidean distance, while the second focuses on clustering products according to their monthly sales volume by using cosine similarity.

The clustering and dimensionality reduction methods are utilized for both. For the former one, The K-means clustering method has been used, and the Elbow method has been employed to determine the optimal number of clusters. Based on the Elbow method graph, four or five clusters appear to be the most appropriate choices. Furthermore, the average silhouette score is plotted for each cluster number, and as observed in Figure 11, the optimal cluster number has been determined to be five with a Silhouette score of 0.56.

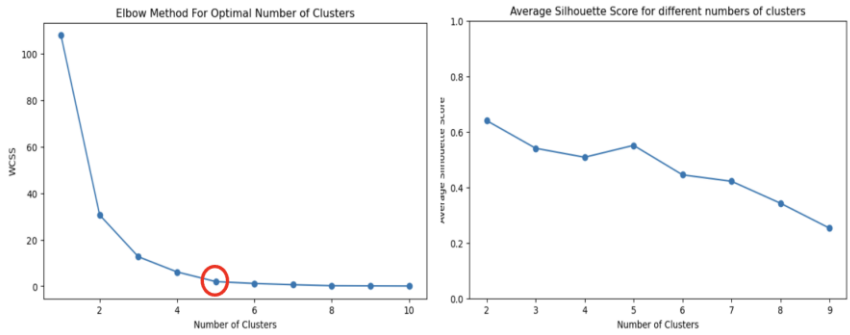


Figure 11: Elbow Method and Average Silhouette Scores for Product Clustering

The clusters created using this method are listed below. Additionally, Figure 12 demonstrates the average yearly sales amounts for each cluster, and it can be seen that Cluster 3 and 4 contains the high-demanded product hierarchies. Figure 13 plots the silhouette score for five clusters. The second and fourth clusters contain only one element, hence the Silhouette score could not be calculated for these clusters.

Cluster 1: 0-52 cm-Freezer, 70-73 cm-Freezer, 64-69 cm-Cooling
Cluster 2: 53-59 cm-Cooling
Cluster 3: 60-63 cm-Cooling, 60-63 cm-Freezer, 53-59 cm-Freezer
Cluster 4: 70-73 cm-Cooling
Cluster 5: 74-79 cm-Cooling,80-89 cm-Cooling,91-92 cm-Cooling,0-52 cm-Cooling

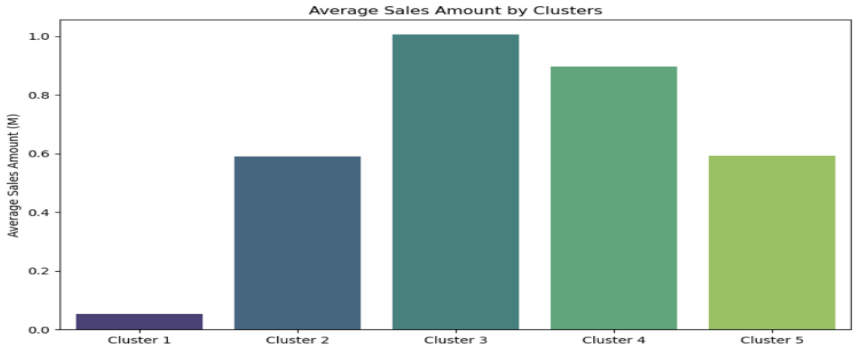


Figure 12: Average Yearly Sales by Product Clusters

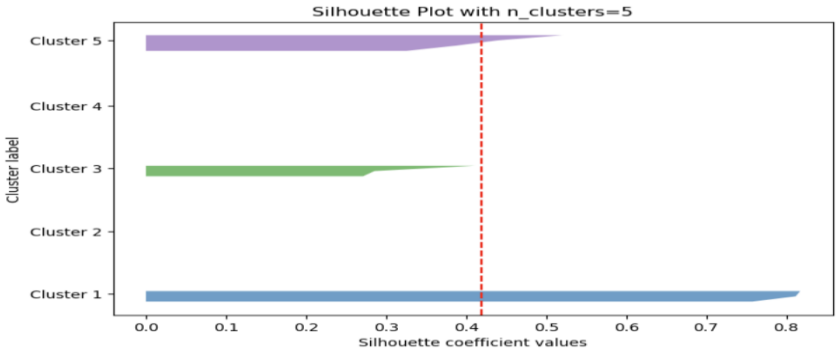


Figure 13: Silhouette Plot for Product Clusters

PCA and t-SNE visualization are used to analyze the separability of different clusters, shown in Figure 14 and Figure 15. In PCA, the separation between clusters is observed more clearly, especially with Cluster 1 and Cluster 4 being distinctly separated. In contrast, t-SNE better represents the intra-cluster distinctions; for example, the proximity and distribution of points within Cluster 1 and Cluster 5 can be seen more clearly.

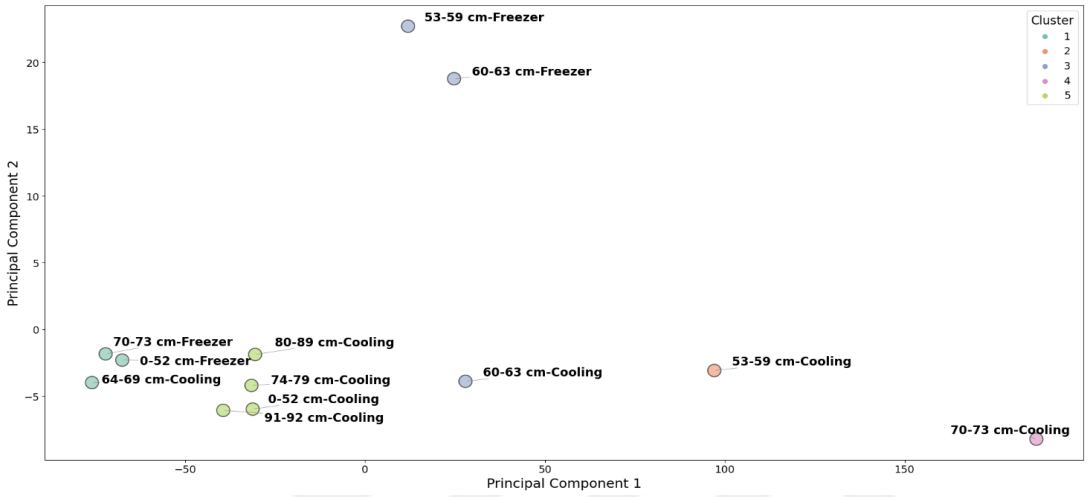


Figure 14: PCA Graph for Product Clustering for First Approach

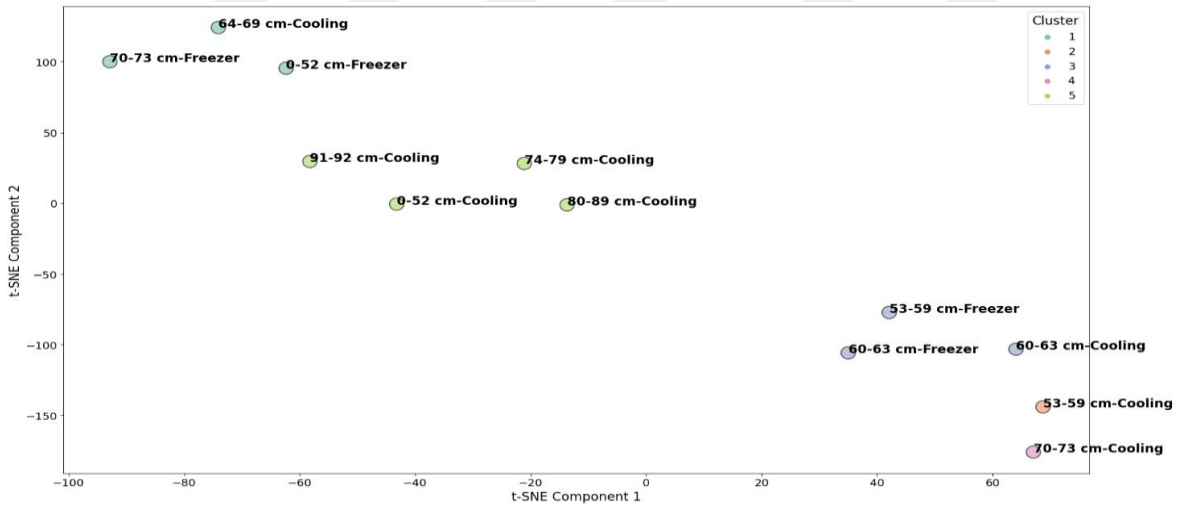


Figure 15: t-SNE Graph for Product Clustering for First Approach

In the second approach, the Agglomerative Clustering method has been applied with average linkage and cosine affinity. The cluster number has been chosen as four based on the average silhouette score plot in Figure 16. The silhouette score is 0.82 when the cluster number is set to 4, indicating strong clustering performance. The second and fourth clusters contain only one element, hence their silhouette scores could not be calculated for these clusters.

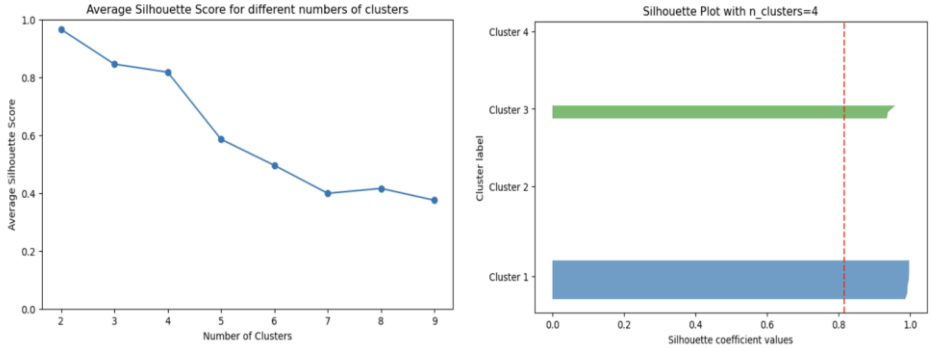


Figure 16: Average Silhouette Scores for different cluster numbers and Silhouette Plot with 4-clusters

The clusters created using this method are listed below and PCA and t-SNE visualizations can be shown in Figure 17 and Figure 18. In PCA, the separation between clusters, especially with Cluster 1 and Cluster 4 is observed more clearly. In contrast, t-SNE better represents the intra-cluster distinctions; for example, the proximity and distribution of points within Cluster 1 can be seen more clearly.

Cluster 1: 74-79 cm-Cooling, 80-89 cm-Cooling, 91-92 cm-Cooling, 0-52 cm-Freezer, 0-52 cm-Cooling, 70-73 cm-Freezer, 64-69 cm-Cooling
Cluster 2: 60-63 cm-Freezer
Cluster 3: 53-59 cm-Cooling, 60-63 cm-Cooling, 70-73 cm-Cooling
Cluster 4: 53-59 cm-Freezer

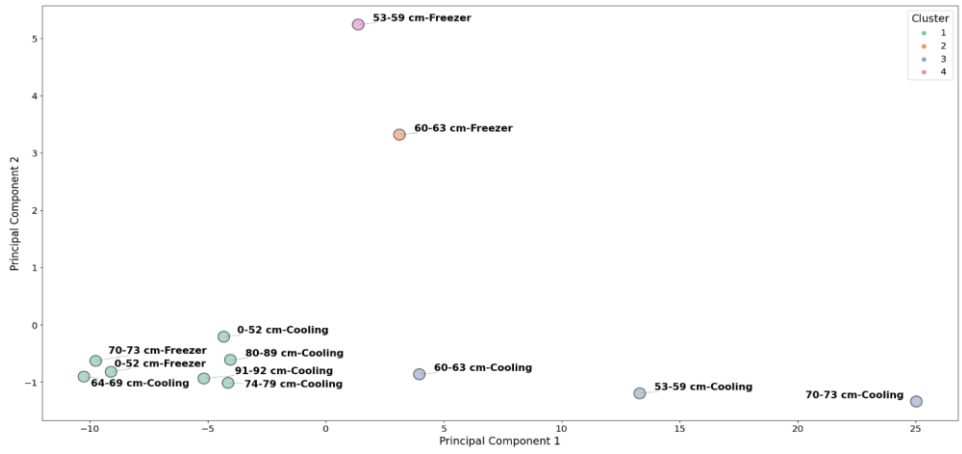


Figure 17: PCA Graph for Product Clustering for Second Approach

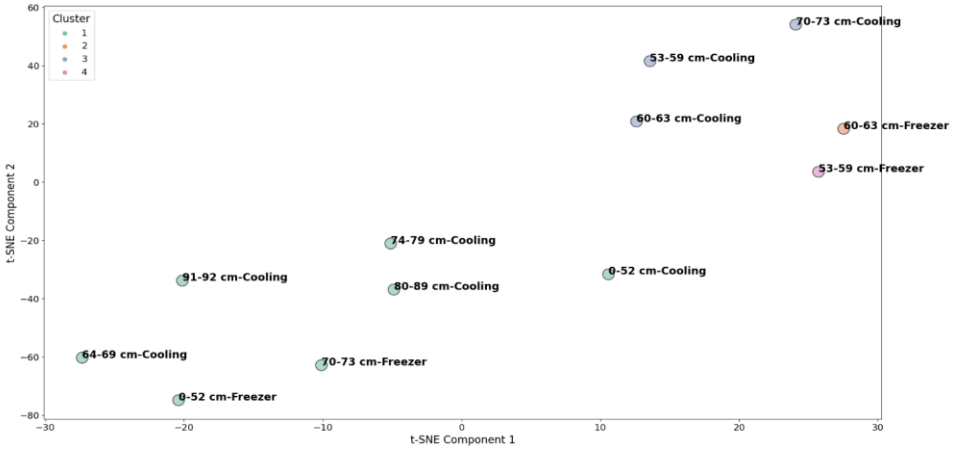
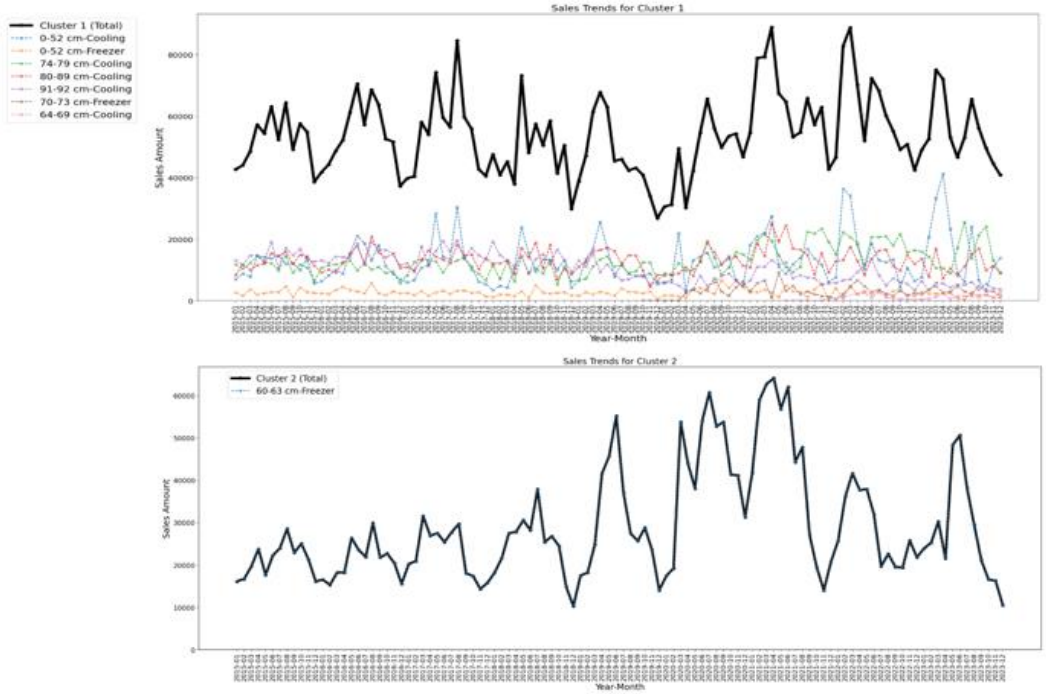


Figure 18: t-SNE Graph for Product Clustering for Second Approach

The monthly sales trend of clusters with the "Product Hierarchy 3" breakdown is represented in Figure 19. According to the graph, the products showing similar trends are grouped in the same cluster. For example, cluster 3 includes the product groups having steady sales trends.



