Faculty of Computer
& Informatics Eng.
Departmant of Software
& Information.

الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

كليـــة هندســـة الحـــاســـوب
والمعلومـــاتيـــــة.
قســم هندســـة البــرمجيــات
ونظـم المعلومـــات.

Store’s Sales Forecasting

Minor Project - Completed the requirements for obtaining a bachelor's degree in Informatics Engineering – Software Engineering and Information Systems

Prepared by

Muhammad Al-Hussein

Muhammad Kenan Al-Akkad

Muhammad Fakher Aqeel

Supervised by

Dr. Kadan Al-joumaa

Academic year

2023-2024

Faculty of Computer
& Informatics Eng.
Departmant of Software
& Information.

كليـــة هندســـة الحـــاســـوب
والمعلومـــاتيـــــة.
قســـم هندســـة البـرمجيـــات
ونظـم المعلومـــات.

الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

التنبوء في مبيعات متجر

مشروع فصلي - قدم إستكمالاً لمتطلبات الحصول على درجة البكالوريوس في هندسة المعلوماتية - هندسة البرمجيات ونظم المعلومات

إعداد

محمد الحسين

محمد كنان العقاد

محمد فاخر عقيل

إشراف

د.كادان جمعة

السنة الدراسية

2024-2023

Faculty of Computer
& Informatics Eng.
Departmant of Software
& Information.

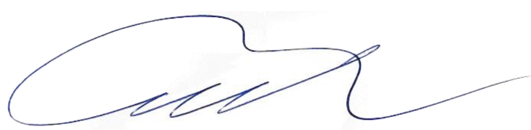
SPU
الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

كليــة هندســة الحــاســوب
والمعلومــاتيــــة.
قســم هندســة البــرمجيــات
ونظـم المعلومـــات.

Supervisor Certification

I certify that the preparation of this project entitled “Store’s Sales Forecasting”, prepared by Muhammad Al-Hussein, Muhammad Kenan Al-Akkad, and Muhammad Fakher Aqeel, was made under my supervision at Department of Software and Information System Engineering / Faculty of Computer and Informatics Engineering in partial fulfillment of the requirements of the Degree of Bachelors of Software and Information System Engineering.

Name	**Signature**	**Date**
Dr. Kadan Al-joumaa		1/6/2024



Abstract

Sales forecasting is a critical component of retail management, enabling businesses to optimize inventory, staffing, and marketing strategies. In this era of data-driven decision-making, traditional methods are giving way to advanced machine-learning techniques that offer greater accuracy and adaptability. This study explores the application of machine learning models, including time series analysis and regression in the context of in-store sales forecasting.

This research aims to provide insights on implementing machine learning-based forecasting models in physical retail stores. It delves into the various models, data preprocessing steps, and feature engineering techniques employed to predict sales accurately. Additionally, the study in upcoming projects will be developed to examines the impact of external factors, such as promotions, seasonality, and economic indicators, on sales predictions and demonstrates how these factors can be incorporated into the forecasting process.

Furthermore, the abstract discusses the benefits of probabilistic forecasting, which offers a more nuanced understanding of uncertainty in sales predictions. The study highlights the importance of model evaluation and selection, emphasizing the need for ongoing model refinement to adapt to changing market conditions and consumer behavior in an upcoming project.

Overall, this research sheds light on the potential of machine learning for in-store sales forecasting and its ability to provide retailers with a competitive edge in today's dynamic retail landscape as much as possible.

الملخص

يعد التنبؤ بالمبيعات عنصرًا حاسمًا في إدارة التجزئة، حيث يمكّن الشركات من تحسين المخزون والتوظيف واستراتيجيات التسويق. في هذا العصر الذي تحكمه القرارات المبنية على البيانات، تتخلى الطرق التقليدية لصالح تقنيات التعلم الآلي المتقدمة التي توفر دقة وقابلية للتكيف أكبر. تستكشف هذه الدراسة تطبيق نماذج التعلم الآلي، بما في ذلك تحليل السلاسل الزمنية والانحدار في سياق التنبؤ بالمبيعات داخل المتاجر.

يهدف هذا البحث إلى توفير رؤى حول تنفيذ نماذج التنبؤ المبنية على التعلم الآلي في المتاجر التجارية الفعلية. وتتعمق في النماذج المختلفة، خطوات معالجة البيانات، وتقنيات هندسة الميزات المستخدمة للتنبؤ بالمبيعات بدقة. بالإضافة إلى ذلك، ستُطور الدراسة في المشاريع القادمة لتُقيم تأثير العوامل الخارجية، مثل العروض الترويجية، والموسمية، والمؤشرات الاقتصادية، على توقعات المبيعات وتُظهر كيف يمكن دمج هذه العوامل في عملية التنبؤ.

علاوة على ذلك، يناقش الملخص فوائد التنبؤ الاحتمالي، الذي يوفر فهمًا أكثر دقة للشك في توقعات المبيعات. تسلط الدراسة الضوء على أهمية تقييم النموذج واختياره، مع التأكيد على الحاجة إلى تحسين النموذج المستمر للتكيف مع تغير ظروف السوق وسلوك المستهلكين في مشروع قادم.

بشكل عام، تسلط هذه الأبحاث الضوء على إمكانات التعلم الآلي في التنبؤ بالمبيعات داخل المتاجر وقدرتها على تزويد تجار التجزئة بميزة تنافسية في المشهد التجاري الديناميكي الحالي بقدر الإمكان.

Table of Contents

CHAPTER ONE XI

INTRODUCTION XI

INTRODUCTION: 2
I. Here are key aspects and methods involved in in-store sales forecasting: 2
THE PROS AND CONS OF STORE SALES FORECASTING: 3
THE AIM OF THE SALES FORECASTING PROJECT: 4
SALES FORECASTING MODELS: 4
I. Moving Averages: 4
II. Time Series Analysis: 4
III. Regression Analysis: 4
IV. Machine Learning Models: 4
V. Neural Networks: 4
VI. Seasonal Decomposition of Time Series (STL): 5
VII. ARIMA-X: 5
VIII. Prophet: 5
IX. Markov Chains: 5
X. Ensemble Methods: 5
REPORT SECTIONS: 5
SUMMARY: 6

CHAPTER TWO I

PROJECT TOPIC & A REFERENCE STUDY I

INTRODUCTION: 8
REFERENCE STUDY: 8
I. Time Series Analysis: 9
II. Machine Learning in Sales Forecasting: 9
III. Sales Forecasting Models: 10
METRICS EXPLORATION: 12
I. Linear Regression Metrics: 12
II. ARIMA Metrics: 13
III. Holt-Winters Metrics: 13
UNDERSTANDING AND LEVERAGING DATA: 13
I. What is Data? 13
II. How to Choose a Dataset? 14
III. Where to Find a Dataset? 14
IV. What is a Good Dataset? 14
V. What the Dataset Should Look Like? 14
VI. What the Dataset Should Have? 14
VII. How to Deal with the Dataset? 14
VIII. What Operations Should be Applied to the Dataset? 15
IX. Data Exploration Techniques: 15
X. Handling Time Series Data: 15

XI. Quality Assurance and Validation: 15
XII. Data Security and Privacy: 15
XIII. Continuous Monitoring and Updates: 15
USED TOOLS: 16
I. Python: 16
II. Anaconda & Jupyter Notebook: 16
III. Google: 16
IV. Excel: 17
V. Power BI: 17
VI. Data Warehouse: 18
VII. SQL Server: 18
VIII. Kaggle: 18
IX. YouTube: 19
X. ChatGPT: 19
SUMMARY: 20

CHAPTER THREE 9

ANALYTICAL STUDY & ENTRY-LEVEL MODELS 9

GANNT CHART: 22
MODEL SELECTION: 23
DATASET SELECTION: 24
DATASET VISUALIZATION: 24
VISUALIZATION BY POWER PI: 26
MODEL BUILDING: 31
MODEL RESULTS: 33
I. Linear regression model results: 33
II. ARIMA model results: 37
III. Holt-Winters Exponential Smoothing model results: 40
SUMMARY: 50

CHAPTER FOUR I

PRACTICAL APPLICATION I

MODEL MODIFICATION: 52
I. Comparison table for the entry-level model and primary model made by Linear regression: 52
II. Comparison table for the entry-level model and primary model made by ARIMA: 53
III. Comparison table for the entry-level model and primary model made by the Exponential smoothing model: 54
DATASET MANIPULATION: 55
MODEL RESULTS: 55
I. Linear regression results: 55
II. ARIMA results: 60
III. Exponential smoothing results: 65
SUMMARY: 73

CHAPTER FIVE I

RESULTS .... **I**
MODEL EVALUATION METRICS: LINEAR REGRESSION: 75
I. Mean Absolute Error (MAE): *75*
II. Root Mean Squared Error (RMSE): *75*
III. Mean Absolute Percentage Error (MAPE): *76*
IV. R-squared (R²) / Accuracy: *76*
MODEL EVALUATION METRICS: ARIMA: 76
I. Mean Absolute Error (MAE): *76*
II. Root Mean Squared Error (RMSE): *76*
III. R-squared (R²): *77*
IV. Adjusted R-squared (Adjusted R²): *77*
V. Accuracy: *77*
MODEL EVALUATION METRICS: HOLT-WINTERS TRIPLE EXPONENTIAL SMOOTHING: 77
I. Holt-Winters Model with No Damping (pred_fit_1): *77*
II. R-squared (R²): *78*
III. Accuracy: *78*
IV. Holt-Winters Model with Damping (pred_fit_2): *78*
V. Total Actual Sales and Predicted Sales: *78*
COMPARISON TABLE: 78
I. Item-Level Analysis: *79*
II. Store-Level Analysis: *80*
WHAT IS NEXT? 82
REFERENCES: 83

List of Tables

Table 1. 1. Forecasting Methods. 1
Table 1. 2. Forecasting Methods. 1
Table 1. 3. Key Aspects & Forecasting Methods. 2
Table 1. 4. Pros & Cons of Store Sales Forecasting. 3

Table 3. 1. Linear regressions Entry-level & Primary Models Comparison. 53
Table 3. 2. ARIMA Entry-level & Primary Models Comparison. 54
Table 3. 3. Exponential Smoothing Entry-level & Primary Models Comparison. 55

Table 5. 1. Comparing Models Results. 79
Table 5. 2. Comparing Models Results by Percentages. 81

List of Figures

Figure 3. 1. Gantt Chart. 23
Figure 3. 2. Yearly Sales for Store Items. 24
Figure 3. 3. yearly Sales for Each Item. 25
Figure 3. 4. Data in CSV File. 27
Figure 3. 5. Running Query to Import Data. 27
Figure 3. 6. Location of The Dataset for Import. 28
Figure 3. 7. Data in SQL Server. 28
Figure 3. 8. Schema Definition and Modification. 29
Figure 3. 9. Data Successfully Imported. 29
Figure 3. 10. Schema Diagram (Star Flake). 30
Figure 3. 11. Data Visualization by Power Pi. 30
Figure 3. 12. LR-ELM Forecast for Item One. 33
Figure 3. 13. LR-ELM Sales & Metrics for Item One. 33
Figure 3. 14. LR-ELM Forecast for Item Two. 34
Figure 3. 15. LR-ELM Sales & Metrics for Item Two. 34
Figure 3. 16. LR-ELM Forecast for Item Three. 34
Figure 3. 17. LR-ELM Sales & Metrics for Item Three. 35
Figure 3. 18. LR-ELM Forecast for Item Four. 35
Figure 3. 19. LR-ELM Sales & Metrics for Item Four. 35
Figure 3. 20. LR-ELM Forecast for Item Five. 35
Figure 3. 21. LR-ELM Sales & Metrics for Item Five. 36
Figure 3. 22. LR-ELM Forecast for Item Six. 36
Figure 3. 23. LR-ELM Sales & Metrics for Item Six. 36
Figure 3. 24. LR-ELM Forecast for Store Items. 36
Figure 3. 25. LR-ELM Sales & Metrics for Store Items. 37
Figure 3. 26. A-ELM Forecast for Item One. 37
Figure 3. 27. A-ELM Sales & Metrics for Item One. 37
Figure 3. 28. A-ELM Forecast for Item Two. 37
Figure 3. 29. A-ELM Sales & Metrics for Item Two. 38
Figure 3. 30. A-ELM Forecast for Item Three. 38
Figure 3. 31. A-ELM Sales & Metrics for Item Three. 38
Figure 3. 32. A-ELM Forecast for Item Four. 38
Figure 3. 33. A-ELM Sales & Metrics for Item Four. 39
Figure 3. 34. A-ELM Forecast for Item Five. 39
Figure 3. 35. A-ELM Sales & Metrics for Item Five. 39
Figure 3. 36. A-ELM Forecast for Item Six. 39
Figure 3. 37. A-ELM Sales & Metrics for Item Six. 40
Figure 3. 38. A-ELM Forecast for Store Items. 40
Figure 3. 39. A-ELM Sales & Metrics for Store Items. 40