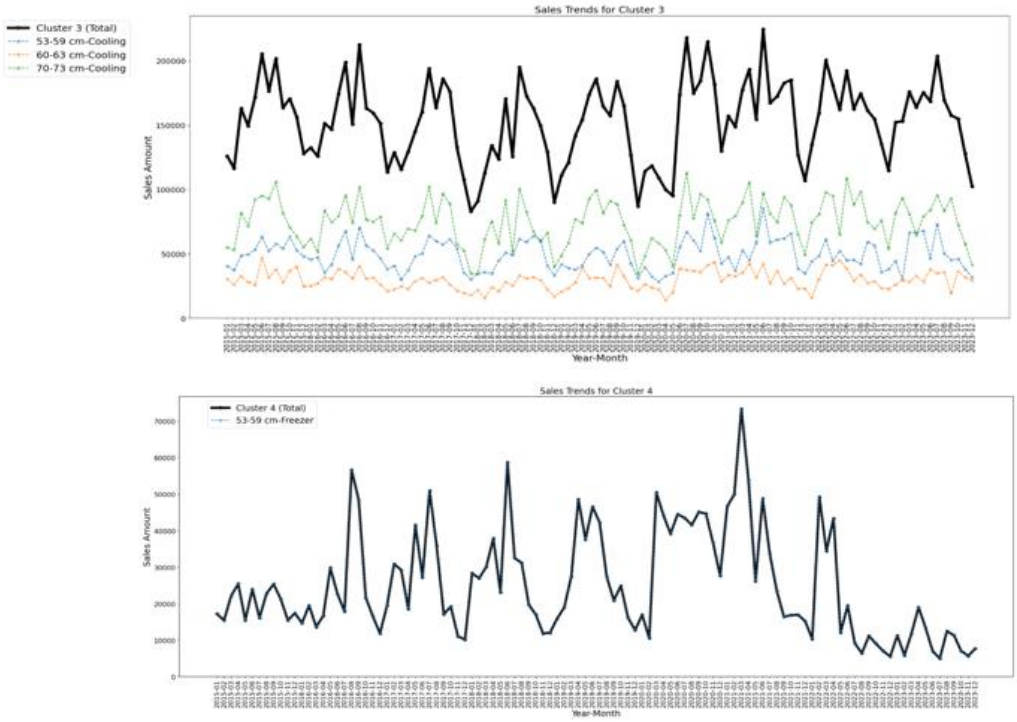


Figure 19: Sales Trend for Product Clustering

The clustering approaches have been evaluated with four experts, each specializing in different areas: marketing, sales, demand planning, and product management. The marketing expert, who has been working at this company for 12 years, has been consulted to evaluate the alignment of product clusters with market segmentation and customer preferences. From the sales perspective, the expert with 9 years of experience has provided feedback on how the clusters could represent sales transitions and align with product dynamics. The supply chain specialist has been working for 14 years, and she has reviewed the clusters to see how they could improve forecasting accuracy and support inventory management. Additionally, a product management expert with 8 years of experience has reviewed the clusters to assess how well they match product development strategies and potential innovation opportunities. Based on these multidisciplinary evaluations, the second clustering approach has been found to be more effective. Because this clustering method groups products based on their sales trends, the variation in their demand patterns also becomes similar. This leads to improved forecasting accuracy, as products with similar demand fluctuations can be predicted more effectively. Also, it is more compatible with product market dynamics and can capture product transitions compared to the first approach. Also, it enables the application of similar promotional campaigns, product launch timing, and stocking strategies for similar products.

3.4.2 Clustering Countries

In this part, the sales dataset has been grouped by country, and for each of the 127 countries, a feature vector with 108 columns- each representing the sales for a specific month- has been created. In all feature vectors, months without sales were imputed

with a value of zero. Two clustering approaches mentioned above have been applied. Both approaches utilize the monthly sales volume of countries, and the sales data by country breakdown is presented in Table 6.

Table 6: Country Basis Sales Data Sample

	Year	Month	Country	Sales Region	Cooling/Freezer	Product Hierarchy 4	Width	Product Hierarchy 3	Sales Amount	Quarter
0	2015	1	Albania	Europe	Cooling	53-59 cm - Çift Kapılı Dondurucu Üstte - NF	53-59 cm	53-59 cm-Cooling	174	Q1
1	2015	1	Albania	Europe	Cooling	60-63 cm - Çift Kapılı Dondurucu Altta - NF	60-63 cm	60-63 cm-Cooling	74	Q1
2	2015	1	Albania	Europe	Cooling	70-73 cm - Çift Kapılı Dondurucu Altta - NF	70-73 cm	70-73 cm-Cooling	29	Q1
3	2015	1	Albania	Europe	Cooling	70-73 cm - Çift Kapılı Dondurucu Üstte - NF	70-73 cm	70-73 cm-Cooling	40	Q1
4	2015	1	Algeria	Russia&CA&Caucasus&MEA	Cooling	60-63 cm - Çift Kapılı Dondurucu Üstte - ST	60-63 cm	60-63 cm-Cooling	1530	Q1
5	2015	1	Australia	APAC&AHHA	Cooling	53-59 cm - Çift Kapılı Dondurucu Üstte - NF	53-59 cm	53-59 cm-Cooling	560	Q1
6	2015	1	Australia	APAC&AHHA	Cooling	53-59 cm - Tezgah seviyesi buzdolabı	53-59 cm	53-59 cm-Cooling	234	Q1
7	2015	1	Australia	APAC&AHHA	Cooling	60-63 cm - Çift Kapılı Dondurucu Üstte - NF	60-63 cm	60-63 cm-Cooling	205	Q1
8	2015	1	Australia	APAC&AHHA	Cooling	60-63 cm - Tek Kapılı Buzdolabı - NF	60-63 cm	60-63 cm-Cooling	19	Q1
9	2015	1	Australia	APAC&AHHA	Cooling	70-73 cm - Çift Kapılı Dondurucu Üstte - NF	70-73 cm	70-73 cm-Cooling	597	Q1

For the first approach, the K-means clustering method has been used and Euclidean distance has been applied. To determine the optimal number of clusters, the Elbow method has been employed, and average silhouette scores have been plotted for different cluster numbers. while three or four clusters appeared to be appropriate, four clusters were selected to avoid having too low number of clusters.

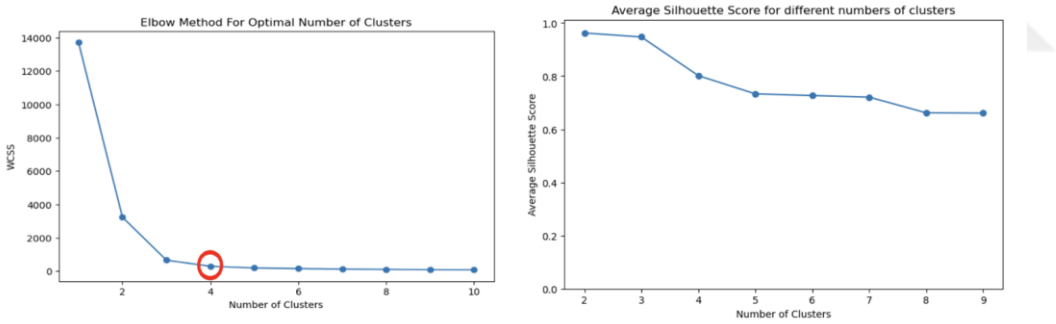


Figure 20: Elbow Method and Average Silhouette Scores for Country Clustering

When the cluster number is set to 4, the Silhouette score is calculated as 0.80, indicating strong clustering performance. Silhouette scores for clusters are plotted below. Because the cluster 2 and 3 have one element, their silhouette scores couldn't be calculated.

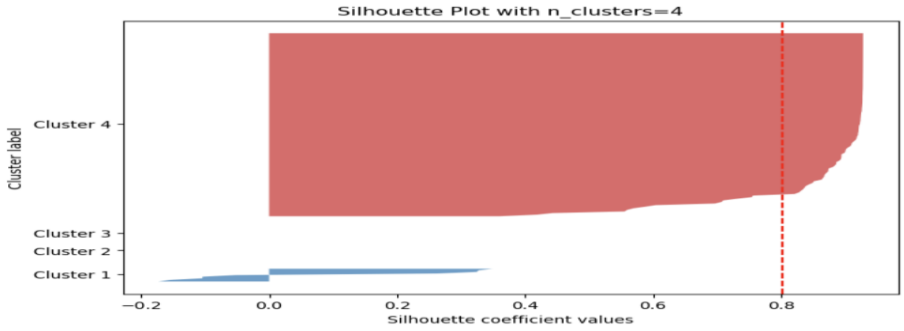


Figure 21: Silhouette Plot for Country Clusters

The clusters created using this method are listed below. Additionally, the average yearly sales amounts for each cluster are demonstrated in Figure 22. Cluster 1 and Cluster 2 consist of the countries that account for the majority of sales.

Cluster 1: Algeria, Libya, Italy, France, Belgium, Poland, Germany, Egypt, Spain
Cluster 2: Turkey
Cluster 3: England
Cluster 4: Kazakhstan, Madagascar, Albania, Sri Lanka, Uganda, Brunei, Malaysia, Oman, Mauritania, Qatar, Argentina, Uruguay, Tunisia, Togo, Mauritius, Kosovo, Ivory Coast, Equatorial Guinea, Benin, Bangladesh, Afghanistan, Sudan, Senegal, Liberia, Cuba, Philippines, Nepal, Guinea, Gambia, Panama, Uzbekistan, Kirghizstan, Taiwan, Holland, The Gambia, Cameroon, Pakistan, Peru, Paraguay,...

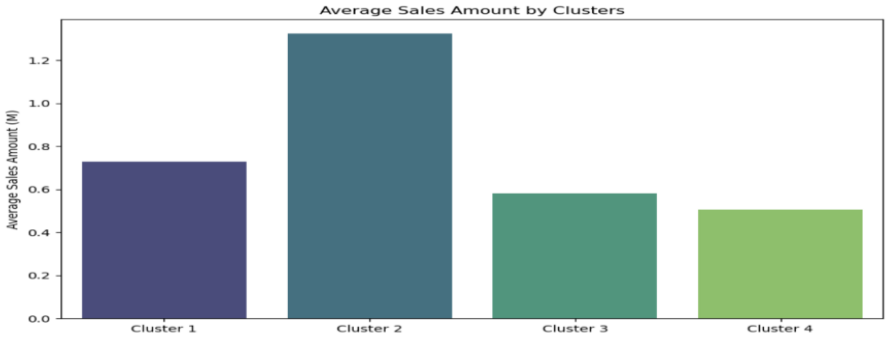


Figure 22: Average Yearly Sales by Country Clusters

PCA and t-SNE visualizations are used to analyze the separability of different clusters, shown in Figure 23 and Figure 24. Whereas the separation between clusters can be seen clearly in PCA, for example, cluster 2 and cluster 3 are plotted far apart, t-SNE better captures the separation within clusters, such as the distribution of points in Cluster 4.

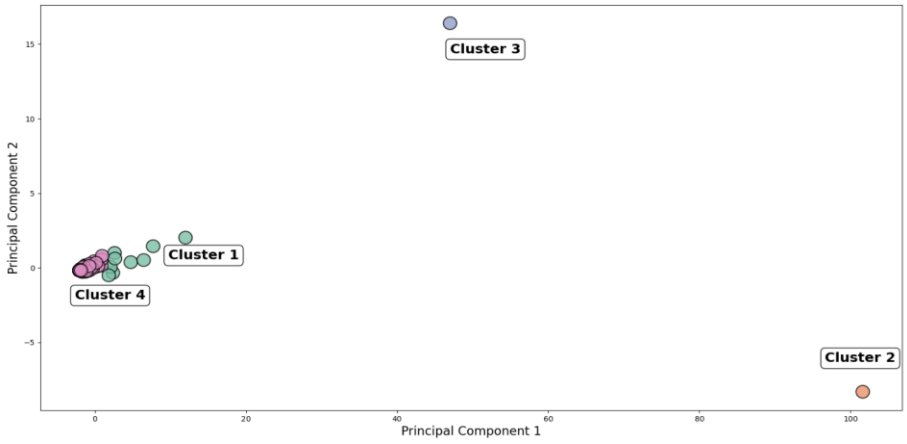


Figure 23: PCA Graph for Sales Region Clustering for First Approach

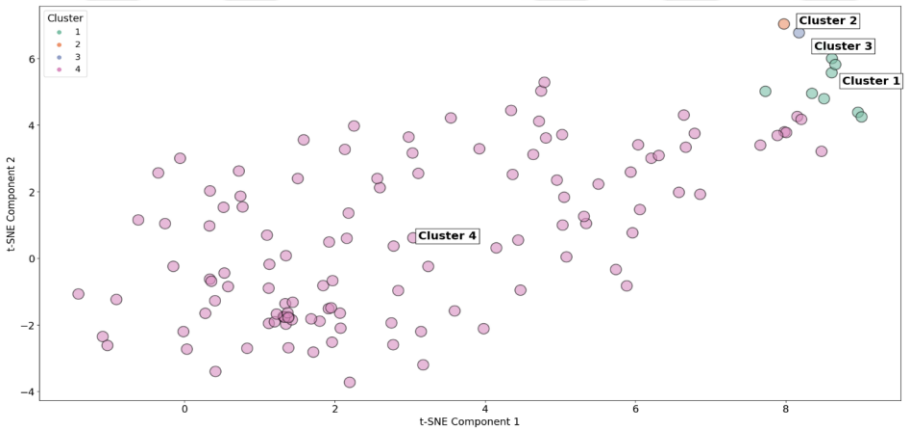


Figure 24: t-SNE Graph for Sales Region Clustering for First Approach

In the second approach, feature vectors representing monthly sales have been created for each country. Then, a cosine similarity matrix has been computed using the cosine similarity function from the Scikit-learn library. K-means method has been applied by using (1-cosine matrix) as distance matrix. Average Silhouette Plot has been employed, and cluster number has been chosen as four and the Silhouette score for this approach with cluster number 4 is 0.86, demonstrating strong clustering performance.

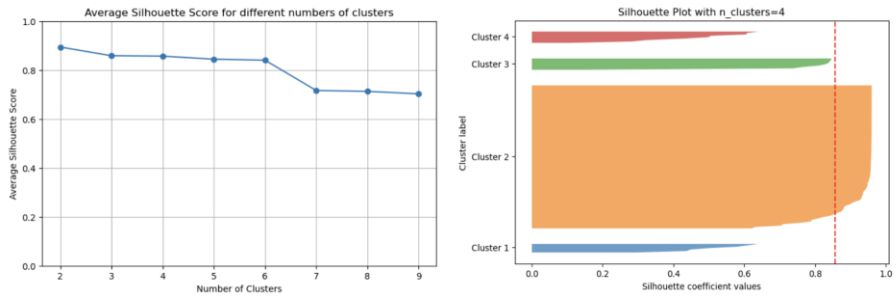


Figure 25: Average Silhouette Scores for different cluster numbers and Silhouette Plot with 4-clusters

The clusters created using this method are listed below and PCA and t-SNE visualizations can be shown in Figure 26 and Figure 27. While the separation between clusters and the distribution of the countries within Cluster 1, Cluster 3, and Cluster 4 can be seen better in the PCA graph, the t-SNE graph captures the distribution within Cluster 2 more effectively.

Cluster 1: Jordan, Norway, Serbia, United Arab Emirates, Iran, Tunisia, Australia
Cluster 2: Sri Lanka, Qatar, Brunei, Madagascar, Kazakhstan, Cuba, Argentina, Malaysia, Uruguay, Uganda, Mauritania, Albania, Kosovo, Ivory Coast, Equatorial Guinea, Benin, Bangladesh, Afghanistan, Sudan, Senegal, Liberia, Lebanon, India, Ghana, Yemen, Togo, Mauritius, …
Cluster 3: France, Sweden, Spain, Belgium, Germany, Turkey, England, Poland, Italy
Cluster 4: Algeria, Austria, Romania, Egypt, Iraq, USA, Libya, China, Israel

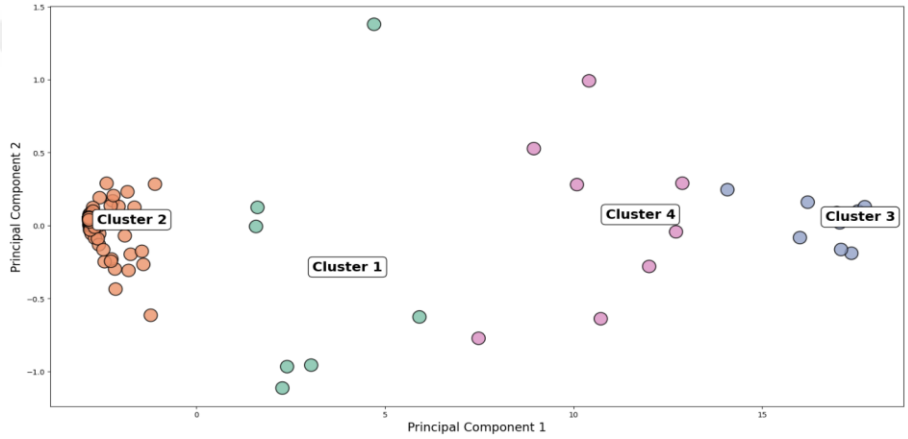


Figure 26: PCA Graph for Sales Region Clustering for Second Approach

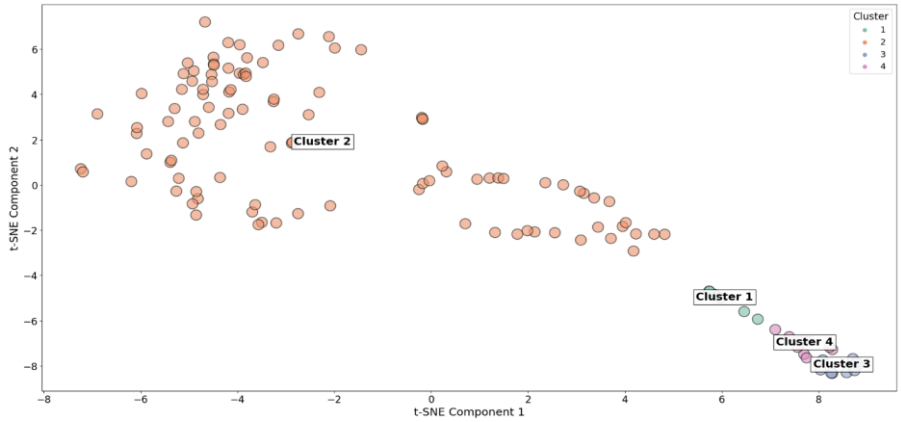


Figure 27: t-SNE Graph for Sales Region Clustering for Second Approach

Furthermore, the monthly sales trend of clusters with the country breakdown is represented in Figure 28. According to the graph, the countries showing similar trends are grouped in the same cluster.

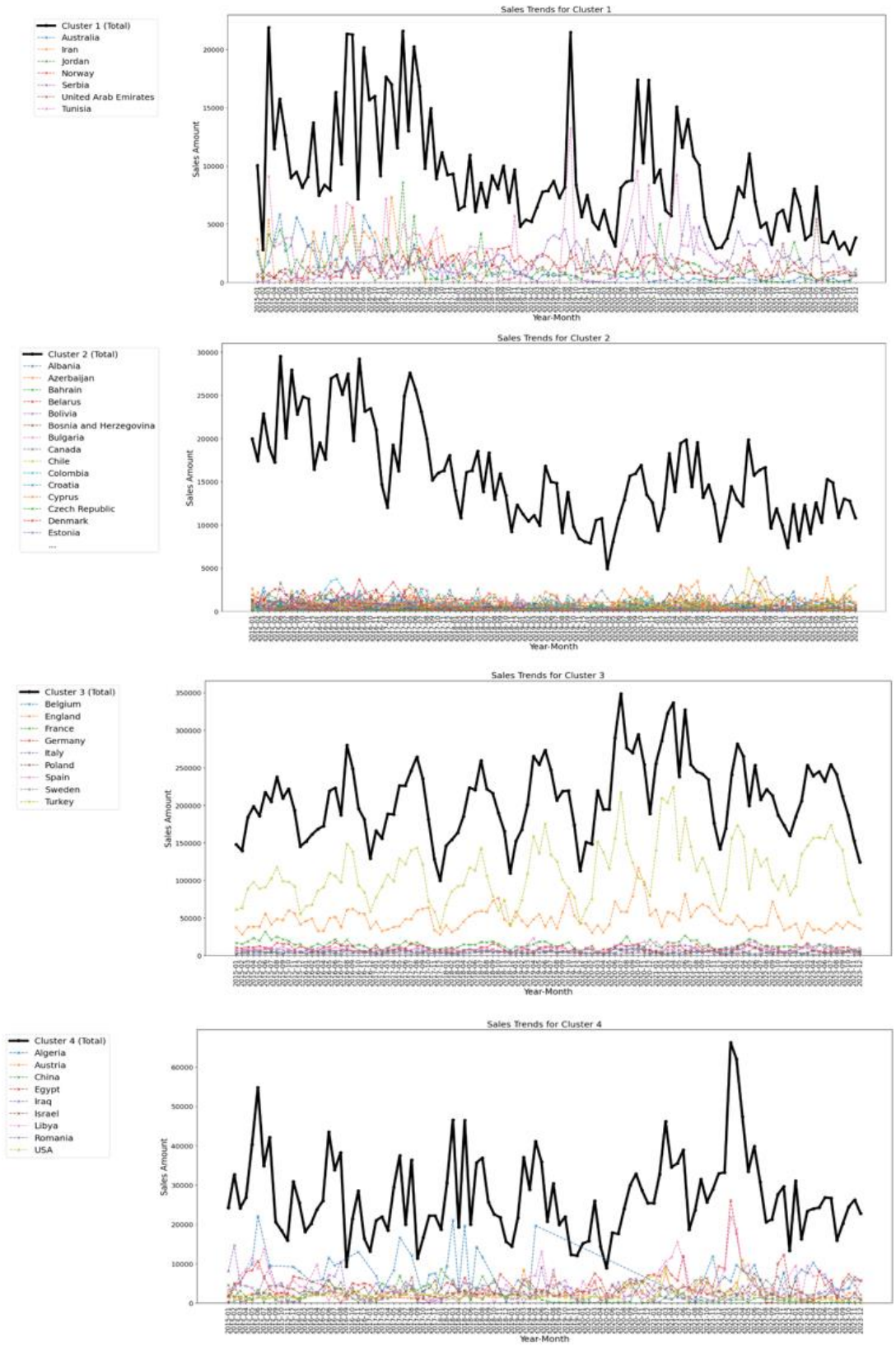


Figure 28: Sales Trend for Sales Region Clustering for Second Approach

The clustering of sales regions has also been evaluated through discussions with the four experts from the sales, marketing, supply chain, and product management departments. The sales expert has examined the clusters from the perspective of sales performance and trends, focusing on how these clusters could reflect changes in regional sales. The marketing expert has contributed by analyzing how the regional clusters align with market segmentation and customer behavior. The supply chain specialist has evaluated how this clustering improves forecasting accuracy and distribution processes. Additionally, the product management expert has analyzed the clusters in terms of their usefulness in planning sales region strategies. Based on their feedback, the second clustering approach has been selected, as it better understands sales patterns, reflects market dynamics, and enables to construct more effective regional strategies. Following this decision, the dataset has been reorganized: the "Sales Region" column has been replaced with the corresponding sales region cluster number, and the "Product Hierarchy 3" column has been replaced with the respective product cluster number.

3.5 Feature Engineering

In the above parts, only the sales dataset and its analysis are mentioned. Moreover, additional datasets have been generated to enhance the forecasting process within the thesis scope. Based on the sales trends, it is observed that sales peaked in 2021 before declining, which is hypothesized to be a result of the COVID-19 pandemic. To investigate this further, monthly COVID-19 death numbers (Ct) have been compiled.

- $\mathbf{C_t}$: the total number of reported COVID-19 deaths globally for a given month t, where t=2015-01,2015-02,.., 2023-12.

Then, based on domain knowledge, the stock level is correlated with the sales amount. Therefore, a monthly stock level ($\text{Stock}_{p,r,t}$) has been gathered, and the stock dataset has been created.

- **$\text{Stock}_{p,r,t}$**: Stock level for product p, sales region r in month t, where t=2015-01,2015-02,…2023-12.

Thirdly, to include the purchasing power effect of countries in the forecasting process, economic indicators ($\mathbf{E_{c,t}}$) have been collected based on the country level. To calculate the indicators for each sales region cluster ($\mathbf{E_{r,t}}$), the country-level indicators have been weighted by their respective sales proportions within the cluster, and the weighted values have been aggregated. Because the sales region clusters include multiple countries, each country's economic index should impact sales proportionally. For instance, a country with very high inflation but very low sales should not have a significant influence on the sales region clusters. The calculation of $\mathbf{E_{r,t}}$ is demonstrated as follows:

$$E_{r,t} = \frac{\sum_{c \in r}(Sales_{c,t} \times E_{c,t})}{\sum_{c \in r}(Sales_{c,t})} \quad (8)$$

where c denotes the country, and t denotes the month, $E_{c,t}$ denotes an economic indicator of the set of economic indicators based on country, and $Sales_{c,t}$ is the sales amount of country c in month t

The economic indicators include the inflation rate, GDP growth rate, unemployment rate, and interest rate, which are critical measures of a country's macroeconomic environment.

$$E_{r,t} \in \{Inflation_{r,t}, GDPGrowth_{r,t}, Unemployment_{r,t}, Interest\ rate_{r,t}\}$$

- **Inflation $_{r,t}$**: Inflation rate for sales region cluster r in month t, where t=2015-01,2015-02, …2023-12.
- **GDPGrowth $_{r,t}$**: GDPGrowth rate for sales region cluster r in month t, where t=2015-01,2015-02, …2023-12.
- **Unemployment $_{r,t}$**: Unemployment rate for sales region cluster r in month t, where t=2015-01,2015-02,…2023-12.
- **Interest rate $_{r,t}$**: Interest rate for sales region cluster r in month t, where t=2015-01,2015-02, …2023-12.

After the dataset generation, all features -monthly stock level, economic features, and covid-death numbers- have been combined with the sales dataset. The descriptive statistic of the combined dataset is presented below.

Table 7: Combined Dataset Descriptive Statistics

	Product Cluster	Sales Region Cluster	Covid-19	Inflation	Interestrate	GDPGrowth	Unemployment	Sales	Stock
count	1725	1725	1725	1725	1725	1725	1725	1725	1725
mean	2.5	2.5	65085.7	5.5	3.5	1	7.1	16378.9	7889.5
std	1.1	1.1	111636.7	8.9	3.4	2.4	4.4	30951.4	14439.1
min	1	1	0.0	-0.8	0.0	-13	0	5	0
25%	1	1	0.0	0	0.0	0	0	1374	752
50%	2	2	0.0	3.3	3.2	0.8	8.6	4055	2294
75%	3	3	85772	6.7	4.6	1.4	9.9	14573	8328
max	4	4	484000	55.9	21.6	15.4	17.1	196319	134231

After the features have been added to the sales dataset, to approach the problem from two different perspectives, it has been decided to generate two distinct datasets from the compact dataset. The first dataset, referred to as the "Product Sales Dataset" is product-cluster based, aggregating sales and stock levels for each product without considering the sales region breakdown. The Covid death number changes year-monthly independently of the product and sales region, so it hasn't been aggregated, and included directly. The economic indicators are based on year-month and sales region clusters. Furthermore, to incorporate these features into the product sales dataset, the weighted average of the indicators has been calculated for each month by assigning a weight for the sales region, taking account of the sales amount concerning the month, and it has been multiplied by the economic features values concerning a month. The same logic is also valid in this calculation; the economic index of a cluster

with a higher sales volume in a given month should have a greater impact. The mathematical formulation and an example of the calculation for a single economic indicator are presented below:

$$\mathrm{WAEI}_t = \frac{\sum_r (\mathrm{Sales}_{r,t} \times \mathrm{E}_{r,t})}{\sum_r \mathrm{Sales}_{r,t}} \quad (9)$$

where r denotes the sales region cluster, and t denotes the month, $\mathrm{E}_{r,t}$ denotes an economic indicator from the set of economic indicators based on the sales region cluster, and $\mathrm{Sales}_{r,t}$ is the sales amount of the sales region cluster r in month t

Table 8: The inflation rate calculation for the product dataset

Weights for sales regions

Sales Region	1	2	3	4	Total_Sales
Year-Month					
2015-01	0.050	0.099	0.732	0.120	1.000
2015-02	0.015	0.090	0.725	0.170	1.000
2015-03	0.086	0.090	0.728	0.095	1.000
2015-04	0.045	0.074	0.776	0.105	1.000
2015-05	0.061	0.067	0.717	0.155	1.000

Monthly inflation rates for sales regions

Sales Region	1	2	3	4
Year-Month				
2015-01	0.6	0.0	2.9	1.8
2015-02	1.6	0.0	3.3	2.3
2015-03	3.5	0.0	3.6	5.2
2015-04	2.4	0.0	3.9	6.6
2015-05	2.1	0.0	4.0	6.6

Monthly weighted average inflation rates

	Year-Month	Weighted_Average_Inflation
0	2015-01	2.367
1	2015-02	2.807
2	2015-03	3.418
3	2015-04	3.826
4	2015-05	4.022

Then, the number of product clusters covering at least %80 of the sales was investigated. The cumulative sales contribution is calculated as follows:

$$\text{Cumulative Contribution}_p = \frac{\sum_{i=1}^{p} Sales_i}{\sum_{i=1}^{P} Sales_i} \quad (10)$$

where Sales_i are sorted in descending order, P denotes the total number of product clusters, and p denotes the minimum product cluster number such that the cumulative contribution satisfies the threshold condition (at least 80% of total sales)

It has been found that the first three product clusters' sales cover the 90% of the total sales. Therefore, the fourth product cluster's data have been removed from the dataset. The second dataset, referred to as the "Product-Region Sales Dataset" includes a breakdown by product and sales region (a new 'Product-Region' column has been created. This dataset aggregates sales and stock levels for each product-region combination. COVID-19 death number changes year-monthly regardless of product and sales region; it has been included directly. Economic indicators are based on year-month and sales region clusters, and they haven't been affected by product type, so they have also been included directly. The same approach applied to this dataset; five out of sixteen product-region groups account for 86.5% of the total sales. Therefore, other product-region groups have been removed from the dataset.