Figure 3. 40. ES-ELM Forecast for Item One (without Dumping). 40
Figure 3. 41. ES-ELM Forecast for Item One (with Dumping). 41
Figure 3. 42. ES-ELM Sales & Metrics for Item One. 41
Figure 3. 43. ES-ELM Forecast for Item Two (without Dumping). 41
Figure 3. 44. ES-ELM Forecast for Item Two (with Dumping). 42
Figure 3. 45. ES-ELM Sales & Metrics for Item Two. 42
Figure 3. 46. ES-ELM Forecast for Item Three (without Dumping). 43
Figure 3. 47. ES-ELM Forecast for Item Three (with Dumping). 43
Figure 3. 48. ES-ELM Sales & Metrics for Item Three. 44
Figure 3. 49. ES-ELM Forecast for Item Four (without Dumping). 44
Figure 3. 50. ES-ELM Sales & Metrics for Item Four (with Dumping). 45
Figure 3. 51. ES-ELM Sales & Metrics for Item Four. 45
Figure 3. 52. ES-ELM Forecast for Item Five (without Dumping). 46
Figure 3. 53. ES-ELM Forecast for Item Five (with Dumping). 46
Figure 3. 54. ES-ELM Sales & Metrics for Item Five. 47
Figure 3. 55. ES-ELM Forecast for Item Six (without Dumping). 47
Figure 3. 56. ES-ELM Forecast for Item Six (with Dumping). 48
Figure 3. 57. ES-ELM Sales & Metrics for Item Six. 48
Figure 3. 58. ES-ELM Forecast for Store Items (without Dumping). 49
Figure 3. 59. ES-ELM Forecast for Store Items (with Dumping). 49
Figure 3. 60. ES-ELM Sales & Metrics for Store Items. 50

Figure 4. 1. LR-PM Forecast for Item One. 56
Figure 4. 2. LR-PM Forecast for Item Two. 56
Figure 4. 3. LR-PM Forecast for Item Three. 57
Figure 4. 4. LR-PM Forecast for Item Four. 57
Figure 4. 5. LR-PM Forecast for Item Five. 58
Figure 4. 6. LR-PM Forecast for Item Six. 58
Figure 4. 7. LR-PM Forecast for Store Items. 59
Figure 4. 8. LR-PM Metrics for Item One. 59
Figure 4. 9. LR-PM Metrics for Item Two. 59
Figure 4. 10. LR-PM Metrics for Item Three. 59
Figure 4. 11. LR-PM Metrics for Item Four. 59
Figure 4. 12. LR-PM Metrics for Item Five. 59
Figure 4. 13. LR-PM Metrics for Item Six. 59
Figure 4. 14. LR-PM Metrics for Store Items. 60
Figure 4. 15. LR-PM Sales for Item One. 60
Figure 4. 16. LR-PM Sales for Item Two. 60
Figure 4. 17. LR-PM Sales for Item Three. 60
Figure 4. 18. LR-PM Sales for Item Four. 60
Figure 4. 19. LR-PM Sales for Item Five. 60

Figure 4. 20. LR-PM Sales for Item Six. 60
Figure 4. 21. LR-PM Sales for Store Items. 60
Figure 4. 22. A-PM Forecast for Item One. 60
Figure 4. 23. A-PM Forecast for Item Two. 61
Figure 4. 24. A-PM Forecast for Item Three. 61
Figure 4. 25. A-PM Forecast for Item Four. 62
Figure 4. 26. A-PM Forecast for Item Five. 62
Figure 4. 27. A-PM Forecast for Item Six. 63
Figure 4. 28. A-PM Forecast for Store Items. 63
Figure 4. 29. A-PM Metrics for Item One. 64
Figure 4. 30. A-PM Metrics for Item Two. 64
Figure 4. 31. A-PM Metrics for Item Three. 64
Figure 4. 32. A-PM Metrics for Item Four. 64
Figure 4. 33. A-PM Metrics for Item Five. 64
Figure 4. 34. A-PM Metrics for Item Six. 64
Figure 4. 35. A-PM Metrics for Store Items. 64
Figure 4. 36. A-PM Sales for Item One. 64
Figure 4. 37. A-PM Sales for Item Two. 64
Figure 4. 38. A-PM Sales for Item Three. 64
Figure 4. 39. A-PM Sales for Item Four. 64
Figure 4. 40. A-PM Sales for Item Five. 64
Figure 4. 41. A-PM Sales for Item Six. 64
Figure 4. 42. A-PM Sales for Store Items. 64
Figure 4. 43. ES-PM Forecast for Item One (without Dumping). 65
Figure 4. 44. ES-PM Forecast for Item One (with Dumping). 65
Figure 4. 45. ES-PM Forecast for Item Two (without Dumping). 66
Figure 4. 46. ES-PM Forecast for Item Two (with Dumping). 66
Figure 4. 47. ES-PM Forecast for Item Three (without Dumping). 67
Figure 4. 48. ES-PM Forecast for Item Three (with Dumping). 67
Figure 4. 49. ES-PM Forecast for Item Four (without Dumping). 68
Figure 4. 50. ES-PM Forecast for Item Four (with Dumping). 68
Figure 4. 51. ES-PM Forecast for Item Five (without Dumping). 69
Figure 4. 52. ES-PM Forecast for Item Five (with Dumping). 69
Figure 4. 53. ES-PM Forecast for Item Six (without Dumping). 70
Figure 4. 54. ES-PM Forecast for Item Six (with Dumping). 70
Figure 4. 55. ES-PM Forecast for Store Items (without Dumping). 71
Figure 4. 56. ES-PM Forecast for Store Items (with Dumping). 71
Figure 4. 57. ES-PM Sales & Metrics for Item One. 72
Figure 4. 58. ES-PM Sales & Metrics for Item Two. 72
Figure 4. 59. ES-PM Sales & Metrics for Item Three. 72
Figure 4. 60. ES-PM Sales & Metrics for Item Four. 72
Figure 4. 61. ES-PM Sales & Metrics for Item Five. 72

Figure 4. 62. ES-PM Sales & Metrics for Item Six. 72
Figure 4. 63. ES-PM Sales & Metrics for Store Items. 72

List of Abbreviations

ARIMA	Autoregressive Integrated Moving Average
MAE	The average of the absolute differences between predicted and actual values. It gives a measure of the average magnitude of errors.
RMSE	Similar to MAE but emphasizes larger errors. It's the square root of the average of squared differences between predicted and actual values.
MAPE	Measures the percentage difference between predicted and actual values. It provides a relative error as a percentage.
R2	Measures the proportion of the variance in the dependent variable that is predictable from the independent variable.
Pros	Advantages
cons	Disadvantages
LR-ELM	Linear Regression Entry-Level Model
A-ELM	ARIMAR Entry-Level Model
ES-ELM	Exponential Smoothing Entry Level Model
LR-PM	Linear Regression Primary Model
A-PM	ARIMA Primary Model
ES-PM	Exponential Smoothing Primary Model

Chapter One

Introduction

Chapter One – Introduction

Forecasting is a process of making predictions or estimates about future events based on past and present data and analysis. It involves using various methods and techniques to anticipate future trends, outcomes, or developments in a specific area. Forecasting is commonly applied in various fields, including finance, economics, weather, sales, demand planning, and many others.

There are different methods of forecasting, and the choice of method depends on the nature of the data, the context of the prediction, and the available resources. Some common forecasting methods include:

Time Series Analysis	Regression Analysis	Machine Learning Models
This involves analyzing historical data to identify patterns and trends over time. Techniques such as moving averages, exponential smoothing, and autoregressive integrated moving average (ARIMA) models are often used in time series analysis.	Regression models are used when there is a relationship between the variable to be predicted and one or more independent variables. It helps in understanding the strength and nature of the relationship.	Advanced techniques like machine learning algorithms, including decision trees, neural networks, and support vector machines, can be employed for forecasting tasks. These models are capable of handling complex patterns and relationships in data.

Table 1. 1. Forecasting Methods.

Qualitative Methods	Simulation and Scenario Analysis
In situations where historical data is limited or not available, qualitative methods involve expert judgment, market research, and subjective opinions to make predictions.	Simulation involves creating a model that represents a system and then running different scenarios to observe the outcomes. This is often used in complex systems where various factors interact.
The accuracy of forecasts can vary, and it's important to consider the limitations and uncertainties associated with any forecasting method. While forecasting provides valuable insights for decision-making, it's essential to regularly update predictions as new data becomes available and conditions change.	

Table 1. 2. Forecasting Methods.

In this chapter, we will be looking at:

1. **Introduction.**
2. **The Pros and Cons of Sales Forecasting.**
3. **The Aim of sales forecasting.**
4. **Sales forecasting models.**
5. **Report sections.**
6. **Summary.**

Introduction:

What is store sales forecasting? Store sales forecasting is a specific type of forecasting that focuses on predicting the future sales of a retail store. This process involves estimating the amount of revenue a store is likely to generate over a specific period, typically on a daily, weekly, monthly, or seasonal basis. Accurate sales forecasting is crucial for effective inventory management, staff scheduling, and overall business planning.

I. Here are key aspects and methods involved in in-store sales forecasting:

Historical Data Analysis	**Time Series Analysis**	**Seasonal Adjustments**	**Promotion and Marketing Impact**
Retailers often analyze historical sales data to identify patterns, trends, and seasonality. This involves looking at sales figures for comparable periods in the past to understand how sales have behaved under similar conditions.	Time series forecasting techniques, such as moving averages, exponential smoothing, and ARIMA models, are commonly used in in-store sales forecasting. These methods take into account past sales data to project future sales.	Many retail businesses experience seasonality, with sales patterns influenced by factors like holidays, weather, or cultural events. Adjusting for these seasonal variations helps in making more accurate predictions.	Anticipating the impact of promotions, discounts, and marketing campaigns is essential. Store sales forecasting should consider how these factors influence customer behavior and purchasing patterns.
Market Trends and External Factors	**Inventory Levels**	**Customer Behavior Analysis**	**Technology and Data Analytics**
External factors, such as economic conditions, industry trends, and changes in consumer behavior, can impact store sales. Analyzing and incorporating these factors into the forecasting process enhances its accuracy.	Sales forecasting is closely linked to inventory management. Predicting sales accurately helps retailers maintain optimal inventory levels, reducing the risk of stockouts or overstock situations.	Understanding customer behavior, preferences, and purchasing patterns is crucial. This can involve analyzing data from loyalty programs, customer surveys, and other sources to gain insights into what drives sales.	increasingly used in in-store sales forecasting. These technologies can analyze large datasets quickly and identify complex patterns that may not be apparent through traditional methods.

Table 1. 3. Key Aspects & Forecasting Methods.

Accurate store sales forecasting helps retailers optimize their operations, improve customer satisfaction, and ultimately maximize profitability. It's an ongoing process that requires continuous refinement and adjustment based on actual sales data and changing market conditions.

The Pros and Cons of Store Sales Forecasting:

While store sales forecasting offers numerous benefits, it's important for businesses to be aware of its limitations and continually refine their methods based on real-time data and market feedback. The goal is to strike a balance between precision and adaptability in response to changing conditions.

Pros	Cons
Improved Inventory Management: Accurate sales forecasting enables retailers to maintain optimal inventory levels. This reduces the risk of stockouts, excess inventory, and associated costs.	**Inaccuracy and Uncertainty:** Forecasting is inherently uncertain, and inaccuracies can arise due to unforeseen events, changes in consumer behavior, or external factors.
Effective Staff Scheduling: With reliable sales predictions, retailers can schedule staff more efficiently, aligning workforce levels with expected customer traffic.	**Dependency on Historical Data:** Over-reliance on historical data may result in inaccurate predictions if significant changes occur in the market or if there are disruptions that deviate from historical patterns.
Cost Reduction: Efficient inventory management and staff scheduling lead to cost savings. Retailers can avoid unnecessary holding costs for excess inventory and optimize labor costs.	**Complexity of External Factors:** External factors such as economic conditions, political events, or unexpected market trends can be challenging to predict accurately, making it difficult to account for them in forecasts.
Enhanced Customer Satisfaction: Maintaining adequate stock levels ensures that customers can find the products they want when they visit the store, improving overall customer satisfaction.	**Data Quality Issues:** Poor data quality, incomplete data, or inaccurate data can negatively impact the accuracy of forecasts. It's crucial to ensure that the data used in forecasting is reliable.
Strategic Decision-Making: Sales forecasting provides valuable insights for strategic planning. It helps retailers make informed decisions about promotions, marketing campaigns, and expansion plans.	**Overemphasis on Short-Term Results:** Some forecasting methods may focus too much on short-term results, potentially overlooking long-term trends or shifts in consumer behavior.
Optimized Supply Chain: Retailers can work closely with suppliers to align production and delivery schedules with forecasted demand, creating a more streamlined and efficient supply chain.	**Resistance to Change:** Implementing forecasting processes may face resistance from employees or management who may be skeptical of the accuracy of predictions or resistant to adopting new methods.
Resource Allocation: Knowing when peak sales periods are likely to occur allows for better allocation of resources, both in terms of personnel and marketing efforts.	**Dynamic Market Conditions: R**apid changes in market conditions, technological advancements, or competitive landscapes can make it challenging for forecasts to keep pace with the evolving business environment.
Improved Cash Flow: By preventing overstock situations and reducing holding costs, accurate forecasting contributes to improved cash flow for the business.	**Model Complexity:** Advanced forecasting models, including machine learning algorithms, can be complex and require specialized expertise. Implementing and maintaining such models may pose challenges for some businesses.

Table 1. 4. Pros & Cons of Store Sales Forecasting.