Following the finalization of datasets, lag variables have been created for the sales, stock, and economic indicators. For the sales amount feature, lag variables have been created for the 10th, 11th, and 12th of previous months. Vector Autoregressive Model (VAR) has been applied, and the Akaike Information Criterion (AIC) has been chosen to determine the optimal lag number for the stock and economic indicators (Bauerová, Starzyczná, & Pražákák, 2022; Bukkapatnam, Kim, & Iyer, 2012).

$$\mathrm{AIC}(k) = -2 \times \ln(\hat{L}) + 2k \tag{11}$$

$\hat{L}$ denotes the likelihood of the model, and k represents the number of lags in the VAR model

This analysis has been conducted for each product and product-region group and according to this analysis, the values belonging to Lag_12 yield the lowest AIC for each cluster for both datasets. The results for the "Product Sales Dataset" can be shown in Table 9. Therefore, lagged variables have been determined by for the 12th previous month (Lag_12).

Table 9: VAR Analysis for Product Clusters

Product Cluster: 1
Lag Order Selection for Product Cluster 1:
VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	69.96	70.15	2.419e+30	70.04
1	61.72	63.22*	6.390e+26	62.32
2	62.19	65.00	1.043e+27	63.33
3	62.12	66.24	1.013e+27	63.79
4	61.78	67.20	7.775e+26	63.97
5	61.96	68.69	1.061e+27	64.68
6	61.44	69.48	7.755e+26	64.69
7	61.12	70.47	7.726e+26	64.90
8	60.83	71.49	9.167e+26	65.14
9	59.98	71.95	7.732e+26	64.82
10	58.85	72.12	6.810e+26	64.21
11	56.05	70.64	1.997e+26	61.95
12	48.38*	64.27	1.380e+24*	54.80*

Product Cluster: 2
Lag Order Selection for Product Cluster 2:
VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	67.59	67.78	2.261e+29	67.67
1	59.44	60.94*	6.566e+25	60.05
2	59.68	62.48	8.404e+25	60.81
3	59.53	63.64	7.556e+25	61.19
4	59.30	64.72	6.500e+25	61.49
5	59.24	65.98	7.024e+25	61.97
6	58.76	66.80	5.328e+25	62.01
7	58.25	67.60	4.376e+25	62.03
8	57.51	68.16	3.310e+25	61.81
9	56.53	68.49	2.446e+25	61.36
10	54.06	67.33	5.670e+24	59.42
11	51.72	66.31	2.623e+24	57.62
12	46.85*	62.74	2.999e+23*	53.27*

Product Cluster: 3
Lag Order Selection for Product Cluster 3:
VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	73.22	73.40	6.279e+31	73.29
1	65.04	66.54*	1.770e+28	65.65
2	65.36	68.17	2.484e+28	66.50
3	65.29	69.41	2.404e+28	66.95
4	65.07	70.49	2.081e+28	67.26
5	64.94	71.67	2.090e+28	67.66
6	64.38	72.42	1.471e+28	67.63
7	63.90	73.25	1.244e+28	67.68
8	63.14	73.80	9.299e+27	67.45
9	61.70	73.67	4.332e+27	66.54
10	60.39	73.67	3.195e+27	65.76
11	57.77	72.35	1.110e+27	63.66
12	51.14*	67.03	2.179e+25*	57.56*

These lagged features are intended to be used instead of the original stock and economic indicators in the analysis to avoid data leakage when making predictions on the test data. For the first year, null values happened. For lagged sales features, these values were filled with (t+12-lag). variables to maintain the same data size. For example, the Sales_Lag_10 feature for product cluster 1 in 2015-01 is assigned to the sales value from 2015-03. For the first year, the original stock and economic indicators have been used in place of the lagged features. The lagging method is summarized below.

$$\mathrm{Sales_Lag}_{\mathrm{lag}}(t) = \begin{cases} \mathrm{Sales}(t-\mathrm{lag}) & \text{if Sales}(t-\mathrm{lag}) \text{ is not null} \\ \mathrm{Sales}(t+(12-\mathrm{lag})) & \text{otherwise} \end{cases}$$

$$\mathrm{Stock_Lag}_{\mathrm{lag}}(t) = \begin{cases} \mathrm{Stock}(t-\mathrm{lag}) & \text{if Stock}(t-\mathrm{lag}) \text{ is not null} \\ \mathrm{Stock}(t) & \text{otherwise.} \end{cases} \tag{12}$$

After the features were created, a correlation analysis has been conducted. For both datasets, the correlation matrix reveals that sales are highly correlated with lagged sales values and lagged stock values. In contrast, economic indicators and COVID-19

show weak or negligible correlations with sales. However, these factors will still be included in the models to investigate their potential impact on sales prediction further.

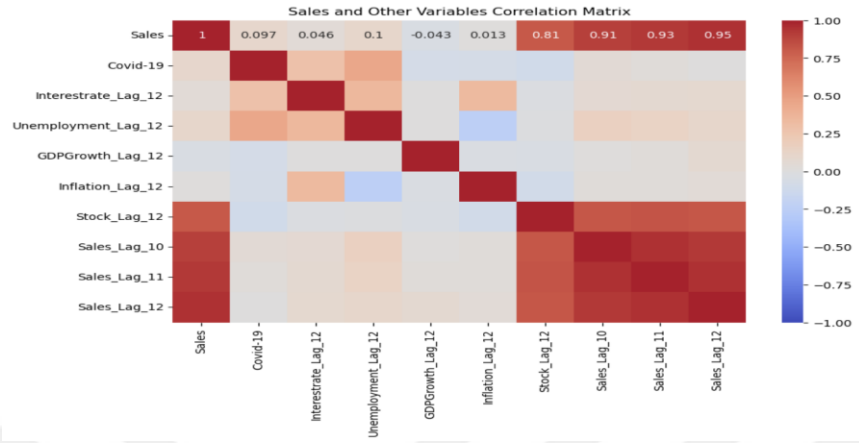


Figure 29: Correlation Analysis between Sales Amount and Exogenous Variables for Product Sales Data

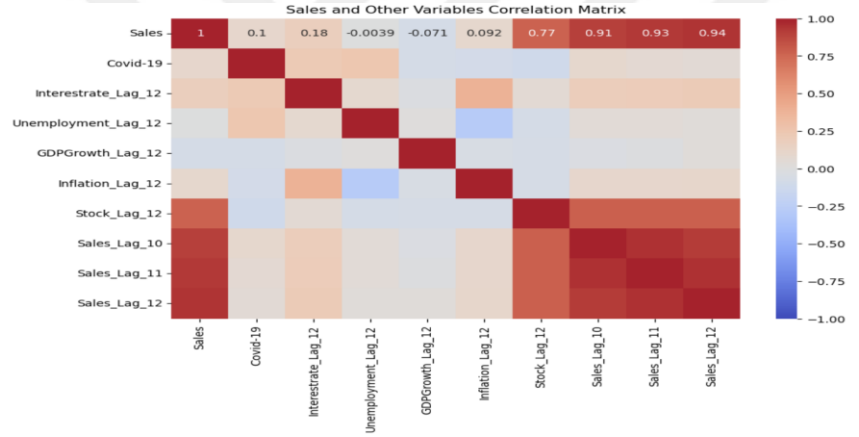


Figure 30: Correlation Analysis between Sales Amount and Exogenous Variables for Product-Region Sales Data

After the feature creation, outlier detection and handling methods have been reviewed. According to the literature, statistical methods or machine learning methods can be used to detect outliers. There are three statistical methods, such as the Interquartile range method, the standard deviation method, and the modified Z Score method. Machine learning methods also contain different methods such as Isolation Forest, K-Nearest Neighbor, and Random Forest (Cihan Çılgın, 2023; Ribeiro, Seruca, & Durão, 2017). In this study, the IQR method was selected and modified using the sliding window approach, with the window parameter set to 12, and it is formulated as follows:

$$\mathrm{IQR_w} = \mathrm{Q3_w} - \mathrm{Q1_w} \tag{13}$$

where $\mathrm{Q1_W}$ denotes the 25th percentile of the data within the window and $\mathrm{Q3_W}$ denotes the 75th percentile of the data within the window

An observation x_i is an outlier if it satisfies the following condition:

$$x_i < Q1_w - 1.5 \times IQR_W \text{ or } x_i > Q3_w + 1.5 \times IQR_W$$

Following the outlier detection, outlier handling methods have been investigated. According to the literature, the handling methods include removing the outliers; transformation methods -transforming the data to decrease the impact of the outliers-, imputation methods, namely changing the outlier values with mean, median, or different values-; the winsorization method which sets the outliers to upper or lower thresholds (Aguinis, Gottfredson, & Joo, 2013; Nyitrai & Virág, 2019). The winsorization method has been selected, and detected outliers have been set to lower or upper values of the respected window.

$$x_i = \begin{cases} Q1_w - k \times IQR_W & \text{if } x_i < Q1_w - k \times IQR_W \\ Q3_w + k \times IQR_W & \text{if } x_i > Q3_w + k \times IQR_W \end{cases} \quad (14)$$

3.6 Modeling

In this section, the construction of the models mentioned in section 2.2 and the hyperparameter optimization process are explained.

3.6.1 Model Construction

The model construction part started with experiments based on datasets and methods described in various research papers, as mentioned in Section 3.1. ARIMA, LSTM, and Prophet models have been chosen as the forecasting models, and these models have been constructed to work with five different datasets obtained from different resources. Important parameters for the models have been identified through a mix of literature reviews and tests on these datasets. Observations have been made about how different parameters affected the models' performance.

After building the initial models, they have been applied to the main datasets of the thesis. During this step, the parameters have been adjusted to fit the dataset better. For example, since the data shows repeating yearly seasonality, the yearly seasonality parameter has been used in the Prophet model. Similarly, the ARIMA model has been replaced with SARIMA to capture the seasonal patterns in the data. These adjustments ensured the models were compatible with the dataset's characteristics. Additionally, whenever extra data preprocessing steps are needed for the models, the data preprocessing step has been revisited. For example, stationarity was checked for the SARIMA model, and the data has been made stationary. Then the model has been performed on the dataset again.

After fine-tuning, additional improvements have been made by including external variables, as mentioned earlier. The selection of external variables was guided by business insights. Economic factors and COVID-19-related variables have been considered critical factors and added to the dataset as features. Moreover, the dataset has been divided into training-validation and test sets. The sales data from 2015 to

2021 have been used as the training set, the sales data from 2021 to 2022 have been used as the validation set, and the sales data from 2022 to 2023 have been used as the test set. After these steps, the models have been applied to the datasets and the models' structures are summarized below.

Firstly, the base model is explained; in this model, forecasts are generated by simply using the previous actual sales value as the prediction for the next period. This approach states that future sales will closely follow the most recent observed value, serving as a simple benchmark for comparison with other models.

One of the main models in this thesis is SARIMA; as explained above, before applying this model, the datasets were assessed for stationarity using the Augmented Dickey-Fuller (ADF) test. A moving average smoothing technique was applied to reduce short-term fluctuations. Although not a standard method for achieving stationarity, the ADF test was reapplied after smoothing and confirmed that the series had become stationary eliminating the need for further transformations. Then two SARIMA sub-models were developed. The first adopts a univariate forecasting approach, where for each forecasting level (e.g., product category, product category cluster, product category-sales region, and product category cluster-sales region cluster), a one-dimensional vector of 84 monthly sales values is used as input. The second employs a multivariate forecasting approach. At each forecasting level, both a one-dimensional vector of 84 monthly sales values and a matrix with n_features rows and 84 columns are used as input. In this matrix, each row corresponds to a different exogenous feature, and each column represents a monthly time step. The SARIMA model can be formulated like below, and it includes seven parameters.

$$y = \text{SARIMA}(\text{p}, \text{d}, \text{q}) \times (\text{P}, \text{D}, \text{Q}, \text{s}) \tag{15}$$

The SARIMA model formula consists of two main parts: non-seasonal parameters and seasonal parameters. The non-seasonal parameters include the general trends and short-term dynamics in the data. These include p (Auto-Regressive, AR), which represents the number of past observations used to predict the current value; d (Differencing), the number of times the data is differenced to make it stationary; and q (Moving Average, MA), which accounts for the number of past forecast errors influencing the current value.

On the other hand, the seasonal parameters focus on recurring patterns and periodicities. These include P (Seasonal Auto-Regressive), the number of past seasonal observations affecting the current seasonal value; D (Seasonal Differencing), the number of seasonal differences applied to remove seasonality; and Q (Seasonal Moving Average, SMA), the number of past seasonal forecast errors incorporated into the model; and lastly s defines the seasonal period. Because the data was made stationary before using the model, differencing parameters have been set to zero, and the data period is the month, and yearly seasonality existed, so s has been set to 12. Therefore, four parameters (p, q, P, Q) have been included in the hyperparameter optimization part.

The other model used is the stacked LSTM model; the architecture consists of two LSTM layers stacked on top of each other, allowing the model to learn complex patterns in the data. The first LSTM layer outputs sequences, which are processed by the second LSTM layer, and a dropout layer is incorporated to prevent overfitting. Finally, a dense layer with one neuron is used to generate the final prediction. The model is compiled using the Huber loss function for robustness to outliers and the Adam optimizer.