The Aim of the Sales Forecasting Project:

The primary aim of this project is to develop a robust and accurate sales forecasting model. By doing so, we hope to mitigate the risks associated with inaccurate predictions and enable businesses to make informed decisions regarding their inventory management, production planning, and budget allocation.

Sales Forecasting Models:

Several forecasting models can be used for store sales forecasting, ranging from simple to more complex methods. The choice of a particular model depends on the nature of the data, the available resources, and the level of accuracy required. Here are some common store sales forecasting models:

I. **Moving Averages:**
   - **Description:** Moving averages involve calculating the average of a certain number of past data points to predict future sales.
   - **Application:** Simple Moving Average (SMA), Weighted Moving Average (WMA), and Exponential Moving Average (EMA) are commonly used.

II. **Time Series Analysis:**
   - **Description:** Time series analysis involves studying past sales data to identify patterns, trends, and seasonality. This forms the basis for predicting future sales.
   - **Application:** Autoregressive Integrated Moving Average (ARIMA) models are frequently used for time series forecasting.

III. **Regression Analysis:**
   - **Description:** Regression models establish relationships between the sales variable and other relevant independent variables, such as marketing spend, promotional activities, or economic indicators.
   - **Application:** Linear regression, multiple regression, and polynomial regression are examples.

IV. **Machine Learning Models:**
   - **Description:** Advanced machine learning algorithms can be employed for more complex and dynamic forecasting tasks. These models can handle non-linear relationships and intricate patterns in the data.
   - **Application:** Decision trees, Random Forests, Support Vector Machines, Neural Networks, and Gradient Boosting are examples of machine learning models used in sales forecasting.

V. **Neural Networks:**
   - **Description:** Neural networks, a subset of machine learning, simulate the human brain's ability to learn. They are capable of handling complex patterns and relationships in data.
   - **Application:** Feedforward neural networks and recurrent neural networks (RNNs) are commonly used in sales forecasting.

VI. **Seasonal Decomposition of Time Series (STL):**
- **Description:** STL decomposes time series data into three components: seasonal, trend, and remainder. This allows for a more granular analysis of patterns.
- **Application:** Particularly useful when dealing with data with strong seasonality.

VII. **ARIMA-X:**
- **Description:** Extends the ARIMA model by incorporating exogenous variables, such as external factors or marketing activities, to enhance forecasting accuracy.
- **Application:** Suitable when external variables significantly impact sales.

VIII. **Prophet:**
- **Description:** Developed by Facebook, Prophet is designed for forecasting with daily observations that display patterns on different time scales. It can handle missing data and outliers.
- **Application:** Particularly useful for retail sales forecasting due to its ability to handle holidays and special events.

IX. **Markov Chains:**
- **Description:** Markov chains model transitions between different states based on probabilities. It's useful when the future state depends only on the current state and not on the sequence of events that preceded it.
- **Application:** Suitable for predicting sales transitions between different product categories or states.

X. **Ensemble Methods:**
- **Description:** Ensemble methods combine the predictions of multiple models to improve overall accuracy and robustness.
- **Application:** Bagging (Bootstrap Aggregating), Boosting, and Stacking are examples of ensemble methods.

When implementing store sales forecasting models, it's crucial to continually evaluate and refine the models based on the latest data and business insights. The choice of a specific model depends on factors like the availability of data, the complexity of the sales patterns, and the business context.

Report Sections:

This report is divided into several Chapters. After this introduction, we will discuss the Project Topic & and a Reference Study in more detail. This will be followed by a chapter on the analytical Study and entry-level Models. We will then delve into the Practical Application. Finally, we will conclude the results of the project.

Chapter One – Introduction

Summary:

The chapter presents a comprehensive view of sales forecasting, emphasizing its importance in business planning. It discusses the complexity of predicting sales, impacted by various factors, and aims to develop an accurate forecasting model to reduce risks from incorrect predictions. The report covers different models like Time Series Analysis, Regression Analysis, and Machine Learning, each with their strengths and weaknesses. It also addresses challenges in sales forecasting and proposes enhancements such as improving data quality and leveraging advanced analytics. The goal is to enable informed business decisions through robust sales forecasting.

Chapter Two

Project Topic & a Reference Study

Chapter Two – Project Topic & Reference Study.

Store sales forecasting is a critical process in the retail industry that involves predicting a retail store's future sales performance. This strategic undertaking relies on the analysis of historical sales data, market trends, and various influencing factors to anticipate consumer demand accurately. The primary goal is to optimize operational aspects such as inventory management, staff scheduling, and resource allocation. Different forecasting methods, including time series analysis, regression models, and advanced machine learning algorithms, are employed to generate precise predictions. By understanding past sales patterns, seasonal variations, and external influences like promotions or economic conditions, retailers can make informed decisions to enhance customer satisfaction, streamline supply chains, and ultimately maximize profitability. Store sales forecasting is an ongoing and dynamic process, requiring continuous refinement and adaptation to changing market conditions and consumer behavior.

In this chapter, we will be looking at:

1. **Introduction.**
2. **Reference Study.**
3. **Metrics exploration.**
4. **Understanding and Leveraging Data.**
5. **Used Tools.**
6. **Summary.**

Introduction:

In the realm of business analytics, the art of store sales forecasting plays a pivotal role in strategic decision-making for retail enterprises. This chapter serves as the foundation for a comprehensive exploration into the essential domain of store sales forecasting. The chapter begins by elucidating the critical significance of store sales forecasting in optimizing various facets of retail operations. It then delves into a substantial reference study that meticulously dissects three prominent forecasting models—linear regression, ARIMA, and triple exponential smoothing. These models are chosen for their proven efficacy in predicting store sales with varying degrees of complexity. Furthermore, the chapter features a dedicated section named "Main Concepts," which illuminates the fundamental principles of store sales forecasting. This section elucidates the core concepts underpinning the forecasting process and provides an in-depth overview of the three selected models, laying the groundwork for the subsequent analytical exploration.

Reference study:

Embarking on the intricate landscape of sales forecasting methodologies, Time Series Analysis and Machine Learning emerge as powerful tools, each bringing its unique set of characteristics, advantages, and considerations to the table. Furthermore, the specific forecasting models—Linear Regression, ARIMA, and Holt-Winters Triple Exponential Smoothing—contribute distinctive strengths and limitations to the predictive analytics toolkit.

Chapter Two – Project Topic & Reference Study.

I. Time Series Analysis:

- **Definition:** Time Series Analysis is a method meticulously designed to model time-dependent data, focusing on forecasting future trends by scrutinizing critical patterns such as seasonality, trends, and cyclic behavior.
- **Pros:**

  i. **Effectiveness in Capturing Patterns:** Proficient in capturing and leveraging seasonal and historical patterns.
  ii. **Clear Trend Analyses:** Provides lucid trend analyses for better strategic planning.
  iii. **Suitability for Stable Markets:** Particularly effective in stable markets exhibiting consistent trends.

- **Cons:**

  i. **Assumption of Stability:** Assumes stability in historical patterns, posing challenges in rapidly changing markets.
  ii. **Sensitivity to Outliers:** May not adapt well to unexpected events and can be sensitive to outliers.
  iii. **Comparison:** Time Series Analysis excels in stable datasets with clear trends and seasonality. However, it may be less adaptable to sudden market changes compared to the flexibility offered by machine learning approaches.

- **Comparison:** Highly effective for stable datasets with clear trends and seasonality but less adaptable to sudden market changes or non-linear patterns compared to machine learning approaches.
- **Examples:** Predicting monthly sales figures based on historical data, and forecasting quarterly financial performance.

II. Machine Learning in Sales Forecasting:

- **Definition:** Machine Learning involves utilizing various algorithms to learn from and make predictions based on data, identifying complex patterns and relationships that traditional methods might miss.
- **Pros:**

  i. **Handling Complex Relationships:** Capable of handling non-linear and intricate relationships in data.
  ii. **Adaptation to New Trends:** Quick adaptation to new trends and incorporation of diverse data types and sources.
  iii. **Flexibility:** More flexible and powerful in handling dynamic, non-linear patterns.

- **Cons:**

  i. **Data Requirements:** Requires large amounts of data for effective model training.
  ii. **Complex Setup:** this can be complex to set up and understand.
  iii. **Data Dependency:** The accuracy of predictions heavily depends on the quality and quantity of the data.
  iv. **Comparison:** Machine learning stands out for its flexibility and adaptability to dynamic markets, especially suited for handling complex, non-linear patterns. However, it requires a substantial amount of data and computational resources compared to traditional Time Series Analysis.

- **Comparison:** More flexible and powerful in handling complex, non-linear patterns and adapting to new data, making it suitable for dynamic and rapidly changing markets. However, it requires more data and computational resources than traditional time series analysis.
- **Examples:** Using algorithms to predict future sales based on customer behavior, incorporating social media data for sales forecasts.

III. Sales Forecasting Models:

1. Linear Regression:

- **Definition:** Linear Regression is a statistical approach that models the relationship between a dependent variable and one or more independent variables.
- **Pros:**

  i. **Simplicity and Interpretability:** Simple and interpretable, providing a clear understanding of relationships.
  ii. **Transparent Results:** Offers transparent results, making it accessible for non-technical users.

- **Cons:**

  i. **Oversimplification:** May oversimplify complex trends and ignore seasonal or cyclic patterns.
  ii. **Limited Flexibility:** Less flexible in capturing intricate relationships compared to more complex models.
  iii. **Comparison:** Linear Regression, while straightforward, may not capture complex trends and patterns as effectively as more intricate models.

- **Comparison:** Straightforward to interpret but may lack sophistication in capturing intricate patterns compared to time series or machine learning methods.
- **Examples:** Predicting sales based on advertising spend, estimating revenue growth based on marketing efforts.

2. **ARIMA (Autoregressive Integrated Moving Average):**

- **Definition:** A complex model combining autoregression, differencing (to achieve stationarity), and a moving average. Effective for a wide range of time series data, especially non-seasonal patterns.
- **Pros:**

  i. **Versatility:** Effective for a wide range of time series data, especially non-seasonal patterns.
  ii. **Comprehensive Approach:** Incorporates autoregression, differencing, and a moving average for a holistic analysis.

- **Cons:**

  i. **Parameter Sensitivity:** Requires careful parameter selection, and may struggle with seasonal data.
  ii. **Complexity:** The complexity of the model may be challenging for some users.
  iii. **Comparison:** ARIMA is versatile, especially for non-seasonal patterns, but demands careful consideration of parameters and may not be ideal for datasets with strong seasonality.

- **Comparison:** This offers more flexibility for non-seasonal data compared to linear regression but may require more fine-tuning for seasonality.
- **Examples:** Forecasting monthly sales with a focus on non-seasonal trends, and predicting stock prices.

3. **Holt-Winters Triple Exponential Smoothing:**

- **Definition:** Tailored for data with trends and seasonal patterns, applying three smoothing equations to capture level, trend, and seasonality.
- **Pros:**

  i. **Seasonal Handling:** Particularly strong in handling seasonality, making it suitable for data with trends and seasonality.
  ii. **Holistic Approach:** Applies three smoothing equations for comprehensive analysis.

- **Cons:**

  i. **Configuration Complexity:** Can be complex to configure, especially for users unfamiliar with the method.
  ii. **Effectiveness with Non-Seasonal Data:** Less effective when dealing with datasets lacking strong seasonality.
  iii. **Comparison:** Holt-Winters excels in capturing seasonal fluctuations but requires careful configuration and may not be as effective with non-seasonal data.

- **Comparison:** Excels in capturing seasonal fluctuations but may be less versatile in scenarios without clear seasonality.
- **Examples:** Predicting quarterly sales with distinct seasonality, forecasting holiday season retail demand.

the choice of a sales forecasting methodology and model hinges on the specific characteristics of the data and the objectives of the business. Time Series Analysis and Machine Learning offer distinct advantages and considerations, and the selection of Linear Regression, ARIMA, or Holt-Winters depends on the nuances of the dataset at hand. The comprehensive overview provides valuable insights for businesses navigating the complexities of sales forecasting.

Metrics Exploration:

Evaluating the performance of forecasting models is essential to gauge their accuracy and reliability in predicting future outcomes. Metrics play a crucial role in quantifying the effectiveness of models across different dimensions. In this analysis, we delve into the key metrics employed to assess the performance of three distinct forecasting models: Linear Regression, ARIMA (Autoregressive Integrated Moving Average), and Holt-Winters. Each metric offers unique insights into the models' predictive capabilities, shedding light on their strengths and areas for improvement.

I. Linear Regression Metrics:

- **Mean Absolute Error (MAE):** The average of the absolute differences between predicted and actual values. It gives a measure of the average magnitude of errors.
- **Root Mean Squared Error (RMSE):** Similar to MAE but emphasizes larger errors. It's the square root of the average of squared differences between predicted and actual values.
- **Mean Absolute Percentage Error (MAPE):** Measures the percentage difference between predicted and actual values. It provides a relative error as a percentage.
- **R-squared (Accuracy):** Represents the proportion of the variance in the dependent variable that is predictable from the independent variable. R-squared ranges from 0 to 1, where 1 indicates perfect predictions.

II. ARIMA Metrics:

- **Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):** Similar to linear regression, they measure the average magnitude of errors.
- **R-squared (R2):** Measures the proportion of the variance in the dependent variable that is predictable from the independent variable.
- **Adjusted R-squared (Adjusted R2):** Similar to R2 but adjusts for the number of predictors, providing a more accurate measure when there are multiple predictors.

III. Holt-Winters Metrics:

- **Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2):** Similar to linear regression, they measure the average magnitude of errors and the proportion of predictable variance.
- **Adjusted R-squared (Adjusted R2):** Adjusts R2 for the number of predictors, providing a more accurate measure in the presence of multiple predictors.

These metrics help assess the accuracy and reliability of forecasting models, providing insights into how well the model's predictions align with actual data.

Understanding and Leveraging Data:

In the realm of store sales forecasting, the foundation lies in the effective understanding and utilization of data. Data, in its raw form, serves as the bedrock upon which accurate predictions and informed decision-making rest. In this section, we delve into the intricacies of what constitutes good data, how to choose an appropriate dataset, and the essential considerations when dealing with data for store sales forecasting. From dataset selection criteria to preprocessing operations and continuous monitoring, a comprehensive exploration awaits to guide businesses through the pivotal process of harnessing the power of data for optimal forecasting outcomes.

I. What is Data?

- Define data as raw facts and figures that can be processed to obtain meaningful information. In the context of store sales forecasting, data includes information about historical sales, customer behavior, promotions, economic indicators, and other relevant factors.

II. How to Choose a Dataset?

- Discuss the criteria for selecting a dataset. Consider factors such as the relevance of the data to your forecasting goals, the timeframe of the data, and the availability of key variables like sales figures, promotional data, and external factors.

III. Where to Find a Dataset?

- Explore potential sources for datasets. This could include internal data from the business, publicly available datasets from government agencies, industry-specific databases, or platforms like Kaggle that host a variety of datasets for different domains.

IV. What is a Good Dataset?

- Define a good dataset as one that is relevant, comprehensive, and representative of the factors influencing store sales. A good dataset should have a sufficient volume of data, minimal missing values, and include all relevant variables for accurate forecasting.

V. What the Dataset Should Look Like?

- Describe the ideal structure of a dataset for sales forecasting. It should typically have a timestamp for each data point, variables such as sales, promotional indicators, and any other relevant features. The dataset should be organized, with clear labels and consistent formatting.

VI. What the Dataset Should Have?

- Enumerate the essential variables a dataset for store sales forecasting should contain. This may include:
- Time-related variables (e.g., timestamps).
- Sales figures.
- Customer-related data.
- Information on promotions or marketing activities.
- External factors (e.g., economic indicators).

VII. How to Deal with the Dataset?

- Discuss the initial steps in handling a dataset for sales forecasting. This includes data cleaning to address missing values and outliers, exploring the distribution of variables, and checking for data consistency.

VIII. What Operations Should be Applied to the Dataset?

- Describe preprocessing operations such as normalization, scaling, or encoding categorical variables. Explain why these operations are necessary to ensure that the dataset is suitable for input into forecasting models.

IX. Data Exploration Techniques:

- Discuss exploratory data analysis (EDA) techniques to gain insights into the dataset. This involves visualizing trends, distributions, and correlations among variables to inform the modeling process.

X. Handling Time Series Data:

- If dealing with time series data, elaborate on techniques for handling temporal aspects. This includes identifying seasonality, trends, and any cyclical patterns in the sales data.

XI. Quality Assurance and Validation:

- Stress the importance of quality assurance in the dataset and how validation techniques, such as cross-validation, can be applied to ensure the reliability of the data for training and testing forecasting models.

XII. Data Security and Privacy:

- Touch on the importance of data security and privacy, especially if dealing with sensitive customer information. Emphasize compliance with relevant regulations and the ethical handling of data.

XIII. Continuous Monitoring and Updates:

- Highlight that the dataset is not static and should be continuously monitored and updated. New data points should be incorporated to improve the accuracy of forecasting models over time.

By addressing these aspects, we can handle and utilize data for store sales forecasting, from dataset selection to ongoing maintenance and improvement.

Used Tools:

The evolution of our sales forecasting models was empowered by a diverse array of tools, each playing a distinctive role in shaping the project's success:

I. **Python:**

- **Definition:** Python, a versatile programming language, served as the backbone for data analysis and model building.
    - **Pros:**

        i. **Versatility:** Python's versatility allowed for comprehensive data analysis and efficient model development.

        ii. Rich Ecosystem: A rich ecosystem of libraries (e.g., NumPy, Pandas, Scikit-learn) enhanced functionality for statistical analysis and machine learning.

    - **Cons:**

        i. **Learning Curve:** For those unfamiliar, there might be an initial learning curve associated with Python.

II. **Anaconda & Jupyter Notebook:**

- **Definition:** Anaconda, alongside Jupyter Notebook, provides an interactive environment for coding, data visualization, and model development.
- **Pros:**

    i. **Interactivity:** An interactive platform fosters real-time adjustments and exploration during the coding and modeling process.

    ii. **Ease of Sharing:** Jupyter Notebook facilitates easy sharing and collaboration

- **Cons:**

    i. **Resource Intensive:** Anaconda can be resource-intensive, requiring substantial disk space.

III. **Google:**

- **Definition:** Google tools were utilized for research and accessing cloud-based data and services.
- **Pros:**

i. **Cloud Integration:** Seamless integration with cloud-based resources facilitated efficient data access and storage.
ii. **Collaboration Tools:** Tools like Google Docs and Sheets provide collaborative platforms for team interaction.

- **Cons:**
- **Privacy Concerns:** Depending on the nature of data, there might be privacy concerns associated with cloud-based services.

IV. Excel:

- **Definition:** Excel played a pivotal role in preliminary data analysis and visualization.
- **Pros:**

i. **User-Friendly:** Excel's user-friendly interface made it accessible for various team members.
ii. **Quick Analysis:** Rapid data analysis and visualization capabilities for initial insights.

- **Cons:**

i. **Limitations with Big Data:** Excel may have limitations when dealing with large datasets.

V. Power BI:

- **Definition:** Power BI contributed to creating dynamic dashboards and visualizations for interpreting forecasting results.
- **Pros:**

i. **Visualization Capabilities:** Powerful visualization tools facilitated clear and interactive representation of forecasting outcomes.
ii. **Integration with Other Tools:** Seamless integration with other Microsoft products enhanced overall efficiency.

- **Cons:**

i. **Learning Curve:** Power BI's advanced features may have a learning curve for new users.

VI. Data Warehouse:

- **Definition:** A centralized data storage system for comprehensive analysis.
- **Pros:**

  i. **Centralization:** Centralized data storage streamlines data accessibility and management.
  ii. **Comprehensive Analysis:** Enables comprehensive analysis by consolidating data from various sources.

- **Cons:**

  i. **Implementation Complexity:** Setting up and maintaining a data warehouse can be complex.

VII. SQL Server:

- **Definition:** SQL Server was employed for managing and querying large datasets efficiently.
- **Pros:**

  i. **Query Performance:** Efficient handling of large datasets through optimized query performance.
  ii. **Scalability:** Scalability to accommodate growing data needs.

- **Cons:**

  i. **Licensing Costs:** Depending on the version, SQL Server may involve licensing costs.

VIII. Kaggle:

- **Definition:** Kaggle provided access to a diverse range of datasets and community-driven insights for model training and benchmarking.
- **Pros:**

  i. **Diverse Datasets:** Access to diverse datasets enriched the model training process.
  ii. **Community Insights:** Community discussions and competitions offered valuable insights and benchmarking opportunities.

- **Cons:**

    i. **Limited Control:** The datasets on Kaggle may not always align with specific project requirements.

IX. YouTube:

- **Definition:** YouTube served as a resource for tutorials and expert talks, enhancing understanding and implementation of complex forecasting concepts.
- **Pros:**

    i. **Visual Learning:** Video tutorials facilitated visual learning for complex concepts.
    ii. **Expert Insights:** Access to expert talks provided deeper insights into advanced forecasting techniques.

- **Cons:**

    i. **Quality Variability:** The quality of tutorials may vary, impacting the consistency of learning.

X. ChatGPT:

- **Definition:** ChatGPT assisted in providing theoretical knowledge, answering queries, and offering guidance on data analysis techniques.
- **Pros:**

    i. **Information Assistance:** ChatGPT served as a valuable resource for theoretical knowledge and guidance.
    ii. **Query Resolution:** Provided real-time answers to specific queries during the project.

- **Cons:**

    i. **Limited to Text-Based Interaction:** Interaction limitations to text-based queries may be a constraint for some users.

These tools, in concert, facilitated every stage of our sales forecasting process, from the initial stages of data gathering and processing to the intricate phases of analysis, visualization, and interpretation. Their collective strengths and considerations formed a robust foundation for a comprehensive and successful sales forecasting endeavor.

Summary:

Store sales forecasting is a crucial process in retail, predicting a store's future sales by analyzing historical data and market trends. This optimizes operations like inventory management. The chapter introduces the significance of forecasting, a reference study on key models (linear regression, ARIMA, triple exponential smoothing), and explores main concepts. Time Series Analysis and Machine Learning are highlighted, with models offering unique strengths and limitations. Then comes the part that talks about data, it's important to understand and leverage data for Store Sales Forecasting. The tools used in the process, such as Python, Google, and Kaggle, are outlined. Together, they form a robust foundation for sales forecasting, aiding in data analysis, model development, and decision-making in the retail sector.

Chapter Three

Analytical Study & Entry-Level Models

Embarking on sales forecasting requires adept navigation through a meticulously crafted roadmap, where each step enhances precision in predictive models. Initial model curation strategically selects from diverse forecasting models to align with specific business goals. Core to this process is the meticulous dataset selection, considering pertinence, thoroughness, and alignment with business objectives. Visualizing datasets through graphics unveils critical patterns, guiding subsequent modeling decisions. The methodical model-building phase, from preprocessing to parameter tuning, ensures a model adept at capturing business intricacies. Culminating in presenting model outcomes through graphics, this iterative process underscores the dynamic relationship between each phase, achieving a refined and accurate sales forecasting model.

In this chapter, we will be looking at:

1. **Gantt Chart.**
2. **Model selection.**
3. **Dataset Selection**
4. **Dataset visualization.**
5. **Visualization by Power Pi.**
6. **Models building.**
7. **Models results.**
8. **Summary.**

Gannt Chart:

This Gantt chart provides a visual timeline of the project's key milestones and activities planned out over the coming months. It outlines the schedule for seminars, phases of research, model development, and the stages of report preparation. Each task is carefully plotted to reflect its start date, duration, and end date, offering a clear overview of the project's trajectory. This chart serves as a dynamic tool to track progress and ensure that all components of the project are completed in a timely and systematic fashion.

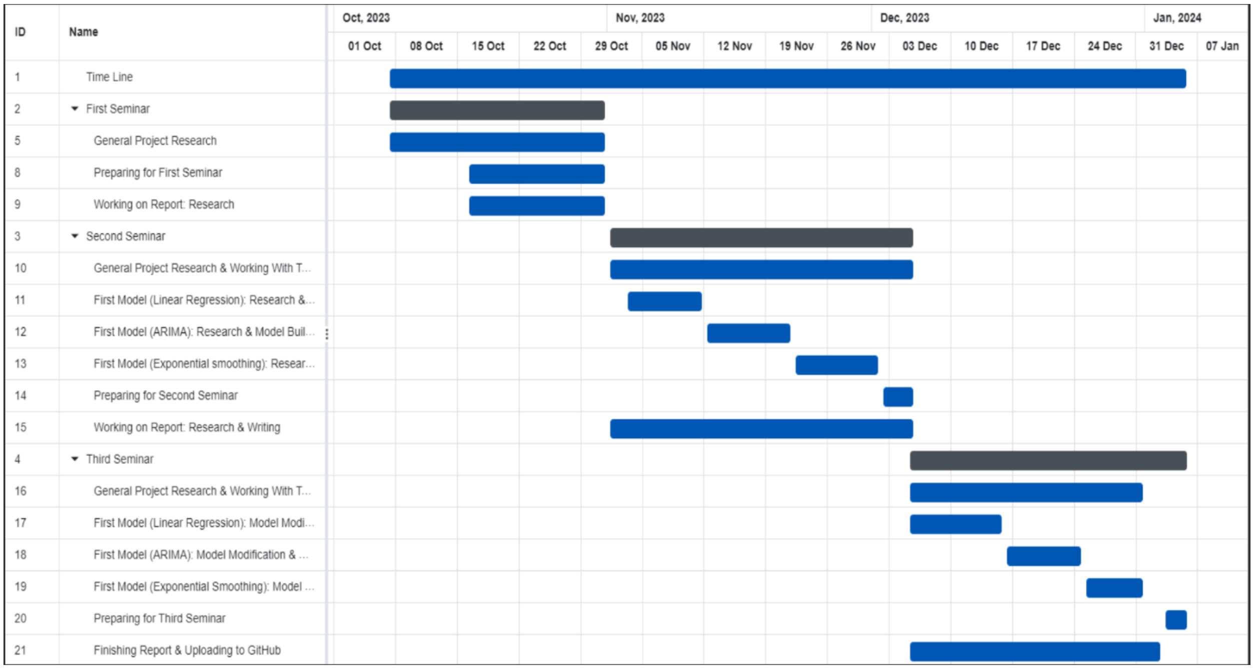


Figure 3. 1. Gantt Chart.