Two variations of the LSTM model were constructed. The first uses a univariate forecasting approach, for each forecasting level a one-dimensional vector of 84 monthly sales values is used as input. Input sequences are then formed based on the n_steps parameter. The second model adopts a multivariate approach, where the input is a matrix with n_features rows and 84 columns, each row representing a different feature (sales amount and exogenous features) and each column corresponding to a monthly time step. As in the univariate model, input sequences are generated using the n_steps parameter.
Hyperparameters such as number of time steps, epoch number, neuron number, and learning rate have been optimized to improve the performance. The number of time steps determines how many past data points are used as input to predict future values. The epoch number specifies how many times the model iterates over the entire training dataset, balancing underfitting and overfitting. The neuron number refers to the number of units in the hidden layers. Lastly, the learning rate controls the step size during optimization, influencing both the speed and stability of the training process.

Another model that is mentioned is the Prophet model. Also, for this part, two sub-models have been created. One adopts univariate forecasting approach and for each clustering level, a two-column matrix consisting of 84 rows was used as input, where the first column (ds) represented the monthly time steps and the second column (y) denoted the corresponding sales values. Other model adopts multivariate forecasting approach and for each clustering level, a matrix consisting of 84 rows and n_features columns was used as input, where the first column (ds) represented the monthly time steps, the second column (y) denoted the corresponding sales values, and the remaining columns denotes used exogenous features.

The season has been chosen as a month, and the yearly seasonality mode has been activated. The optimization process focused on four key parameters: changepoint prior scale, seasonality prior scale, the number of changepoints, and the Fourier order. The changepoint prior scale controls the sensitivity of the model to trend changes, while the seasonality prior scale adjusts how closely the model fits seasonal patterns. The number of changepoints determines the resolution at which trend changes are captured, and the Fourier order defines the complexity of seasonal components. By fine-tuning these parameters, the models can effectively capture both long-term trends and recurring seasonal patterns.

3.6.2 Hyperparameter Optimization

The main models consist of different hyperparameters that can affect the models' performances. These hyperparameters have been optimized using the Optuna library. The Tree-structured Parzen Estimator (TPE) has been used as a sampler to help find the best hyperparameters by focusing on the most promising areas in the search space. Instead of randomly trying different values, it uses past results to predict which combinations are more likely to perform well, making the optimization faster and more efficient. (Watanabe, 2023)

The hyperparameters and their ranges used for each model are shown in Table 10. The hyperparameter ranges and step sizes have been determined based on the structure of the data, model trials, and common practices in existing studies. For example, during model experiments, it has been observed that small changes in the changepoint_prior_scale and seasonality_prior_scale in the Prophet model significantly affected the result, so the step size has been kept small.

Table 10: The hyperparameters range of the models

Model	Hyperparameters	Range/step	Step
SARIMA	p	[1,5]	1
	q	[1,5]	1
	P	[0,2]	1
	Q	[0,2]	1
LSTM	number of steps	[1,6]	1
	number of epochs	[100,150]	10
	number of neurons	[50,100]	10
	learning rate	[0.0001, 0.0002,0.0005,0.0008,0.001]	(list-based)
PROPHET	changepoint_prior_scale	[0.01,10]	0.05
	seasonality_prior_scale	[0.01,10]	0.05
	number of change points	[5,30]	5
	fourier order	[1,5]	2

Because the model complexities and the number of hyperparameter combinations vary, different trial numbers have been used and the time for the hyperparameter optimization has been recorded. According to Table 11, LSTM has the slowest runtime for a single trial, moreover because of a higher number of combinations of hyperparameters for the Prophet model, the trial number for this model has been chosen greater than others.

Table 11: Optimization Time of Forecasting Models

Model	Optimization Duration (s)	Trial Number
SARIMA-1 (Product Sales Data)	697	200
SARIMA-2 (Product-Region Sales Data)	901	
SARIMA-3 (Product Sales Data with exogenous variables)	886	
SARIMA-4 (Product-Region Sales Data with exogenous variables)	824	
LSTM-1 (Product Sales Data)	964	100
LSTM-2 (Product-Region Sales Data)	1633	
LSTM-3 (Product Sales Data with exogenous variables)	772	
LSTM-4 (Product-Region Sales Data with exogenous variables)	1467	
PROPHET-1 (Product Sales Data)	1463	2000
PROPHET-2 (Product-Region Sales Data)	2015	
PROPHET-3 (Product Sales Data with exogenous variables)	1620	
PROPHET-4 (Product-Region Sales Data with exogenous variables)	3370	

Furthermore, these experiments have been monitored on W&B, and the hyperparameter optimization process has been visualized. Parallel coordinates graphs of some examples from the experiments conducted for each model are presented to understand how the change of parameters and combinations of these affect the evaluation results. For some parameters, specific observations can be made; for instance, as shown in Figure 31, an excessively small learning rate leads to worse SMAPE results. However, W&B visualizations indicate that it is not possible to attribute good performance to a single parameter value. Instead, improvements in performance are achieved through variations in parameter combinations.

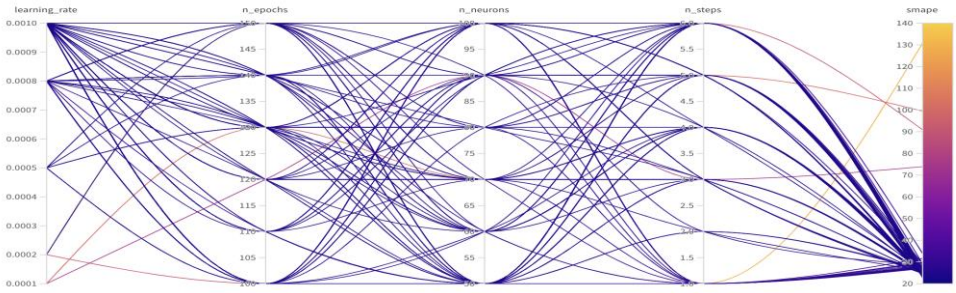


Figure 31: Parallel Coordinates Graph for LSTM

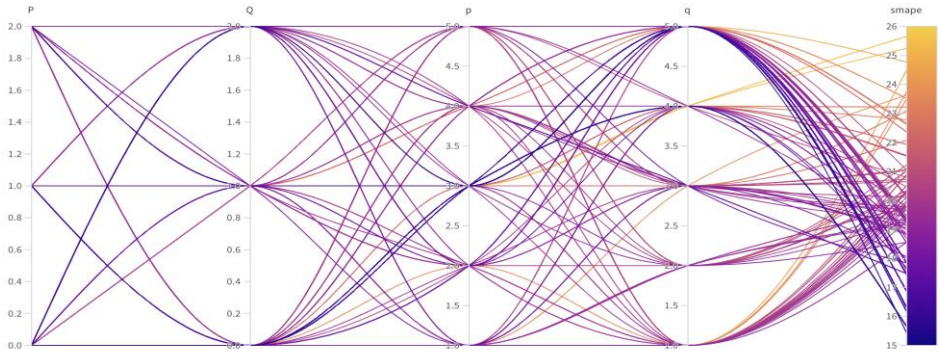


Figure 32: Parallel Coordinates Graph for SARIMA

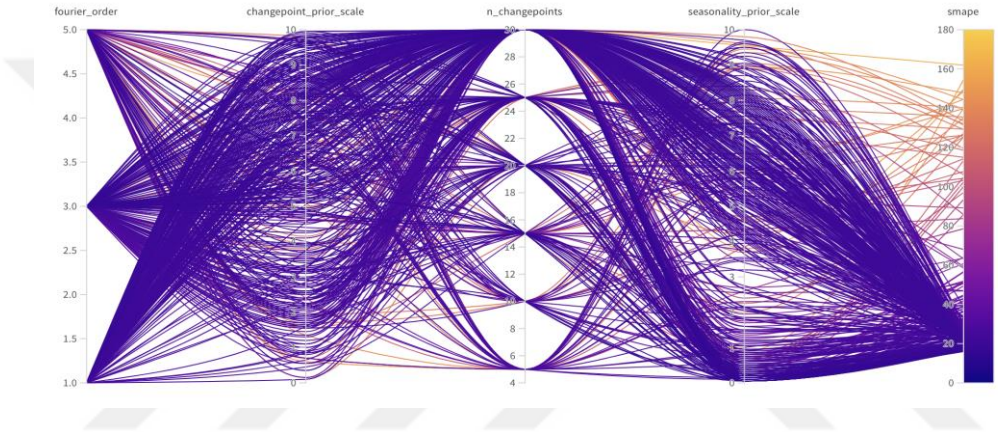


Figure 33: Parallel Coordinates Graph for Prophet

3.7 Evaluation

After training the models on the datasets and performing hyperparameter optimization, their performance needs to be evaluated. Based on the results, certain steps may need to be revisited. For example, an evaluation of SARIMA results indicated the need to revise data preprocessing, leading to modifications in the stationarity method. Similarly, if necessary, the modeling process is reapplied, and the hyperparameter tuning duration is adjusted accordingly.

As detailed in Section 2.4, three distinct evaluation metrics have been utilized to compare the performance of various forecasting models: RMSE, MAE and SMAPE. The developed models have been applied individually to each product group and product-region group, and the corresponding evaluation metrics have been calculated for each group.

Following the model execution, the weighted average of these metrics has been computed to provide an overall assessment. The same calculations have been performed for the training, validation and test sets. An example of the calculation for the test set for each evaluation metric is defined below:

$$\mathrm{WA_RMSE} = \frac{\sum_{p=1}^{P} \mathrm{RMSE}_p * w_p}{\sum_{p=1}^{P} w_p} \tag{16}$$

where $w_p = \sum_{k=1}^{K} \mathrm{Sales}_{p,k}$, $\mathrm{RMSE}_p = \sqrt{\frac{1}{K} \times \sum_{k=1}^{K} (\hat{y}_k - y_k)^2}$,and p denotes the product group, k denotes the k-th observation in test size for product group p, and K denotes the sample size of the test set for product group p, $\mathrm{Sales}_{p,k}$ represents the sales amount for the k-th observation in the test set of the p-th product group.

$$\mathrm{WA_MAE} = \frac{\sum_{p=1}^{P} \mathrm{MAE}_p * w_p}{\sum_{p=1}^{P} w_p} \tag{17}$$

where $w_p = \sum_{k=1}^{K} \mathrm{Sales}_{p,k}$, $\mathrm{MAE}_p = \sqrt{\frac{1}{K} \times \sum_{k=1}^{K} |\hat{y}_k - y_k|}$, and i denotes the product group, k denotes the k-th observation in test size for product group p, and K denotes the sample size of the test set for product group p, $\mathrm{Sales}_{p,k}$ represents the sales amount for the k-th observation in the test set of the p-th product group.

$$\mathrm{WA_SMAPE} = \frac{\sum_{p=1}^{P} \mathrm{SMAPE}_p * w_p}{\sum_{p=1}^{P} w_p} \tag{18}$$

where $w_p = \sum_{k=1}^{K} \mathrm{Sales}_{p,k}$, $\mathrm{SMAPE}_p = \frac{100}{K} \times \sum_{k=1}^{K} \frac{|\hat{y}_k - y_k|}{(|\hat{y}_k| + |y_k|)/2}$, and p denotes the product group, k denotes the k-th observation in test size for product group p and K denotes the sample size of the test set for product group p, $\mathrm{Sales}_{p,k}$ represents the sales amount for the k-th observation in the test set of the p-th product group

CHAPTER 4

RESULTS & DISCUSSION

This section represents the results and their evaluation, along with a comparison across different modeling scenarios. Furthermore, the findings were discussed and compared with relevant literature to highlight both consistencies and differences from existing studies.

4.1 Results and Comparison

As previously mentioned, there are three main models, and these models were first trained to generate forecasts for the company's own product and sales region categorization, which is the first and second scenario in Table 12. Subsequently, products and countries were clustered, and forecasts were generated for the product cluster and product cluster-sales region cluster breakdowns, which is the third and fourth scenario in Table 12. During this forecasting process, different experiments were conducted using various exogenous variables. Table 12 represents all approaches, and the features used.

Table 12: Forecasting Approaches and Utilized Features

Scenario	Forecasting Level	Forecasting Approach	Utilized Features
1	Product Category	Univariate Time Series Forecasting	Monthly Sales for each product category
2	Product Category-Sales region	Univariate Time Series Forecasting	Monthly Sales for each product category-sales region
3	Product Category Cluster	Univariate Time Series Forecasting	Monthly Sales for each product category cluster
		Multivariate Time Series Forecasting (Experiment 1)	Monthly Sales for each product category cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12

Table 12 (cont.)

3	Product Category Cluster	Multivariate Time Series Forecasting (Experiment 2)	Monthly Sales for each product category cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12 GDPGrowth_Lag12, Unemployment_Lag12, Interestrate_Lag12
		Multivariate Time Series Forecasting (Experiment 3)	Monthly Sales for each product category cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12, Unemployment_Lag12
4	Product Category Cluster-Sales Region Cluster	Univariate Time Series Forecasting	Monthly Sales for each product category cluster-sales region cluster
		Multivariate Time Series Forecasting (Experiment 1)	Monthly Sales for each product category cluster-sales region cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12
		Multivariate Time Series Forecasting (Experiment 2)	Monthly Sales for each product category cluster-sales region cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12 GDPGrowth_Lag12, Unemployment_Lag12, Interestrate_Lag12
		Multivariate Time Series Forecasting (Experiment 3)	Monthly Sales for each product category cluster-sales region cluster, Covid-19, Stock_Lag12, Sales_Lag10, Sales_Lag11, Sales_Lag12 Unemployment_Lag12

The models were first applied to product and product-region groups used by the company. There are 12 different product groups, such as "53-59 cm-Cooling", "60-63 cm-Freezer", etc., and the cumulative sales contribution approach defined in Equation 10 in Section 3.5 has been applied to these product groups. Seven of 12 different product groups account for the majority of the sales; therefore these seven product groups have been included in the modeling parts. The models' results are shown in Table 13. The best result is highlighted in bold and marked with * for clarity; PROPHET model outperformed the other models, achieving a SMAPE of 24.52%. The evaluation results based on each product group are shown in Appendix A.1.

Table 13: The evaluation results on the product groups defined by the company

Model	MAE	RMSE	SMAPE
BASE Model	10514	12764	26.58
SARIMA	9129	11024	25.52
LSTM	10690	13177	26.87
***PROPHET**	**8958**	**10558**	**24.52**

Then, the models have been applied for product-region groups used by the company, such as "53-59 cm-Freezer-Europe", "70-73 cm-Cooling-Turkey & South Asia", etc. There are 58 different product-region groups, and according to cumulative sales contribution calculation, 12 of them consist of the majority of the sales. Therefore, these 12 product-region groups have been included in the modeling parts. The evaluation results can be seen in Table 14. The best result is highlighted in bold and marked with * for clarity; SARIMA model outperformed the other models, achieving a SMAPE of 22.27 %. The evaluation results based on each product group are shown in Appendix A.2.

Table 14: The evaluation results on the product-region groups defined by the company

Model	MAE	RMSE	SMAPE
BASE Model	8081	9405	34.95
***SARIMA**	**4839**	**6020**	**22.27**
LSTM	7937	10181	33.02
PROPHET	6294	7953	30.57

After these experiments, these models have been trained, and hyperparameter optimization has been conducted on two datasets: the Product Sales Dataset and the Product-Region Sales Dataset, formed based on the clustering methods mentioned in section 3.4. For both datasets, the models have been applied separately to each product and product-region cluster.