Model selection:

The steps for model selection in sales forecasting typically include:

1. **Define Business Objectives:** Understand what you want to achieve with the forecasting model (e.g., short-term vs. long-term forecasting).
2. **Analyze Data Availability:** Review the data you have, considering its volume, variety, and quality.
3. **Understand Data Characteristics:** Identify patterns in the data like seasonality, trends, and cyclicality.
4. **Choose Relevant Models:** Based on the data and objectives, select potential models (like time series, and machine learning algorithms).
5. **Test and Evaluate Models:** Implement models on a subset of data to assess their accuracy and efficiency.
6. **Consider Model Complexity vs. Performance:** Balance the complexity of the model with the level of accuracy needed.
7. **Iterate and Optimize:** Continuously refine the model based on testing results and business feedback.

The selection of Linear Regression, ARIMA, and Holt-Winters Triple Exponential Smoothing models in sales forecasting follows a structured approach. This involves defining business objectives, analyzing data availability, understanding data characteristics, and selecting models that align with these factors. Linear Regression is chosen for its simplicity in interpreting variable relationships, ARIMA for handling complex, non-seasonal data, and Holt-Winters for its proficiency in seasonal trends. These models are then rigorously tested and evaluated for accuracy, balancing complexity against performance, and continually optimized based on results and feedback. This methodology ensures the chosen models are well-suited for diverse forecasting needs.

Dataset selection:

Choosing the right dataset for sales forecasting based on previous steps is crucial for effective sales forecasting. These steps, including defining objectives, assessing data relevance, and checking data quality, ensure the chosen dataset aligns with specific forecasting goals. This methodical approach led us to Kaggle, a platform known for its diverse and quality datasets. On Kaggle, we identified a dataset that met our defined criteria, including historical sales data with relevant external factors, thereby providing a robust foundation for our forecasting models. This careful selection process is essential for accurate and reliable forecasting outcomes.

In the initial stages of the project, I meticulously examined the dataset to extract valuable insights. This examination was crucial to understanding the nature of the data and informing subsequent decisions. I then split the data into training and testing sets, a standard practice in model development to evaluate performance and generalizability. During the model development phase, I experimented with various modifications to the dataset. This process involved tweaking and adjusting data inputs and model parameters to test and enhance the model's predictive accuracy.

Dataset Visualization:

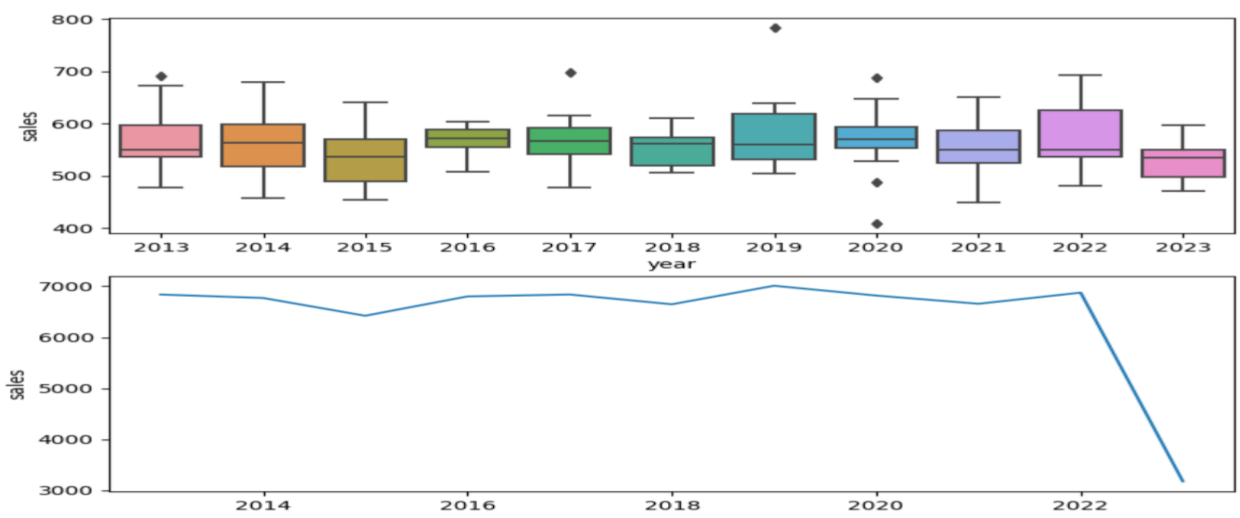


Figure 3. 2. Yearly Sales for Store Items.

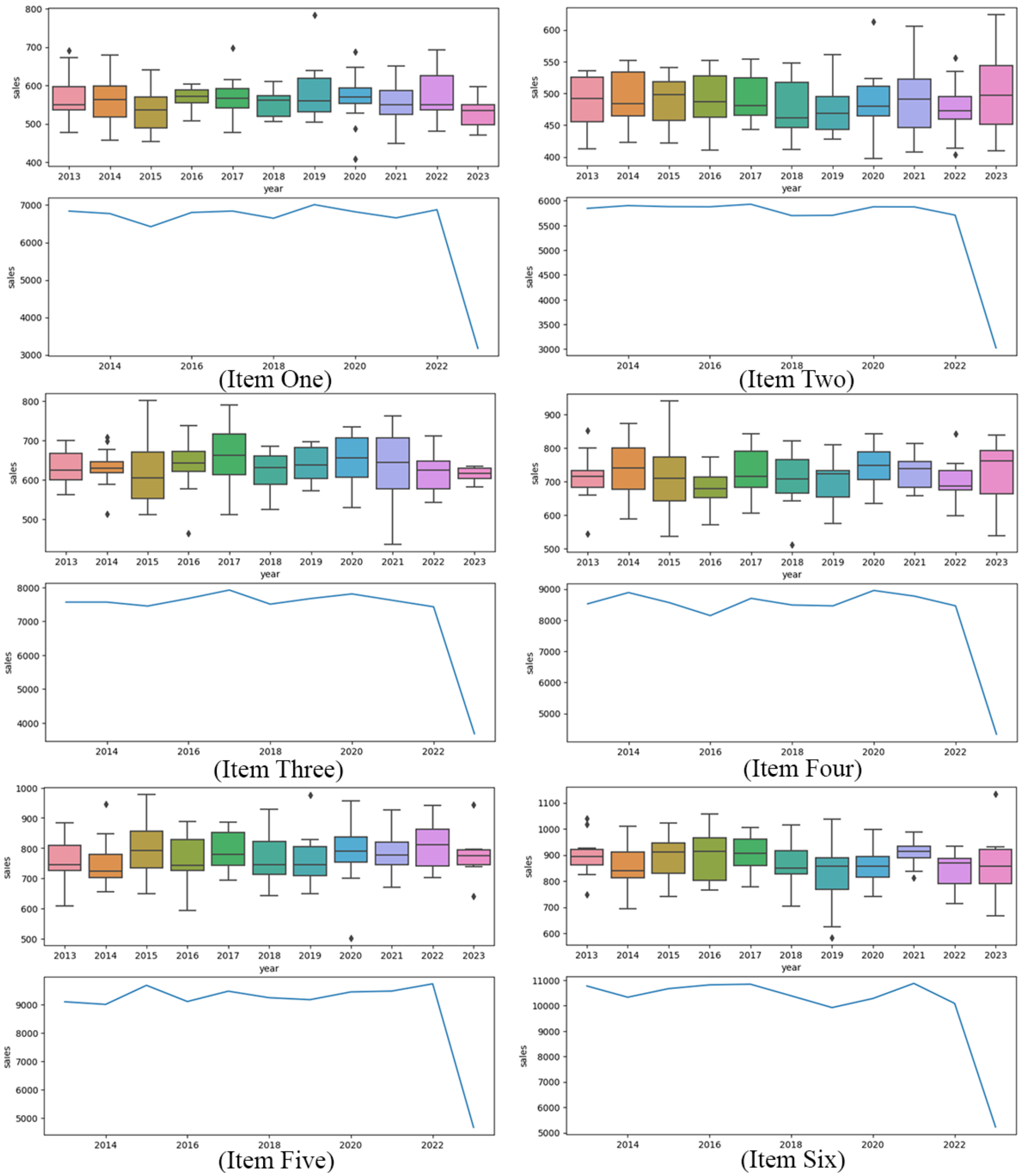


Figure 3. 3. yearly Sales for Each Item.

Visualization by Power Pi:

In the realm of data analysis and business intelligence, the ability to forecast sales is crucial for making informed decisions that drive company growth. The process of analyzing store sales encompasses several technical steps, from the initial data upload to the final visualization. This report documents a comprehensive walkthrough of uploading a sales dataset into Power BI through SQL Server, structuring the data within a defined schema, and leveraging Power BI's robust visualization tools to forecast store sales. The following visual aids illustrate each pivotal stage in the data's journey from raw numbers to actionable insights.

With the foundational dataset in place, the subsequent step involves constructing a data warehouse, a centralized repository designed to facilitate reporting and analysis. The importance of a data warehouse lies in its ability to consolidate data from multiple sources, providing a unified platform for robust data operations. This centralization is crucial for comprehensive analytics, as it enables the performance of complex queries and advanced operations on large datasets without impacting the performance of transactional systems.

In this data warehouse, data operations such as cleaning, transformation, and enrichment take precedence. These operations ensure that the data is accurate, consistent, and in a format that is conducive to insightful analysis. Once the data is prepped and stored in the data warehouse, we move to the visualization step using Power BI.

The process of data visualization in Power BI typically follows these steps:

Connect Power BI to the Data Warehouse: Establish a direct connection to the data warehouse, allowing Power BI to access the prepared datasets.

I. **Model the Data:** Use Power BI's data modeling features to define relationships, calculate columns, measures, and create hierarchies.
II. **Design the Visuals:** Craft visual representations such as charts, graphs, and tables to best illustrate the data's story.
III. **Refine the Dashboard:** Apply filters, slicers, and drill-down capabilities to enhance the interactivity of the dashboard.
IV. **Share Insights:** Share the dashboard with stakeholders, providing them with actionable insights and the ability to make data-driven decisions.

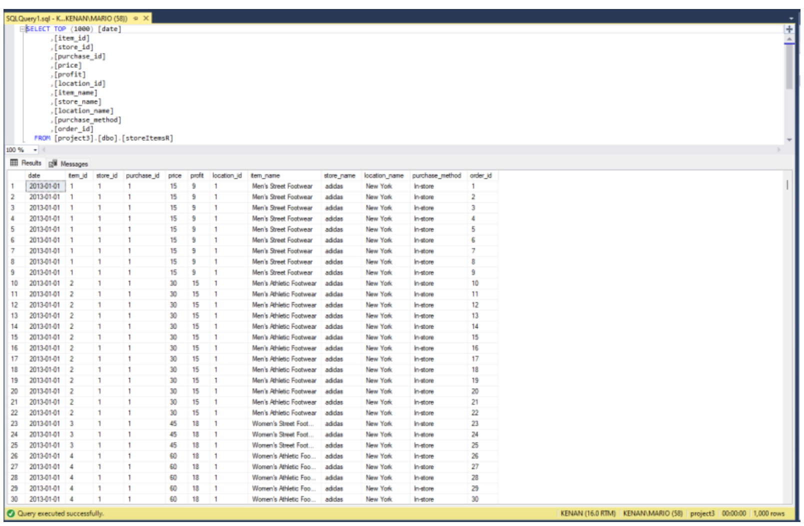
By leveraging the power of a data warehouse and the advanced visualization tools of Power BI, organizations can not only predict future trends but also make informed strategic decisions that align with their business objectives.

I. **Data Upload Overview:** This image displays a spreadsheet with detailed sales data, including order IDs, dates, item specifics, and profit margins, ready for upload to SQL Server for further analysis.

order_id	date	item_id	item_name	price	profit	location_id	location_name	store_id	store_name	purchase_id	purchase_method
1	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
2	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
3	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
4	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
5	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
6	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
7	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
8	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
9	1/1/2013	1	Men's Street Footwear	15	9	1	New York	1	adidas	1	In-store
10	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
11	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
12	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
13	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
14	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
15	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
16	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
17	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
18	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
19	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
20	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
21	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
22	1/1/2013	2	Men's Athletic Footwear	30	15	1	New York	1	adidas	1	In-store
23	1/1/2013	3	Women's Street Footwear	45	18	1	New York	1	adidas	1	In-store
24	1/1/2013	3	Women's Street Footwear	45	18	1	New York	1	adidas	1	In-store
25	1/1/2013	3	Women's Street Footwear	45	18	1	New York	1	adidas	1	In-store
26	1/1/2013	4	Women's Athletic Footwear	60	18	1	New York	1	adidas	1	In-store
27	1/1/2013	4	Women's Athletic Footwear	60	18	1	New York	1	adidas	1	In-store
28	1/1/2013	4	Women's Athletic Footwear	60	18	1	New York	1	adidas	1	In-store

Figure 3. 4. Data in CSV File.

II. **SQL Server Query Execution:** Here we see a SQL Server Management Studio (SSMS) window with a query result set, indicating the successful execution of a SELECT statement that retrieves the top records from the dataset.