The models have been applied by using only historical sales data, then they have been performed by incorporating exogenous variables. As mentioned in section 3.5, according to correlation matrices (Figure 29 and Figure 30) the effects of COVID-19 and economic indicators on sales are negligible, but the models have been executed with different combinations of features to investigate their potential impacts on sales prediction.

Firstly, the models' results on the Product-Sales Dataset are explained. Table 15 shows the results of the models' execution by using only historical sales data, and the best result is highlighted in bold and marked with * for clarity. Prophet model outperformed the other models, achieving a SMAPE of 11.69%, and yielded better MAE and RMSE results than other models. Furthermore, it is observed that the clustering method improves the results. The experiment using the clustering method achieved 52% better results compared to the experiment based on the company's usage scenario demonstrated in Table 13. The evaluation results based on each product group and

comparison with the company's own product groups are shown in Appendix A.3. Moreover, graphical comparison of forecasted and actual values can be found in Figure 34. It demonstrates that Prophet model effectively captures both seasonal patterns and trend changes and produces predictions that are largely consistent with the actual sales values.

Table 15: The model results on the Product Sales Test Dataset by using only historical sales data

Model	MAE	RMSE	SMAPE (%)
BASE Model	13574	17075	13.13
SARIMA	10707	14536	12.10
LSTM	17498	21216	15.75
***PROPHET**	**9988**	**13469**	**11.69**

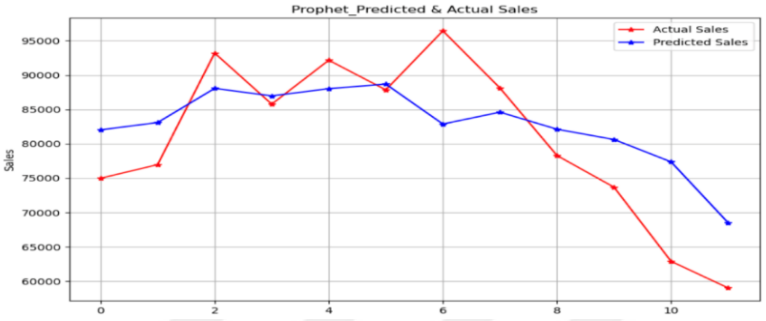


Figure 34: Average Predicted vs. Actual Sales Across Product Category Clusters

The models have been performed by using different combinations of exogenous variables. The three best-performing experiments are presented in Table 16, and the best result is highlighted in bold and marked with * for clarity. Among the models, when evaluated based on SMAPE, the LSTM model produces better results, whereas Prophet appears to perform better according to the MAE and RMSE metrics. However, compared to the results in Table 15, incorporating exogenous variables did not improve upon the best outcome achieved by the Prophet model using only historical sales data. Therefore, for the product breakdown, only historical sales data should be utilized to achieve optimal performance. The evaluation results for the Product Sales Training Dataset can be seen in Appendix B.1, and the results of the second experiment for each product group can be seen in Appendix A.4.

Table 16: The model results on the Product Sales Test Dataset by using different exogenous variables

Exp.	Exogenous Variables	Model	MAE	RMSE	SMAPE (%)
1	Covid-19 Stock_Lag$_{12}$,Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$	SARIMA	14569	18501	19.61
		LSTM	12746	14960	13.34
		PROPHET	10769	14757	14.68

Table 16 (cont.)

2	Covid-19 Stock_Lag$_{12}$,Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ GDPGrowth_Lag$_{12}$, Unemployment_Lag$_{12}$, Interestrate_Lag$_{12}$	SARIMA	20053	24048	22.71
		***LSTM**	**11862**	**14319**	**12.62**
		PROPHET	10623	14187	15.74
3	Covid-19 Stock_Lag$_{12}$,Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ Unemployment_Lag$_{12}$	SARIMA	15972	20467	20.69
		LSTM	12767	15114	13.10
		PROPHET	10085	13975	14.14

Secondly, the models have been performed on the Product-Region Sales Dataset. Table 17 shows the results of the models' execution by using only historical sales data, and the best result is highlighted in bold and marked with *. SARIMA model outperformed the other models, achieving a SMAPE of 15.30%, and yielded less MAE and RMSE results than other models. Furthermore, the clustering method also improves the results for the product-region breakdown. The experiment using the clustering method achieved 31% better results compared to the experiment based on the company's product-region categorization demonstrated in Table 14. The evaluation results based on each product-region group and comparison of the company's own product-region groups are shown in Appendix A.5. Moreover, graphical comparison of forecasted and actual values can be found in Figure 35. It shows that the SARIMA model is capable of accurately modeling both seasonality and trend dynamics, resulting in forecasts that align well with actual sales data.

Table 17: The model results on the Product-Region Sales Test Dataset by using only historical sales data

Model	**MAE**	**RMSE**	**SMAPE (%)**
BASE Model	13334	15396	18.55
***SARIMA**	**9731**	**11190**	**15.30**
LSTM	15750	19210	20.52
PROPHET	11636	14121	16.82

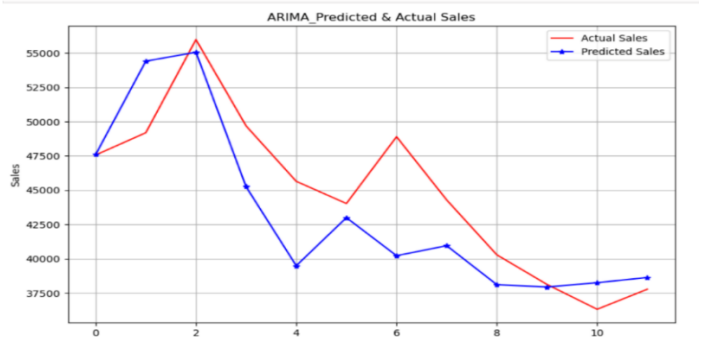


Figure 35: Average Predicted vs. Actual Sales Across Product Category-Sales Region Clusters

The models have been executed using various combinations of exogenous variables, the three best-performing experiments are presented in Table 18, and the best result is highlighted in bold and marked with *. Among the models, the SARIMA model demonstrated superior performance, with the best results observed in the third experiment, while the results of the second experiment are very close. Therefore, the incorporation of exogenous variables significantly enhances the model's performance for product-region breakdown. This improvement is particularly reasonable given that economic indicators vary across countries. The evaluation results for the Product-Region Sales Training Dataset are provided in Appendix B.2, and the results of the third experiment for each product-region group can be seen in Appendix A.6.

Table 18: The model results on the Product-Region Sales Test Dataset by using different exogenous variables

Exp.	Exogenous Variables	Model	MAE	RMSE	SMAPE (%)
1	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$Sales_Lag$_{11}$, Sales_Lag$_{12}$	SARIMA	7805	9763	15.64
		LSTM	15383	17875	22.01
		PROPHET	14870	18117	23.17
2	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ GDPGrowth_Lag$_{12}$, Unemployment_Lag$_{12}$, Interestrate_Lag$_{12}$	SARIMA	7522	9543	14.65
		LSTM	9609	11833	19.08
		PROPHET	10007	13925	18.89
3	Covid-19 Stock_Lag$_{12}$,Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ Unemployment_Lag_12	***SARIMA**	**6649**	**9633**	**14.11**
		LSTM	14345	16699	20.83
		PROPHET	13035	16352	21.79

4.2 Discussion

The model results could be analyzed based on the structure of trend patterns and seasonality. Trend and seasonal graphs shown in Appendix C.1 and C.2, and for both product category cluster level and product category-sales region cluster level, seasonality is consistently high. However, the degree of trend variability differs notably: product category clusters exhibit more dynamic and fluctuating trend structures, whereas product category-sales region clusters demonstrated more stable trends.

In line with these structural distinctions, Prophet performed best in product category clusters, where trend components were less stable. This aligns with Prophet's design, which allows for automatic detection of changepoints and accommodates complex, non-linear trend patterns (Chaturvedi, Rajasekar, Natarajan, & McCullen, 2022; Sulaimaan, Hariprasath, Dhinakaran, & Kose, 2023). By contrast, SARIMA outperformed other models in product category-sales region clusters, where trend behavior was smoother and more consistent. This supports SARIMA's well-documented strengths in modeling stationary and cyclic data with stable autocorrelation structures (Brykin D. , 2024; Ripatti, 2025).

When compared to Prophet and SARIMA, the LSTM model generally showed lower forecasting performance. Although LSTM is designed to capture long-term dependencies in time series data, its accuracy often drops in multi-step forecasting tasks. As noted by Lim et al. (2021), standard LSTM models can struggle in multi-horizon settings due to difficulties in learning complex time-based patterns and accumulating errors across steps (Lİm, Arık, Loeff, & Pfister, 2021).

Still, there are two notable findings about LSTM in this study. First, in some clusters with more volatile trends, LSTM performed better than both SARIMA and Prophet, and second, when exogenous variables were included, LSTM model with exogenous variables showed better results than the version without exogenous variables, though it still did not outperform the classical models overall. These results suggest that LSTM can benefit from trend variability and additional features, aligning with findings in prior studies where deep learning models demonstrated improved performance when capturing complex temporal dependencies (Song, Cai, Ma, & Li, 2024), but its overall performance was still weaker than Prophet and SARIMA.

Interestingly, the improved performance of LSTM model with exogenous variables compared to the version without exogenous variables may offer a different perspective on the findings of Lim et al. (2021). In their study, the authors reported that LSTM performance tends to degrade when future exogenous variables are unknown or must be estimated, which is often the case in real-world multi-step forecasting. However, in this study, the inclusion of such features improved the LSTM model's accuracy. This suggests that, under certain conditions, incorporating relevant past exogenous inputs can still enhance LSTM's forecasting ability

These findings emphasize that SARIMA remains a robust and reliable model under stable conditions characterized by consistent seasonal and trend patterns, while Prophet performs better in environments with more unstable dynamics. Although LSTM has shown the potential to perform well under unstable trend structures, and often improves its univariate performance when enriched with exogenous variables, it still falls short of achieving the forecasting accuracy observed with SARIMA and Prophet in most cases. This observation aligns with findings in prior studies (Brykin D. , 2024; Hewamalage, Bergmeir, & Bandara, 2021) which report that while deep learning models like LSTM can capture nonlinear patterns, they could underperform in comparison to classical statistical models, especially in datasets with strong seasonality and limited data availability.

CHAPTER 5

CONCLUSION

The motivation for this thesis to deal with the increasing complexity of supply chain management and the critical need for accurate sales forecasting in a rapidly changing global market. Technological advancements, fluctuating economic conditions, and global cases, such as the COVID-19 pandemic, have increased the importance of data-driven approaches in demand forecasting. This thesis focused on cooling and freezer products, aiming to enhance forecasting accuracy and provide actionable insights for supply chain optimization by utilizing real-world data from a global home appliances company.

5.1 Summary and Findings

This study highlights the important role of effective forecasting models in sales prediction, emphasizing the importance of integrating diverse methodologies with domain-specific knowledge. Initially, a comprehensive literature review has been conducted to examine various sales forecasting studies, particularly those dealing with seasonal data. Based on this review, three models-SARIMA, LSTM and Prophet- have been selected for analysis. These models have been implemented on the company's dataset, and hyperparameter tuning has been performed to optimize their performance.

The data structure and sales trends of the home appliances company have been analyzed, followed by data preprocessing steps. The company currently employs its own product and country grouping approach. In this thesis, alternative clustering techniques are applied to determine whether improved groupings for products or countries could be identified. After product and country clusters are established, feature engineering steps are undertaken. Potential exogenous variables, including Covid-19 deaths, economic indicators (e.g., unemployment rate, interest rate), inventory levels, and lagged sales data, have been integrated. While correlation analysis revealed that some features had negligible effects on sales, they have been incorporated into the modeling process to assess their potential impact. Finally, outlier detection has been conducted using the IQR method with a sliding window, and outliers have been handled using the winsorization method.

The three models have been applied to two datasets: the "Product Sales Dataset" and the "Product-Region Sales Dataset". These datasets enabled to generate forecasts at both the product cluster and product cluster-sales region cluster levels. MAE, RMSE, and SMAPE have been selected as evaluation metrics. After hyperparameter optimization, the models have been compared. For the product cluster breakdown, the

Prophet model yields the best performance with 11.69 % of SMAPE without using the exogenous variables, and the incorporation of exogenous variables did not improve model performance for this breakdown. In contrast, when exogenous variables were utilized, the best result was achieved by the LSTM model with a SMAPE of 12.62%, indicating that, at the product cluster level, excluding external factors led to better forecasting accuracy, and the best result outperforms the base model by 11%. Additionally, when the models were trained with the company's own categorization, the best performance was achieved by the PROPHET model with a SMAPE of 24.52. As a result, product clustering has improved performance by 52% compared to the company's own product categorization scenario. For the product cluster-sales region cluster breakdown, the SARIMA model, which incorporates exogenous variables, outperformed other models, yielding 14.11% of SMAPE, the best result outperforms the base model by 24%. Furthermore, the experiments show that factors such as Covid-19 and economic indicators, especially the unemployment rate, inventory level, and lagged sales, significantly influence the sales predictions for product cluster-sales region cluster breakdown. Additionally, when the models were trained with the company's own product-region categorization, the best performance was achieved by the SARIMA model with a SMAPE of 22.27%. As a result, product-region clustering has provided a 37% improvement compared to the company's own product-region categorization scenario.

5.2 Theoretical Contributions

This thesis contributes to the demand forecasting literature by focusing on a relatively underexplored area: the home appliance sector, particularly refrigerators and freezers. It further enriches the literature by integrating exogenous variables-such as inflation, unemployment rate, GDP growth, interest rate, and inventory levels-into the forecasting models, offering a more realistic and data-driven framework. In addition, the study provides a broader temporal perspective by examining sales patterns not only during the COVID-19 pandemic, but also across pre-pandemic and post-pandemic periods, which are less frequently addressed in previous research. Moreover, the observation that SARIMA and Prophet models outperformed LSTM provides an alternative perspective to the prevailing emphasis on machine learning in the literature. This finding suggests that traditional time series models may offer superior performance under certain conditions. In addition, the integration of a clustering-based segmentation approach-grouping products and countries by sales trend similarity-adds further methodological value by demonstrating how segmentation can improve forecasting accuracy and strategic decision-making.

5.3 Practical Contributions

This study also offers valuable contributions for the home appliance company. The results have been evaluated with the experts from different disciplines such as sales, marketing, supply chain, and product management, the experts have indicated that the product and country groupings provided by this study are closely aligned with market dynamics, and that having similar sales trends among products and countries within the same cluster would be more beneficial for strategy development. They particularly

emphasized that for country clustering, clustering based on sales trends would be more effective in terms of analyzing sales change and taking actions compared to the location-based grouping currently used by the company. This helps businesses design better sales strategies such as targeted campaigns, optimized stocking plans, specialized regional strategies and more effective product launches. The improved forecasting accuracy supports better inventory control and helps companies avoid stock shortages or excess. Overall, the results of this thesis can help the company to plan more effectively and increase their profitability in a competitive and seasonal market.