SQLQuery1.sql - K...KENAN\MARIO (58))

```sql
SELECT TOP (1000) [date]
      ,[item_id]
      ,[store_id]
      ,[purchase_id]
      ,[price]
      ,[profit]
      ,[location_id]
      ,[item_name]
      ,[store_name]
      ,[location_name]
      ,[purchase_method]
      ,[order_id]
  FROM [project3].[dbo].[storeItemsR]
```

100 %

Results | Messages

	date	item_id	store_id	purchase_id	price	profit	location_id	item_name	store_name	location_name	purchase_method	order_id
1	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	1
2	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	2
3	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	3
4	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	4
5	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	5
6	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	6
7	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	7
8	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	8
9	2013-01-01	1	1	1	15	9	1	Men's Street Footwear	adidas	New York	In-store	9
10	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	10
11	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	11
12	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	12
13	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	13
14	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	14
15	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	15
16	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	16
17	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	17
18	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	18
19	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	19
20	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	20
21	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	21
22	2013-01-01	2	1	1	30	15	1	Men's Athletic Footwear	adidas	New York	In-store	22
23	2013-01-01	3	1	1	45	18	1	Women's Street Foot...	adidas	New York	In-store	23
24	2013-01-01	3	1	1	45	18	1	Women's Street Foot...	adidas	New York	In-store	24
25	2013-01-01	3	1	1	45	18	1	Women's Street Foot...	adidas	New York	In-store	25
26	2013-01-01	4	1	1	60	18	1	Women's Athletic Foo...	adidas	New York	In-store	26
27	2013-01-01	4	1	1	60	18	1	Women's Athletic Foo...	adidas	New York	In-store	27
28	2013-01-01	4	1	1	60	18	1	Women's Athletic Foo...	adidas	New York	In-store	28
29	2013-01-01	4	1	1	60	18	1	Women's Athletic Foo...	adidas	New York	In-store	29
30	2013-01-01	4	1	1	60	18	1	Women's Athletic Foo...	adidas	New York	In-store	30

Query executed successfully. | KENAN (16.0 RTM) | KENAN\MARIO (58) | project3 | 00:00:00 | 1,000 rows

Figure 3. 5. Running Query to Import Data.

III. **Data Import Wizard:** The third picture shows a step in the SQL Server Import and Export Wizard, where the user specifies the input file, pointing to a CSV file on the desktop, which will be used to create a new table in the database.

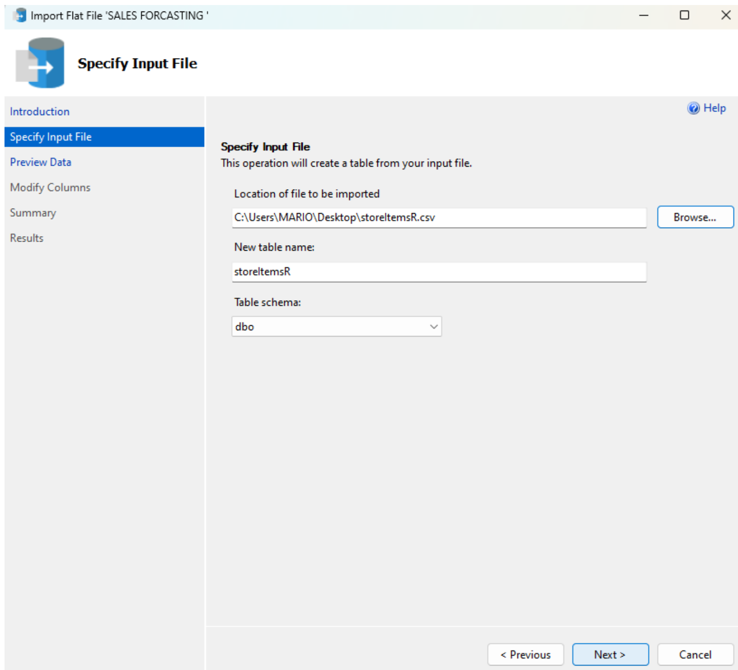


Figure 3. 6. Location of The Dataset for Import.

IV. Data Preview in Import Wizard: This image provides a preview of the data as seen in the SQL Server Import and Export Wizard, highlighting how the data will appear once imported into SQL Server.

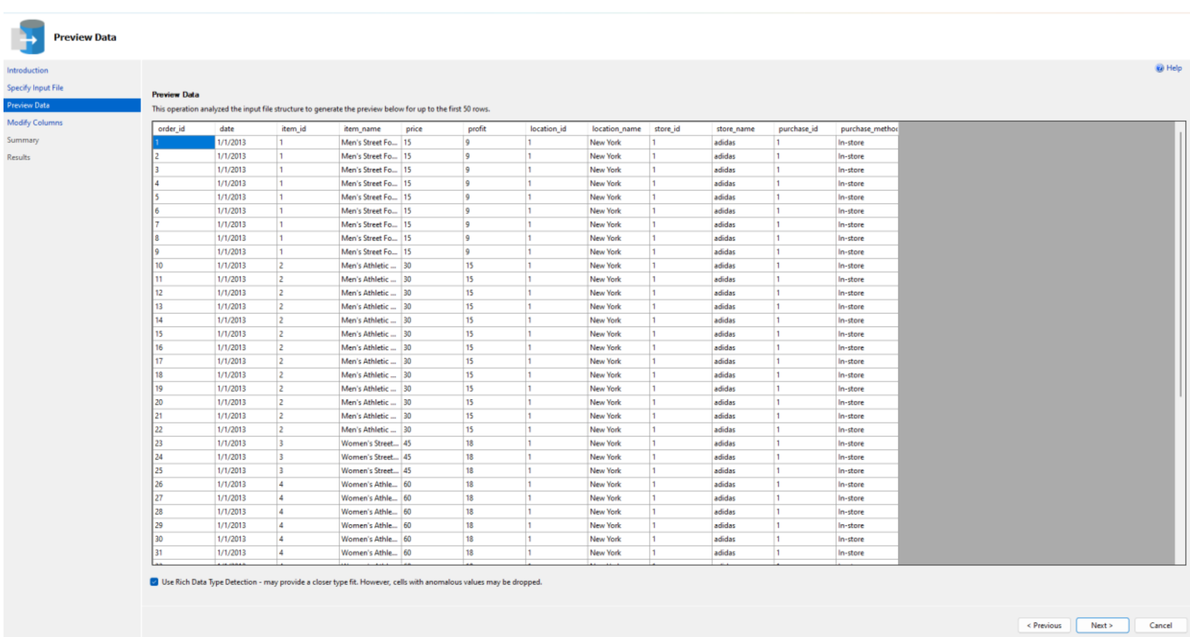
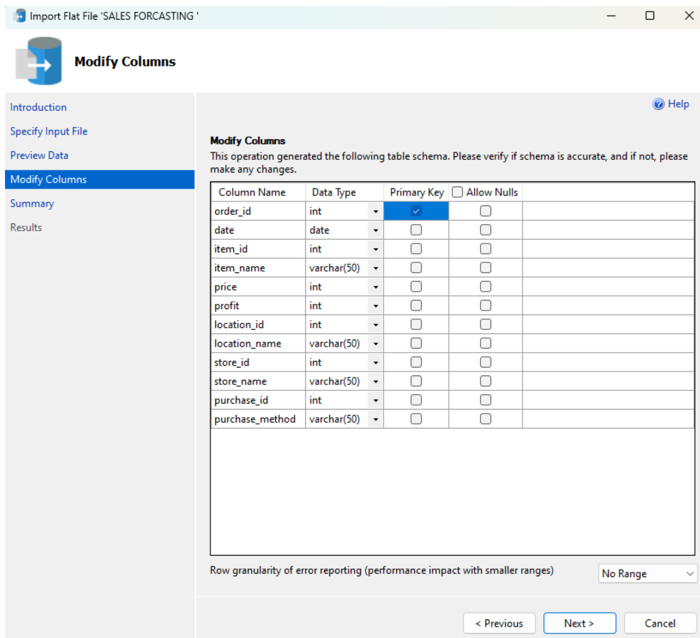


Figure 3. 7. Data in SQL Server.

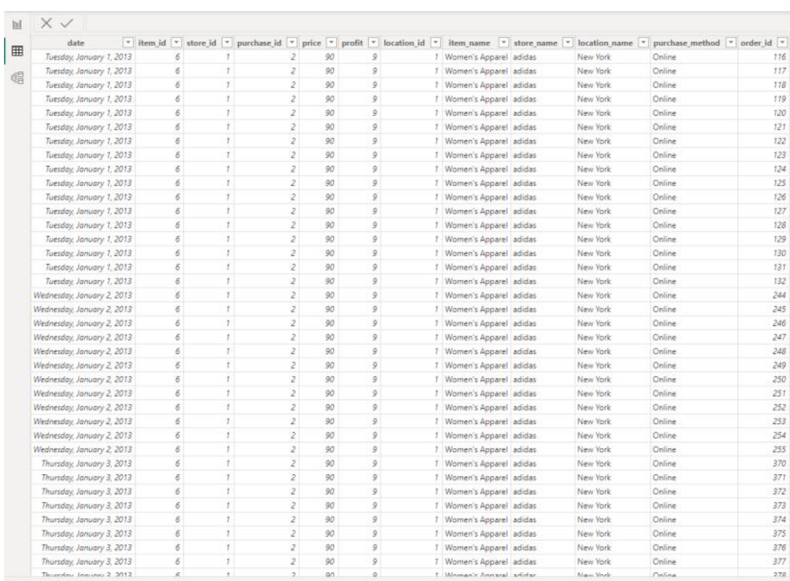
V. **Table Schema Confirmation:** In this screenshot, we're looking at the Modify Columns step of the SQL Server Import and Export Wizard, where the user can confirm or adjust the data types and properties of the columns in the newly created table.



Column Name	Data Type	Primary Key	Allow Nulls
order_id	int	✓	☐
date	date	☐	☐
item_id	int	☐	☐
item_name	varchar(50)	☐	☐
price	int	☐	☐
profit	int	☐	☐
location_id	int	☐	☐
location_name	varchar(50)	☐	☐
store_id	int	☐	☐
store_name	varchar(50)	☐	☐
purchase_id	int	☐	☐
purchase_method	varchar(50)	☐	☐

Figure 3. 8. Schema Definition and Modification.

VI. **Data Integration in SQL Server:** This image presents a table with a comprehensive view of the uploaded dataset in SQL Server, showcasing the data ready for analysis.



date	item_id	store_id	purchase_id	price	profit	location_id	item_name	store_name	location_name	purchase_method	order_id
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	116
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	117
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	118
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	119
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	120
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	121
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	122
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	123
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	124
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	125
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	126
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	127
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	128
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	129
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	130
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	131
Tuesday, January 1, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	132
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	244
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	245
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	246
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	247
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	248
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	249
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	250
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	251
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	252
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	253
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	254
Wednesday, January 2, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	255
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	370
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	371
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	372
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	373
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	374
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	375
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	376
Thursday, January 3, 2013	6	1	2	90	9	1	Women's Apparel	adidas	New York	Online	377

Table: storeItemsR (510,608 rows)

Figure 3. 9. Data Successfully Imported.

VII. **Database Relationships Diagram:** Here we have a visual representation of the database schema in SQL Server, showing the relationships between different tables such as 'store', 'purchases', 'items', and others.

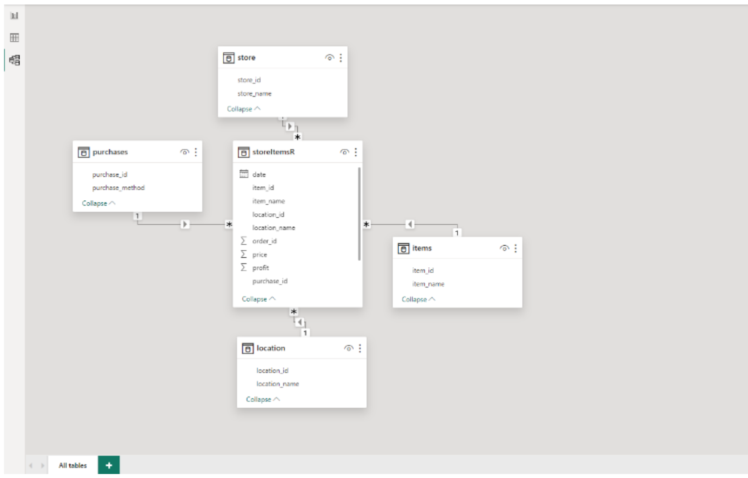


Figure 3. 10. Schema Diagram (Star Flake).

VIII. **Power BI Visualization Dashboard:** The final picture illustrates a dashboard in Power BI, where the data has been transformed into a series of visualizations, including pie charts and line graphs, for sales forecasting purposes.

Figure 3. 11. Data Visualization by Power Pi.

Model building:

Building a sales forecasting model typically involves these steps:

- **Data Collection:** Gather relevant historical sales data and potentially influential factors (like market trends, and customer demographics).
- **Data Preprocessing:** Clean the data to handle missing values, outliers, and format inconsistencies.
- **Feature Selection:** Choose relevant variables (features) that influence sales outcomes.
- **Model Selection:** Decide on a forecasting model (like ARIMA, or machine learning algorithms) based on the data's nature and business objectives.
- **Model Training:** Use historical data to train the model, allowing it to learn patterns.
- **Model Validation and Testing:** Test the model on a separate data set to evaluate its accuracy and reliability.
- **Iterative Improvement:** Refine the model based on test results and feedback, adjusting parameters as necessary.
- **Deployment:** Implement the model in a real-world business environment for actual forecasting.
- **Monitoring and Maintenance:** Continuously monitor the model's performance and update it as necessary to adapt to new data and changing market conditions.