5.4 Limitations

Despite these contributions, this thesis has some limitations that should be considered. First, the study focused only on cooling and freezer products within the white goods sector, and other product categories haven't been analyzed. Second, the dataset used in this study covers a limited period, with consistent data available only from 2015 onward. A larger dataset could have allowed for a deeper understanding of long-term trends and seasonal effects. Moreover, additional exogenous variables, such as campaign information or the sales data of competing brands, could be used in the analysis. However, consistent campaign data and product-region-level sales data of competing brands haven't been accessible. Also, economic indicators were included in the analysis; however, the weighted average values of these indicators were calculated for each cluster. As a result, the direct variability within each country could not be directly included. Furthermore, this thesis explored K-Means clustering with Euclidean distance, which can be limiting in high-dimensional settings. Since each product category and country is represented by 108 monthly values, Euclidean distance may fail to reflect true similarity due to the curse of dimensionality. It is also sensitive to magnitude differences and may overlook trend-based patterns. Lastly, because of the time constraint, in hyperparameter optimization, all hyperparameter combinations couldn't have been explored.

5.5 Future Work

While this thesis proposes an effective forecasting approach focused on refrigerator and freezer products, it also points out potential areas for future research. First, future research with different forecasting or ensemble models may help improve the accuracy and adaptability of the results.

Second, the analysis could be further refined by conducting SKU-level forecasting. A detailed SKU-level analysis would better capture sales trends and demand variations and thereby enable companies to optimize their inventory planning, production scheduling, and distribution systems at a much deeper level. Moreover, this model can be transferred to different product categories which would enable the evaluation of its effectiveness and finding sales-related dynamics characterizing other home appliance sector or other sectors.

There is also room for improvement in the clustering analysis, particularly regarding feature extraction and the selection of distance metrics. Exploring alternative features and distance measures could help mitigate issues related to the curse of dimensionality. For example, extracting summary features or identifying key sub-sequences before performing clustering can enhance the quality of the analysis. In addition, alternative distance metrics, such as Pearson correlation, Spearman rank correlation, cross-correlation, or dynamic time warping, can be utilized to better capture the underlying similarities between data points.

Finally, future studies could focus on a single region or country so that more localized variables and unique socio-economic factors could be considered. For example, measuring the impact of regional changes in politics and marriage rates, or country-specific indicators, may provide better insights on the sales of regional markets. By following these suggestions, future research can improve demand forecasting methods and provide more accurate, localized, and practical insights.

REFERENCES

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 270-298.

Alharbi, F. R., & Csala, D. (2022). A Seasonal Autoregressive Integrated Moving Average with exogenous factors (Sarimax) forecasting model-based time series approach. *Inventions*, 7(94).

Ariyo, A., Adewumi, A., & Ayo, C. (2014). Stock price prediction using the Arima model. *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106-112. Cambridge: IEEE.

B., K., George, D. J., Manikandan, G., & Thomas, T. (2020). A comparative study on K-means clustering and Agglomerative hierarchical clustering. *International Journal of Emerging Trends in Engineering Research*, 1600-1604.

Bandara, K., Shi, P., Bergmeir, C., & Hewamalage, H. (2019). Sales demand forecast in e-commerce using a Long Short-Term Memory Neural Network methodology. *International Conference on Neural Information Processing*, 1-16. Sydney: Springer.

Bauerová, R., Starzyczná, H., & Pražákák, T. (2022). At what lag should economic indicators be applied to predict sales in a transformed economy? Is it worthwhile when e-tailing? *Forum Scientiae Oeconomia*, 29-46.

Brownlee, J. (2019, August 5). How to prepare univariate time series data for Long Short-Term Memory Networks. Retrieved January 12, 2025, from https://machinelearningmastery.com/prepare-univariate-time-series-data-long-short-term-memory-networks/

Brykin, D. (2024). Sales forecasting models: Comparison between ARIMA, LSTM and Prophet. *Journal of Computer Science*, 1222-1230.

Bukkapatnam, S. T., Kim, J., & Iyer, P. S. (2012). Multi-step sales forecasting in automotive industry based on structural relationship identification. *International Journal of Production Economics*, 875-887.

Chai, T., & Draxler, R. R. (2014). Root means square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247-1250.

Chaturvedi, S., Rajasekar, E., Natarajan, S., & McCullen, N. (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy*, 168 (113097).

Chawla, A. (2024, March 16). Breathing Kmeans vs Kmeans. Retrieved January 12, 2025, from https://blog.dailydoseofds.com/

Chopra, S., & Sodhi, M. (2004). Managing risk to avoid supply-chain breakdown. *MIT Sloan Management Review*, 52-63.

Cihan Çılgın, Y. G. (2023). The effect of outlier detection methods in real estate valuation with machine learning. *İzmir Sosyal Bilimler Dergisi*, 9-20.

Daraghmeh, M., Agarwal, A., Manzano, R., & Zaman, M. (2021). Time series forecasting using Facebook Prophet for cloud resource management. *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 1-7. Montreal: IEEE.

Djimasbe, R., Gyamfi, S., & Donatus, C. (2024). Development of an Arimax model for forecasting airport electricity consumption in Accra-Ghana: the role of weather and air passenger traffic. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 9 (100691).

Dmitry, I. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation- based analysis on the coronavirus outbreak (COVID-19 / SARS-CoV-2) case. *Transportation Research Part E Logistics and Transportation Review*, 136 (101922).

Duan, G., & Dong, J. (2024). Construction of ensemble learning model for home appliance demand forecasting. *Applied Sciences*, 1-17.

Dudukcu, H., Taskiran, M., & Taskiran, Z. (2021). Comprehensive comparison of LSTM variations for the prediction of chaotic time series. *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 1-5. Kocaeli: IEEE.

Gopaluni, B. (2023). Large-scale process models using deep learning. *Modelling of Chemical Process Systems*, 315-336. St. John's: Elsevier.

Gustriansyah, R., Ermatita, E., & Rini, D. P. (2022). An approach for sales forecasting. *Expert Systems with Applications*, 207 (118043).

Hamiche, K., Abouaïssa, H., Goncalves, G., & Hsu, T. (2018). A robust and easy approach for demand forecasting in supply chains. *International Federation of Automatic Control*, 1732-1737.

Haque, M. S., Amin, M. S., & Miah, J. (2023). Retail demand forecasting: A comparative study for multivariate time series. *Journal of Mathematics and Statistics Studies*, 40-46.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 388-427.

Houdt, G., Mosquera, C., & Napoles, G. (2020). A review on the Long Short-Term Memory model. *Artificial Intelligence Review*, 5929–5955.

Hyndman, R. J., & Athanasopoulos, G. (2021). Advanced forecasting methods. *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts.

Investment Office. (2024, March). WHITE GOODS HVAC, HOME APPLIANCES INDUSTRY OVERVIEW. Retrieved January 12, 2025, from https://www.invest.gov.tr/en/library/publications/lists/investpublications/white-goods-industry.pdf

Jha, B. K., & Pande, S. (2021). Time series forecasting model for supermarket sales using Fb-Prophet. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 547-554. Erode, India: IEEE.

K. Popat, S., & M., E. (2014). Review and comparative study of clustering techniques. *International Journal of Computer Science and Information Technologies*, 805-812.

Khashei, M., & Bijari, M. (2012). A new class of hybrid models for time series forecasting. *Expert Systems with Applications*, 4344-4357.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 90-95.

Kolambe, M. (2024). Forecasting the future: A comprehensive review of time Series prediction techniques. *Journal of Electrical Systems*, 575-586.

Korenius, T., Laurikkala, J., & Juhola, M. (2007). On principal component analysis, cosine and Euclidean measures in information retrieval. *Information Sciences*, 4893-4905.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon. *International Journal of Forecasting*, 1748-1764.

Market Research Future. (2024). Major home appliances market. Retrieved January 12, 2025, from https://www.marketresearchfuture.com/reports/major-home-appliances-market-27211

McKinsey Global Institute. (2021). The future of work after COVID-19. Retrieved January 12, 2025, from https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19

Menna, F. D., Traini, L., & Cortellessa, V. (2024). Time series forecasting of runtime software metrics: an empirical study. *15th ACM/SPEC International Conference on Performance Engineering*, 48-59. London: ACM Digital Library.

Meta Open Source. (2024). Prophet. Retrieved January 12, 2025, from https://facebook.github.io/prophet/docs/diagnostics.html

Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., & Chandran, V. (2018). Long Short-Term Memory hyperparameter optimization for a Neural Network Based Emotion Recognition framework. *IEEE*, 49325-49338.

Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, 34-42.

Panwar, A. A. (2021). Introduction to time series analysis and key concepts. Retrieved January 12, 2025, from https://medium.com/analytics-vidhya/introduction-to-time-series-analysis-and-key-concepts-dbf6c394984f

Parmezan, A. R., Souza, V., & Batista, G. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences*, 302-337.

Power, D. (2005). Supply chain management integration and implementation: a literature review. *An International Journal of Supply Chain Management*, 253-263.

Qiwei, H., Boylan, J., Huijing, C., & Labib, A. (2018). OR in spare parts management: A review. *European Journal of Operational Research*, 395-414.

R. Bhushan Gopaluni, L. C. (2023). Large-scale process models using deep learning. *In Modelling of Chemical Process Systems*, 315-336.

Rao, U. (2015). Development of statistical time series models for solar power prediction. *Renewable Energy*, 100-109.

Ribeiro, A., Seruca, I., & Durão, N. (2017). Improving organizational decision support: Detection of outliers and sales prediction for a pharmaceutical distribution company. *Procedia Computer Science*, 282-290.

Ripatti, V. (2025). Forecasting Prescription Medication. Espoo: Aalto University.

Singh, B., Kumar, P., Sharma, N., & Sharma, K. P. (2020). Sales forecast for Amazon sales with time series modeling. *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, 38-43. Raipur: IEEE.

Song, Y., Cai, C., Ma, D., & Li, C. (2024). Modelling and forecasting high-frequency data with jumps based on a hybrid nonparametric regression and LSTM model. *Expert Systems with Applications*, 237 (121527).

Statista. (2024, May). Household Appliances - Worldwide. Retrieved January 12, 2025, from https://www.statista.com/outlook/cmo/household-appliances/worldwide

Sulaimaan, R. B., Hariprasath, G., Dhinakaran, P., & Kose, U. (2023). Time-series forecasting of web traffic using Prophet Machine Learning Model. *FMDB Transactions on Sustainable Computer Letter*, 161-177.

Sun, J., Lee, H., & Yang, J. (2021). The Impact of the COVID-19 pandemic on the global value chain of the manufacturing industry. *Sustainability*, 32-54.

Svetunkov, I. (2023). Forecasting and analytics with the Augmented Dynamic Adaptive Model (ADAM). Retrieved January 12, 2025, from openforecast.org: https://openforecast.org/adam/

Tokuda, E. K., Comin, C. H., & Costa, L. d. (2022). Revisiting agglomerative clustering. *Physica A: Statistical Mechanics and its Applications*, 585 (126433).

Wang, F., & Aviles, J. (2023). Contrasting univariate and multivariate time series forecasting methods for sales: a comparative analysis. *Applied Science and Innovative Research*, 127-138.

Watanabe, S. (2023). Tree-Structured Parzen Estimator: Understanding its algorithm components and their roles for better empirical performance. *Arvix*, arXiv:2304.11127.

Weng, T., Liu, W., & Xiao, J. (2020). Supply chainSales forecasting based on lightGBM and LSTM combination model. *Industrial Management & Data Systems*, 265-279.

World Economic Forum. (2022). Annual Report 2021-2022. Retrieved January 12, 2025, from https://www.weforum.org/publications/annual-report-2021-2022/digest/

Yadav, M. (2020, December 30). Hierarchical Agglomerative Clustering. Retrieved January 12, 2025, from https://mrinalyadav7.medium.com/hierarchical-agglomerative-clustering-969cd147dfb5

Zonglei, C., Ma, M., Li, T., Wang, H., & Li, C. (2023). Long sequence time-series forecasting with deep learning: A survey. *Information Fusion*, 97 (101819).

Zunic, E., Korjenic, K., Hodzic, K., & Donko, D. (2020). Application of Facebook's Prophet algorithm for successful sales forecasting based on real-world data. *International Journal of Computer Science & Information Technology*, 23-36.

APPENDICES

APPENDIX A THE EVALUATION RESULTS BASED ON PRODUCT AND PRODUCT-REGION GROUP

A.1 Evaluation Results Based on Product Group used by the company

The evaluation results of each product group can be seen in Table 20, and the best result for each product cluster is highlighted in bold and marked with * for clarity. The average sales of test set observations for each product are demonstrated in Table 19.

Table 19: Average sales in the test set for each product group

Product	Average Sales
0-52 cm Cooling	16428
53-59 cm-Cooling	50513
60-63 cm-Cooling	30758
60-63 cm-Freezer	27472
70-73 cm-Cooling	77659
74-79 cm-Cooling	15229
80-89 cm-Cooling	11474

Table 20: The model results based on each product group used by the company

Product Cluster	Product	Model	MAE	RMSE	SMAPE (%)
1	0-52 cm Cooling	BASE	9800	11447	66.01
		SARIMA	9580	11140	79.38
		LSTM	10581	13253	64.01
		***PROPHET**	**8394**	**10037**	**56.51**
	74-79 cm-Cooling	***BASE**	**5077**	**6107**	**31.64**
		SARIMA	5427	5919	44.89
		LSTM	5375	6224	32.05
		PROPHET	7223	8376	42.63

Table 20 (cont.)

	80-89 cm-Cooling	BASE	4454	5219	38.58
		SARIMA	3382	3744	28.81
		***LSTM**	**2943**	**3984**	**25.73**
		PROPHET	3711	4288	31.02
2	60-63 cm-Freezer	***BASE**	**6929**	**9666**	**24.99**
		SARIMA	8870	11221	31.45
		LSTM	6879	10161	39.42
		PROPHET	8937	10033	32.87
3	53-59 cm-Cooling	BASE	12765	17011	25.84
		SARIMA	11879	13940	23.51
		LSTM	13434	16485	26.94
		***PROPHET**	**11259**	**13426**	**22.27**
	60-63 cm-Cooling	BASE	5599	7704	19.25
		***SARIMA**	**4130**	**5029**	**13.87**
		LSTM	5594	6385	19.73
		PROPHET	5508	6731	20.06
	70-73 cm-Cooling	BASE	14378	15800	19.39
		SARIMA	10889	13446	15.71
		LSTM	14437	17402	16.54
		***PROPHET**	**10071**	**11857**	**13.51**

A.2 Evaluation Results Based on Product-Region Group used by the company

The evaluation results of each product-product region group are shown in Table 22, and the best result for each product cluster is highlighted in bold and marked with * for clarity. The average sales of test set observations for each product-region group are demonstrated in Table 21.

Table 21: Average sales in test set for each product-region group

Product Cluster	Product-Region Group	Average Sales of Test Set
1	0-52 cm-Cooling-Turkey&South Asia	9883
	74-79 cm-Cooling-Turkey&South Asia	14010
	80-89 cm-Cooling-Turkey&South Asia	9158

Table 21 (cont.)