To develop the Linear Regression model, ARIMA model, and Holt Winter's triple exponential smoothing model for sales forecasting, I followed the outlined model-building steps. First, I collected and preprocessed relevant historical sales data, addressing missing values and outliers. Next, I identified key features impacting sales, such as rolling means and lag variables, through feature selection techniques. The model was then selected and trained using this processed data. Subsequent steps involved testing and validating the model's predictions against a separate dataset, followed by iterative improvements based on error analysis. Finally, the model was deployed and continuously monitored for performance, ensuring its adaptability and accuracy over time.

I. The steps in the provided **Linear Regression** model for sales forecasting are:

- **Data Collection:** Gathering relevant sales data along with influential factors.
- **Data Preprocessing:** Cleaning the data, and handling issues like missing values and outliers.
- **Feature Selection:** Identifying significant variables (e.g., rolling means, lag features) that influence sales.
- **Model Selection and Training:** Choosing the Linear Regression model and training it with the selected features.

- **Model Validation and Testing:** Testing the model's predictions against a separate dataset to evaluate accuracy.
- **Iterative Improvement:** Refining the model based on error analysis and feedback.
- **Deployment:** Implementing the model for actual sales forecasting.
- **Monitoring and Maintenance:** Continuously tracking the model's performance and making necessary updates.

Each step is designed to ensure the model is accurate, reliable, and adaptable to changing sales trends.

II. This model involves several steps for sales forecasting using **ARIMA** and **SARIMAX** models:

- **Data Preparation:** The script begins by importing necessary libraries and loading the dataset. It filters data for a specific store and formats the 'date' column.
- **Creating Date Features:** Additional features like 'year', 'month', 'day', and 'weekday' are created for more detailed analysis.
- **Data Splitting:** The dataset is divided into training and testing sets based on dates.
- **Stationarity Testing:** The Augmented Dickey-Fuller test checks if the sales data is stationary, crucial for time series modeling.
- **ARIMA Modeling:** An ARIMA model is fitted to the training data. This model is selected based on its ability to handle non-seasonal data.
- **SARIMAX Modeling:** A SARIMAX model, which accounts for seasonality, is also fitted.
- **Forecasting:** The SARIMAX model is used to forecast sales for the test period.
- **Error Calculation:** Errors between the forecasted and actual sales are calculated using metrics like MAE, RMSE, and MAPE.
- **Visualization:** The sales data, along with forecasts and errors, are plotted for visual analysis.
- **Result Aggregation:** Finally, the results are summarized, providing a comprehensive view of the model's performance.

These steps collectively create a robust forecasting model, considering both seasonal and non-seasonal patterns in sales data.

III. This model applies the **Holt-Winters Exponential Smoothing** method for sales forecasting:

- **Data Preparation:** Imports necessary libraries, loads the dataset, filters for a specific store, and formats the 'date' column.
- **Data Splitting:** Divides the dataset into training and testing sets.
- **Seasonal Decomposition:** Uses seasonal decomposition to understand underlying trends and seasonality in the sales data.

- **Holt-Winters Model Implementation:** Two versions of the Holt-Winters model are applied — one without damping and one with a damping trend component.
- **Forecasting:** Both models are used to forecast sales for the test period.
- **Error Calculation:** Calculates forecast errors using methods like MAE, RMSE, and MAPE.
- **Visualization:** Plots the sales, forecasts, and errors for visual comparison.
- **Result Aggregation:** Summarizes the results in tables, comparing the performance of the two models.

Each step contributes to a nuanced understanding of the model's performance and the impact of incorporating a damping trend in forecasting.

Model results:

I. Linear regression model results:

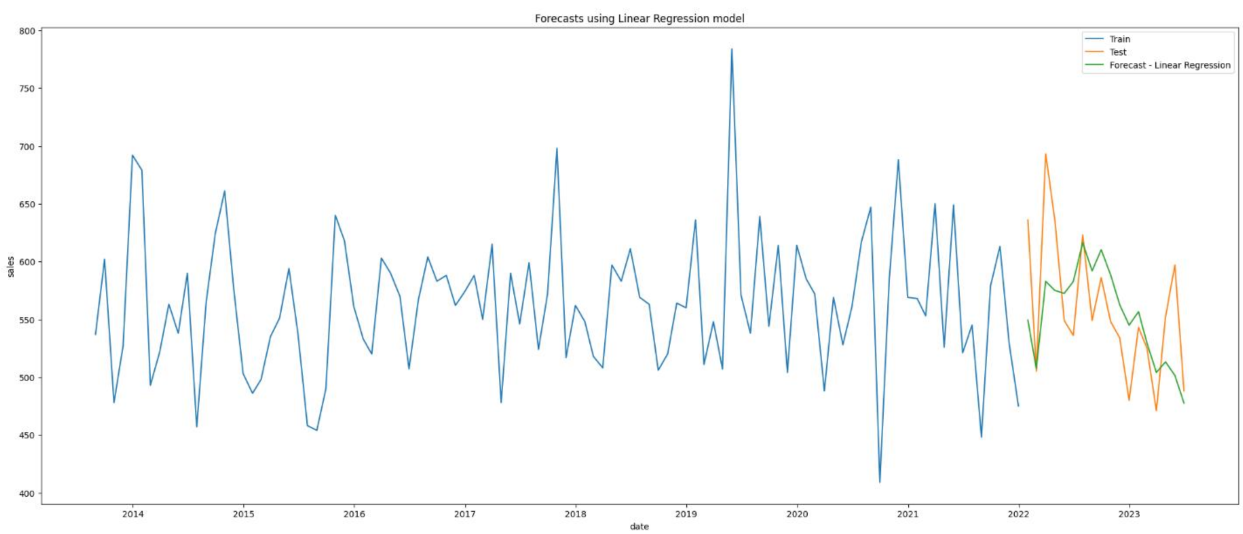


Figure 3. 12. LR-ELM Forecast for Item One.

model	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
LinearRegression	10051	9966.587652	-84.412348	40.780816	51.25074	7.07796

Figure 3. 13. LR-ELM Sales & Metrics for Item One.

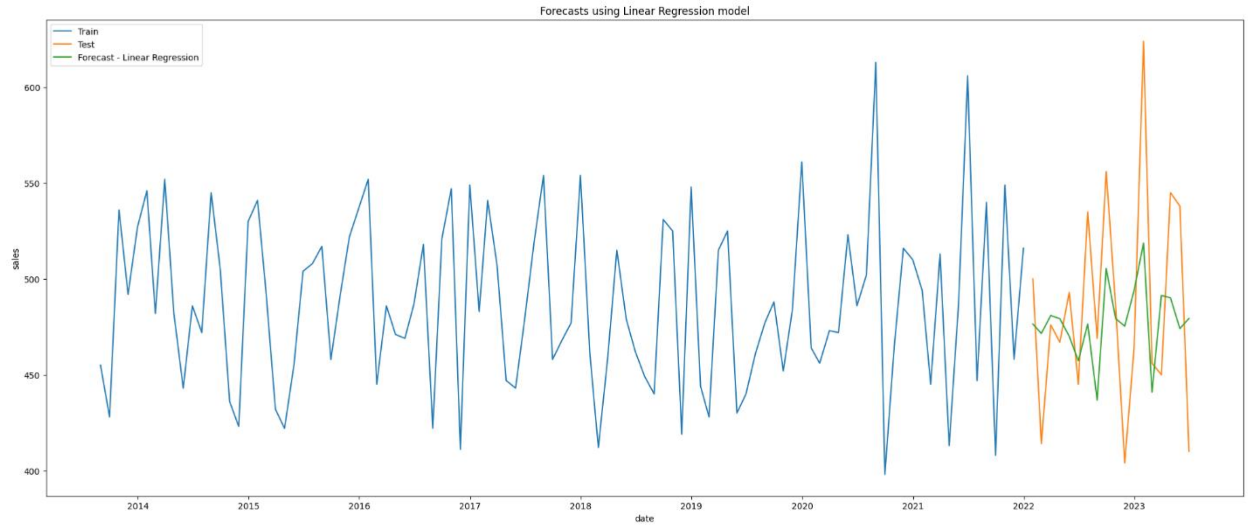


Figure 3. 14. LR-ELM Forecast for Item Two.

	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
model						
LinearRegression	8732	8597.405159	-134.594841	40.708307	48.716055	8.351638

Figure 3. 15. LR-ELM Sales & Metrics for Item Two.

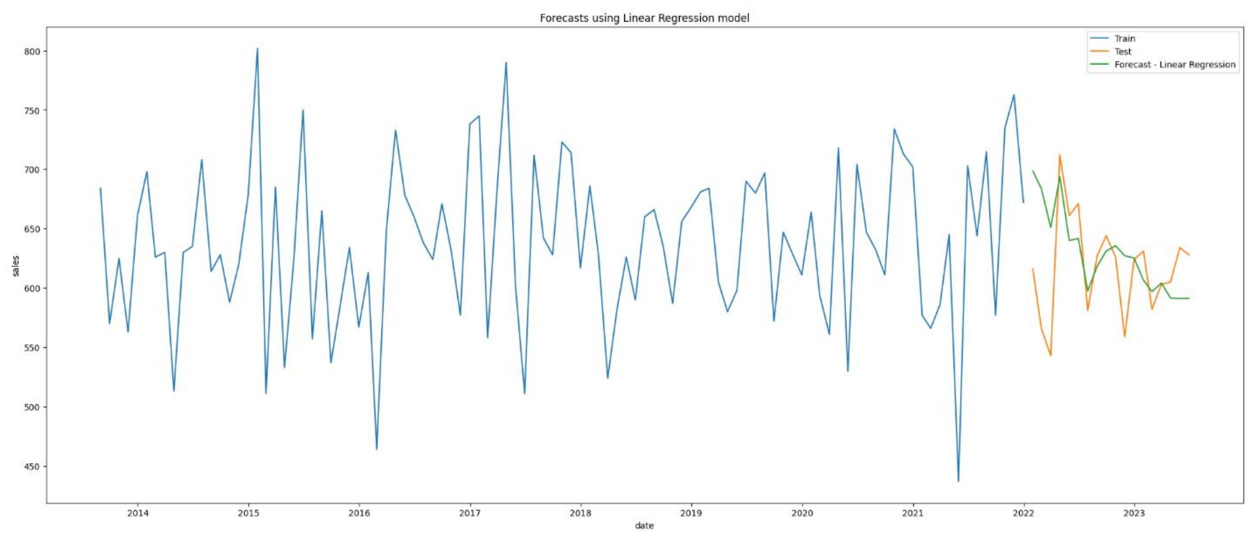


Figure 3. 16. LR-ELM Forecast for Item Three.

model	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
LinearRegression	11113	11325.175196	212.175196	34.860741	49.133694	5.878942

Figure 3. 17. LR-ELM Sales & Metrics for Item Three.

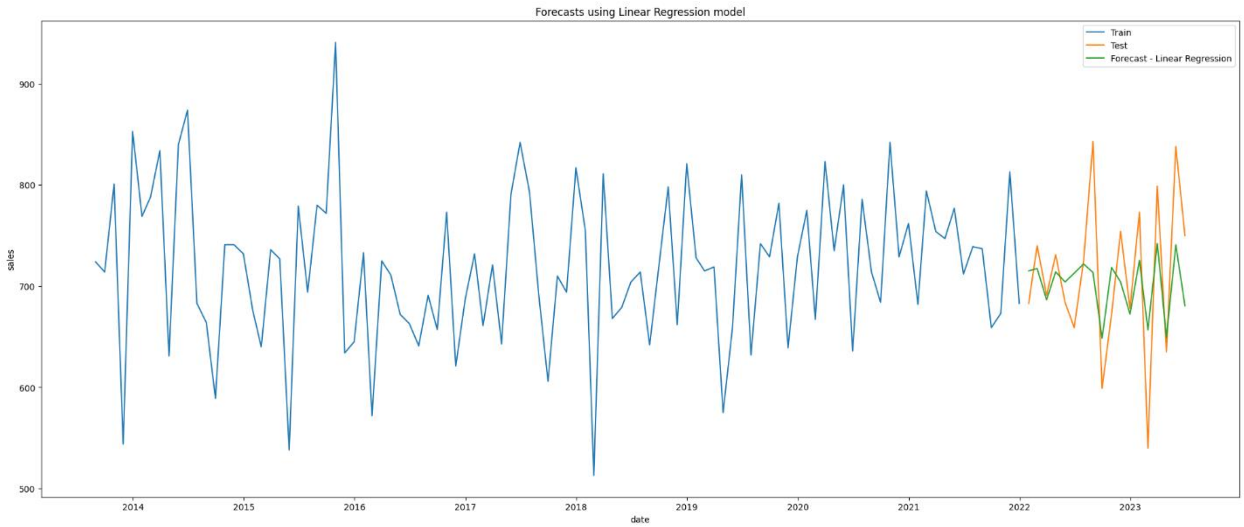


Figure 3. 18. LR-ELM Forecast for Item Four.

model	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
LinearRegression	12794	12624.00848	-169.99152	46.429462	58.985473	6.564892

Figure 3. 19. LR-ELM Sales & Metrics for Item Four.