2	60-63 cm-Freezer-Europe	10646
	60-63 cm-Freezer-Turkey&South Asia	14584
3	53-59 cm-Cooling-Europe	34949
	60-63 cm-Cooling-Europe	24215
	70-73 cm-Cooling-Europe	7417
	70-73 cm-Cooling-Russia&CA&Caucasus&MEA	12027
	70-73 cm-Cooling-Turkey&South Asia	57730
4	53-59 cm-Freezer-Europe	6643
	53-59 cm-Freezer-Turkey&South Asia	2725

Table 22: The model results based on each product-region group used by the company

Product Cluster	Product-Region Group	Model	MAE	RMSE	SMAPE (%)
1	0-52 cm-Cooling-Turkey&South Asia	BASE	9475	11932	126.31
		***SARIMA**	**4415**	**5327**	**84.71**
		LSTM	10678	13220	113.58
		PROPHET	8103	10206	111.43
	74-79 cm-Cooling-Turkey&South Asia	***BASE**	**4465**	**5625**	**32.46**
		SARIMA	3637	4397	36.32
		LSTM	5919	6990	34.11
		PROPHET	4769	5463	33.68
	80-89 cm-Cooling-Turkey&South Asia	BASE	3798	4677	41.01
		SARIMA	3020	3503	34.29
		***LSTM**	**2840**	**3910**	**31.14**
		PROPHET	3876	4197	39.43
2	60-63 cm-Freezer-Europe	BASE	2591	2958	24.92
		***SARIMA**	**1918**	**2248**	**18.60**
		LSTM	2075	2687	25.98
		PROPHET	2946	3442	25.62
	60-63 cm-Freezer-Turkey&South Asia	BASE	6560	7710	61.20
		SARIMA	5084	5711	62.30
		LSTM	10648	13252	85.00
		***PROPHET**	**8141**	**11511**	**61.12**
3	53-59 cm-Cooling-Europe	BASE	9424	10976	27.11
		***SARIMA**	**4245**	**6366**	**10.65**

Table 22 (cont.)

		LSTM	5611	7521	17.29
		PROPHET	5846	7596	16.49
	60-63 cm-Cooling-Europe	BASE	4975	7142	21.09
		SARIMA	4357	5169	18.38
		LSTM	4889	5862	21.20
		***PROPHET**	**3690**	**4636**	**15.97**
	70-73 cm-Cooling-Europe	BASE	1402	1792	20.41
		***SARIMA**	**1237**	**1443**	**14.24**
		LSTM	1129	1364	17.69
		PROPHET	2275	2784	27.89
	70-73 cm-Cooling-Russia&CA&Caucasus&MEA	BASE	2157	2539	18.34
		***SARIMA**	**1300**	**1923**	**10.61**
		LSTM	2948	3535	23.02
		PROPHET	3711	4479	28.55
	70-73 cm-Cooling-Turkey&South Asia	BASE	14184	15618	26.88
		SARIMA	9690	11409	20.12
		LSTM	14625	19050	23.98
		***PROPHET**	**9855**	**12390**	**18.87**
4	53-59 cm-Freezer-Europe	BASE	3057	3668	44.09
		***SARIMA**	**1555**	**2102**	**12.19**
		LSTM	2050	2566	33.36
		PROPHET	2992	3598	41.23
	53-59 cm-Freezer-Turkey&South Asia	BASE	2223	3193	79.76
		***SARIMA**	**1695**	**2485**	**15.52**
		LSTM	2372	3288	100.12
		PROPHET	2724	3837	95.26

A.3 Evaluation Results Based on Product Group Analysis Using Only Historical Sales

The evaluation results of each product cluster are shown below, and the best result for each product cluster is highlighted in bold and marked with * for clarity. When the results are compared with the company's own product categorization scenario in Table 20 in Appendix A.1, the clustering approach produced better results for the product groups in product cluster 1 and product cluster 3. However, product cluster 2, which contains only one product group, showed no difference in results.

Table 23: The model results based on product group by using historical sales data

Product	Model	MAE	RMSE	SMAPE (%)
Product Cluster 1	BASE	8425	10226	15.00
	***SARIMA**	**7209**	**9711**	**12.88**
	LSTM	10687	12062	17.53
	PROPHET	9276	11454	16.44
Product Cluster 2	***BASE**	**6929**	**9666**	**24.99**
	SARIMA	9043	11448	32.00
	LSTM	8579	11182	38.47
	PROPHET	8816	9964	32.46
Product Cluster 3	BASE	16473	20686	10.46
	SARIMA	12264	16833	8.09
	LSTM	21442	26163	11.04
	***PROPHET**	**10433**	**14759**	**6.50**

A.4 Evaluation Results Based on Product Group Analysis Using Exogenous Variables

When incorporating exogenous variables, for the first and second product clusters LSTM yielded the best performance, for the third product cluster, Prophet outperformed the other models.

Table 24: The model results based on product group by using exogenous variables

Product	Model	MAE	RMSE	SMAPE (%)
Product Cluster 1	SARIMA	8837	11238	15.85
	***LSTM**	**6720**	**7272**	**12.15**
	PROPHET	11638	14260	22.72
Product Cluster 2	SARIMA	18263	20364	64.45
	***LSTM**	**9214**	**10287**	**33.94**
	PROPHET	13818	17693	58.12
Product Cluster 3	SARIMA	22627	27879	16.10
	LSTM	14277	17701	9.02
	***PROPHET**	**9728**	**13561**	**6.08**

A.5 Evaluation Results Based on Product-Region Group Analysis Using Only Historical Sales

The evaluation results of each product-sales region cluster can be shown below, and the best result for each product-sales region cluster is highlighted in bold and marked with * for clarity. The countries in the sales regions used by the company do not exactly match the countries in the sales region clusters we created. However, sales region cluster 3 can be matched with "Turkey & South Asia" and "Europe", while sales region cluster 4 generally contains the countries belonging to "Russia, CA, Caucasus, and MEA". Therefore, it would be more appropriate to evaluate the results for the same product cluster in this context. From this perspective, the clustering method has produced better results for the product-region groups in Product Cluster 1–Sales Region Cluster 3, Product Cluster 2–Sales Region Cluster 3, and Product Cluster 3–Sales region Cluster 3. However, it cannot be definitively asserted that better results were achieved for the product-region groups belonging to Product Cluster 3–Sales Region Cluster 4 and Product Cluster 4–Sales Region Cluster 3, as presented in Table 22.

Table 25: The model results based on the product-region group by using historical sales data

Product-Region Group	Model	MAE	RMSE	SMAPE (%)
Product Cluster 1-Sales Region Cluster 3	BASE	8957	10782	18.54
	SARIMA	8464	9880	20.55
	LSTM	9329	12125	18.38
	***PROPHET**	**8061**	**9319**	**16.53**
Product Cluster 2-Sales Region Cluster 3	BASE	7020	8025	31.57
	***SARIMA**	**6298**	**7574**	**30.68**
	LSTM	8544	10731	46.92
	PROPHET	7004	8884	33.87
Product Cluster 3-Sales Region Cluster 3	BASE	17773	20329	13.87
	***SARIMA**	**12913**	**14713**	**10.58**
	LSTM	21223	25623	14.23
	PROPHET	15216	18526	12.01
Product Cluster 3-Sales Region Cluster 4	BASE	3762	4780	23.41
	SARIMA	2973	3391	18.19
	LSTM	5222	5912	31.26
	***PROPHET**	**2962**	**3581**	**18.04**
Product Cluster 4-Sales Region Cluster 3	BASE	4291	4833	47.04
	***SARIMA**	**3228**	**3888**	**14.49**
	LSTM	4688	5672	41.39
	PROPHET	5263	7062	45.98

A.6 Evaluation Results Based on Product-Region Group Analysis Using Exogenous Variables

Incorporating the exogenous variable has improved the model performance except for second and fourth product-region group. Different models yield best performances for distinct product-sales region clusters, the results can be shown below.

Table 26: The model results based on the product-region group by using exogenous variables

Product- Sales Region Cluster	Model	MAE	RMSE	SMAPE (%)
Product Cluster 1-Sales Region Cluster 3	SARIMA	6592	8768	16.14
	***LSTM**	**6664**	**9221**	**13.43**
	PROPHET	7437	9552	15.07
Product Cluster 2-Sales Region Cluster 3	***SARIMA**	**6913**	**9232**	**34.46**
	LSTM	7274	9058	35.79
	PROPHET	12816	16122	71.28
Product Cluster 3-Sales Region Cluster 3	***SARIMA**	**7015**	**10781**	**5.6**
	LSTM	20687	23510	16.13
	PROPHET	16183	20362	12.33
Product Cluster 3-Sales Region Cluster 4	SARIMA	6098	7954	33.01
	***LSTM**	**2908**	**3542**	**17.87**
	PROPHET	9637	10734	44.44
Product Cluster 4-Sales Region Cluster 3	***SARIMA**	**4952**	**6675**	**22.88**
	LSTM	9969	10649	76.72
	PROPHET	3113	4284	34.72

APPENDIX B THE EVALUATION RESULTS ON THE TRAINING SETS

B.1 The Evaluation Results on The Product Sales Dataset

Firstly, models have been trained on the training set, for the product sales dataset, Prophet has outperformed other models, as shown in Table 27. When the exogenous variables were included, SARIMA yielded the best performance with the first experiment, results can be shown in Table 28.

Table 27: The model results on the Product Sales Training Dataset by using only historical sales data

Model	**MAE**	**RMSE**	**SMAPE (%)**
BASE Model	18798	22889	16.88
SARIMA	15266	22016	15.62
LSTM	24558	30385	25.11
***PROPHET**	**13335**	**16357**	**12.46**

Table 28: The model results on the Product Sales Training Dataset by using different exogenous variables

Exp.	**Exogenous Variables**	**Model**	**MAE**	**RMSE**	**SMAPE (%)**
1	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$	***SARIMA**	**10959**	**14690**	**11.05**
		LSTM	18598	23110	16.71
		PROPHET	13004	18338	12.68
2	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ GDPGrowth_Lag$_{12}$, Unemployment_Lag$_{12}$, Interestrate_Lag$_{12}$	SARIMA	11780	15163	11.38
		LSTM	17900	22501	16.46
		PROPHET	11826	16257	11.14
3	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ Unemployment_Lag_12	SARIMA	11345	15171	11.23
		LSTM	17385	21755	15.90
		PROPHET	13068	18271	12.90

B.2 The Evaluation Results on the Product-Region Sales Dataset

Models have been trained on the training set of "Product-Region Sales Dataset", Prophet has outperformed other models, as shown in Table 29. When the exogenous variables were included, SARIMA has yielded the best performance with third experiment, results are demonstrated in Table 30.

Table 29: The model results on the training set of Product-Region Sales Dataset by using only historical sales data

Model	**MAE**	**RMSE**	**SMAPE (%)**
BASE Model	13381	16937	20.59
SARIMA	9392	14325	16.72
LSTM	19623	23915	30.59
***PROPHET**	**11636**	**14121**	**14.17**

Table 30: The model results on the training set of the Product-Region Sales Training Dataset by using different exogenous variables

Exp.	**Exogenous Variables**	**Model**	**MAE**	**RMSE**	**SMAPE (%)**
1	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$	SARIMA	7559	10176	13.28
		LSTM	14552	17728	21.15
		PROPHET	9729	13045	15.47
2	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ GDPGrowth_Lag$_{12}$, Unemployment_Lag$_{12}$, Interestrate_Lag$_{12}$,	SARIMA	7767	9918	13.28
		LSTM	9594	12222	15.32
		PROPHET	9262	12183	14.50
3	Covid-19 Stock_Lag$_{12}$, Sales_Lag$_{10}$, Sales_Lag$_{11}$, Sales_Lag$_{12}$ Unemployment_Lag_12	***SARIMA**	**7463**	**10057**	**12.97**
		LSTM	14312	17200	21.21
		PROPHET	9286	12637	14.63

APPENDIX C THE STL DECOMPOSITION RESULTS

C.1 The Seasonal Decomposition Graphs for Product Category Cluster

Figure 36 presents the STL decomposition results of the sales time series for each product category cluster, illustrating the extracted trend, seasonal, and residual components.

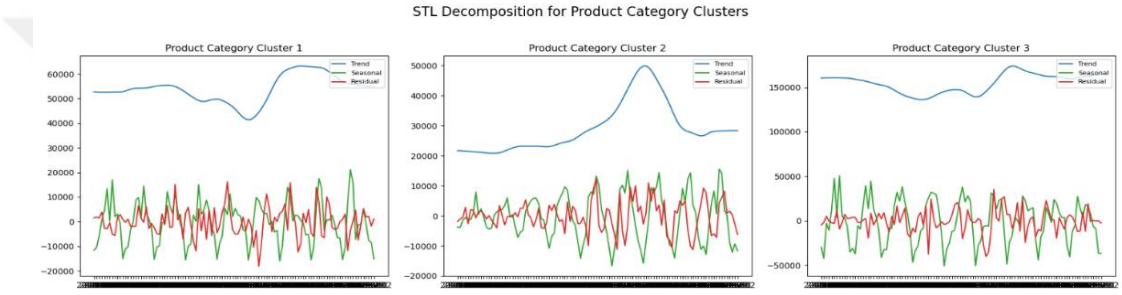


Figure 36: STL Decomposition for Product Category Clusters

C.2 The Seasonal Decomposition Graphs for Product Category Cluster-Sales Region Cluster

Figure 37 presents the STL decomposition results of the sales time series for each product category-sales region cluster, illustrating the extracted trend, seasonal, and residual components.

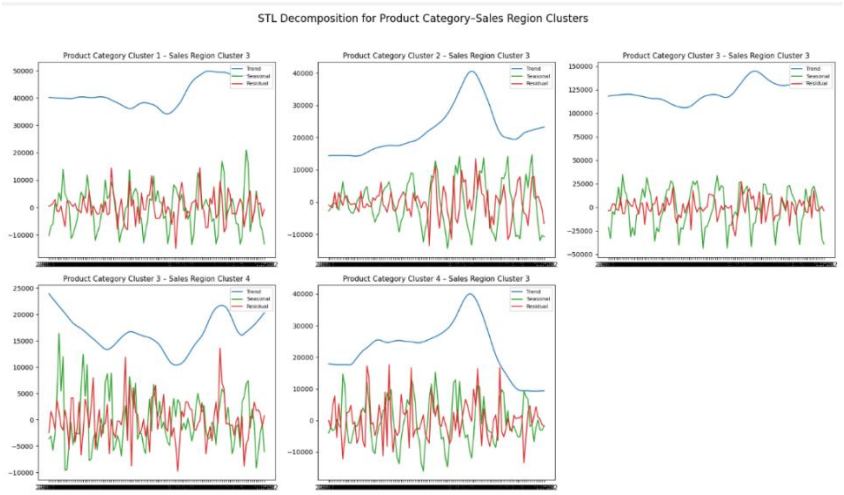


Figure 37: STL Decomposition for Product Category-Sales Region Clusters