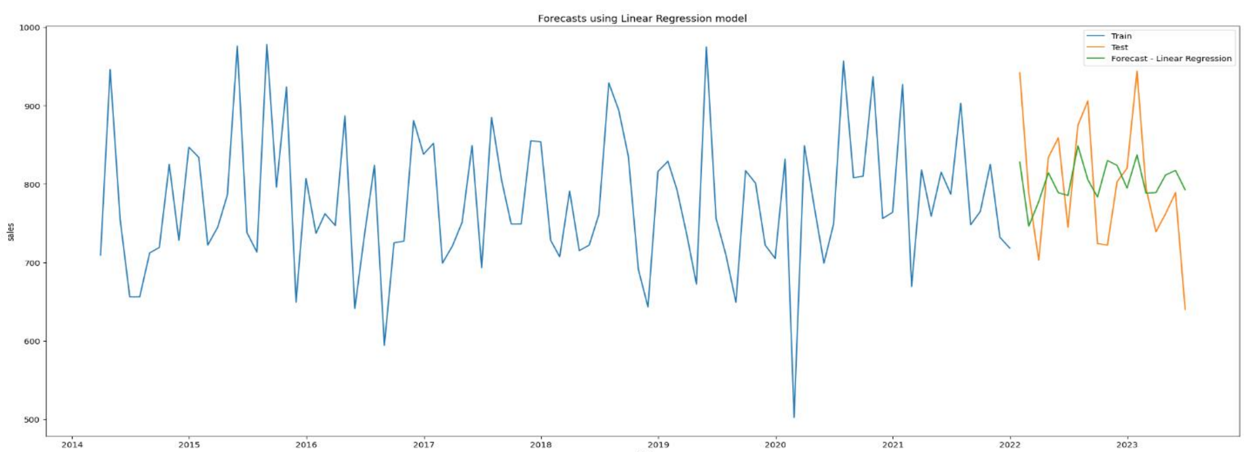


Figure 3. 20. LR-ELM Forecast for Item Five.

	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
model						
LinearRegression	14390	14462.261777	72.261777	60.830506	72.581122	7.764698

Figure 3. 21. LR-ELM Sales & Metrics for Item Five.

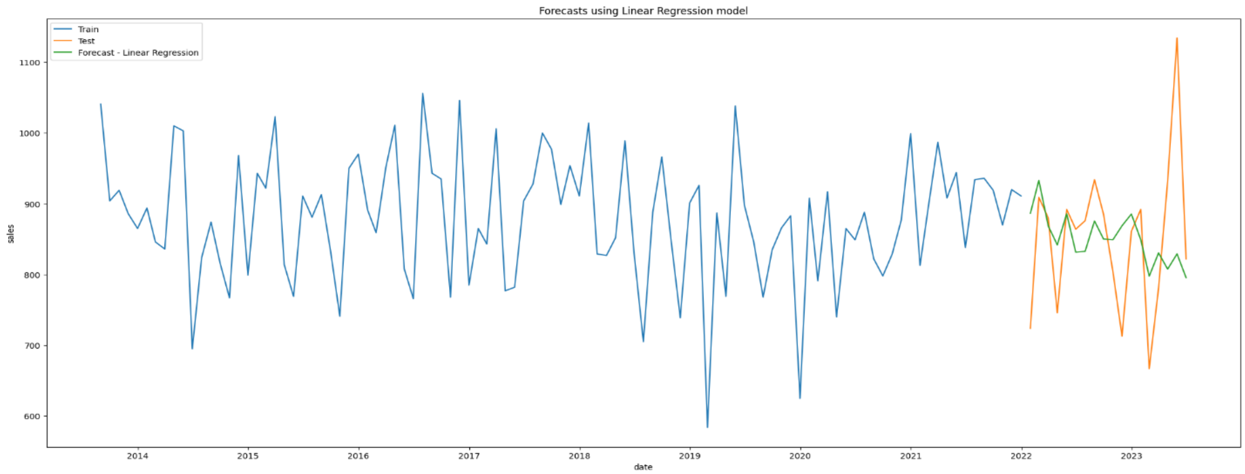


Figure 3. 22. LR-ELM Forecast for Item Six.

	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
model						
LinearRegression	15316	15318.96496	2.96496	76.26244	105.742761	9.070105

Figure 3. 23. LR-ELM Sales & Metrics for Item Six.

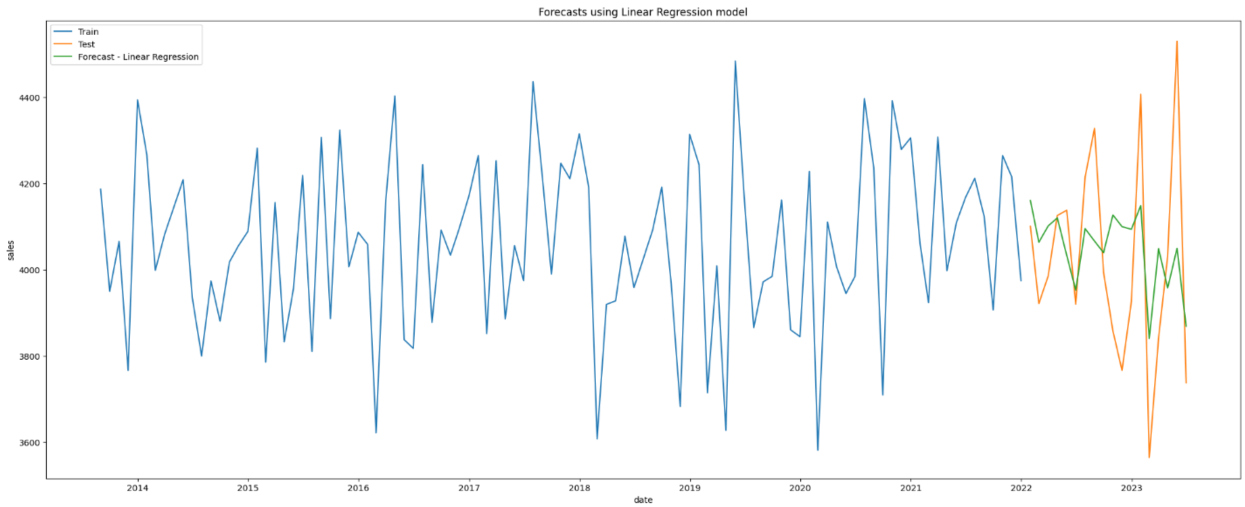


Figure 3. 24. LR-ELM Forecast for Store Items.

	total_sales	total_pred_sales	LR_overall_error	MAE	RMSE	MAPE
model						
LinearRegression	72396	72873.023032	477.023032	171.135578	208.622438	4.246859

Figure 3. 25. LR-ELM Sales & Metrics for Store Items.

II. ARIMA model results:

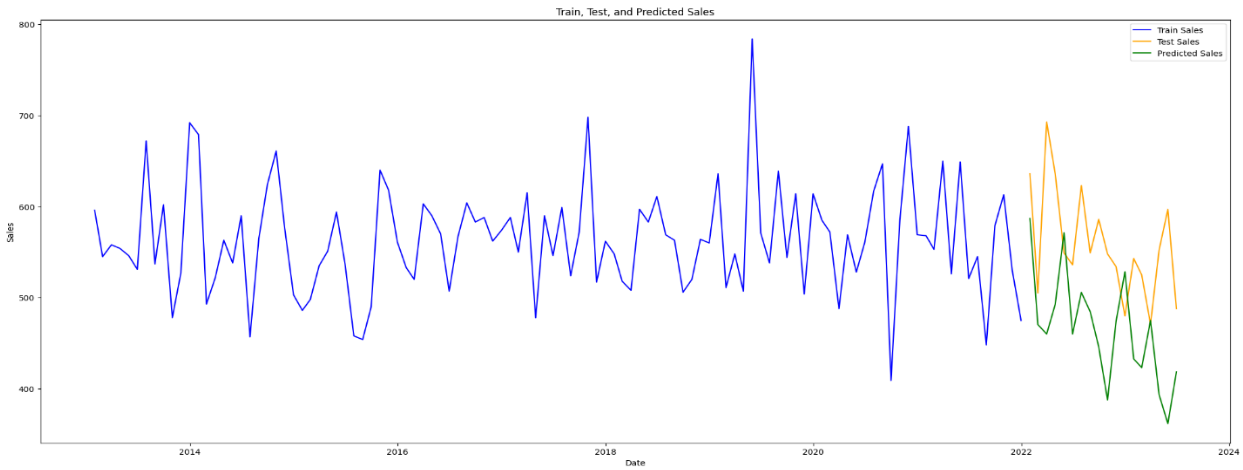


Figure 3. 26. A-ELM Forecast for Item One.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	10051	8372.187641	1678.812359	101.538058	120.586068	17.601034

Figure 3. 27. A-ELM Sales & Metrics for Item One.

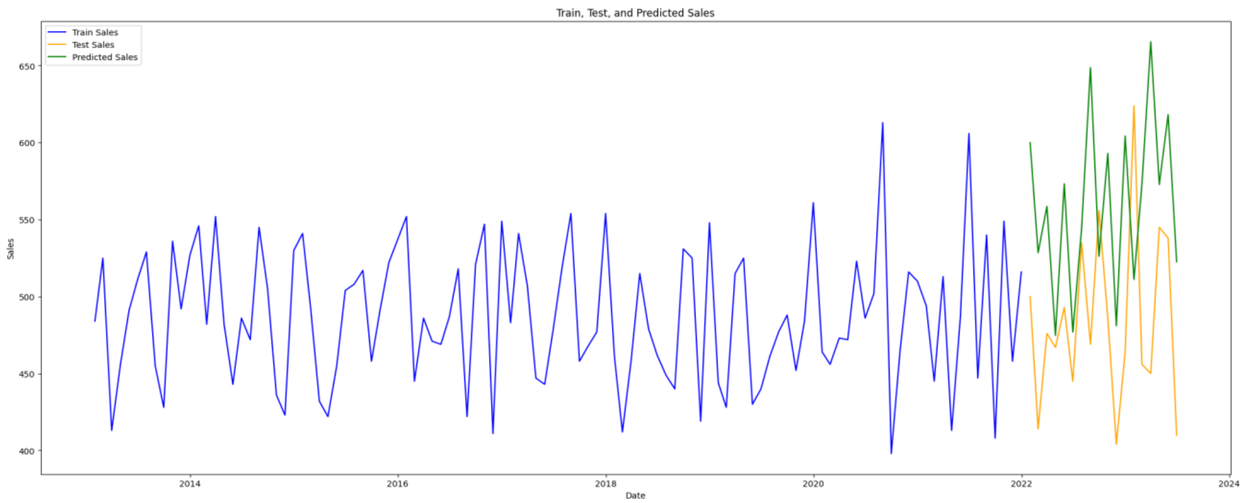


Figure 3. 28. A-ELM Forecast for Item Two.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	8732	10073.034183	-1341.034183	90.358858	105.572899	19.224146

Figure 3. 29. A-ELM Sales & Metrics for Item Two.

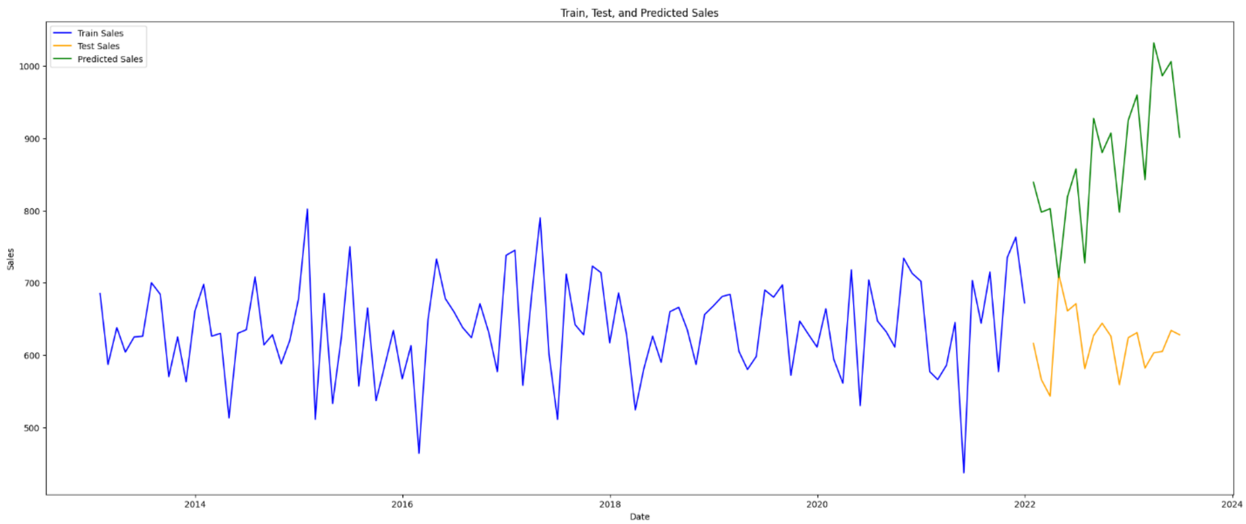


Figure 3. 30. A-ELM Forecast for Item Three.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	11113	15716.177876	-4603.177876	256.259453	273.183768	42.013125

Figure 3. 31. A-ELM Sales & Metrics for Item Three.

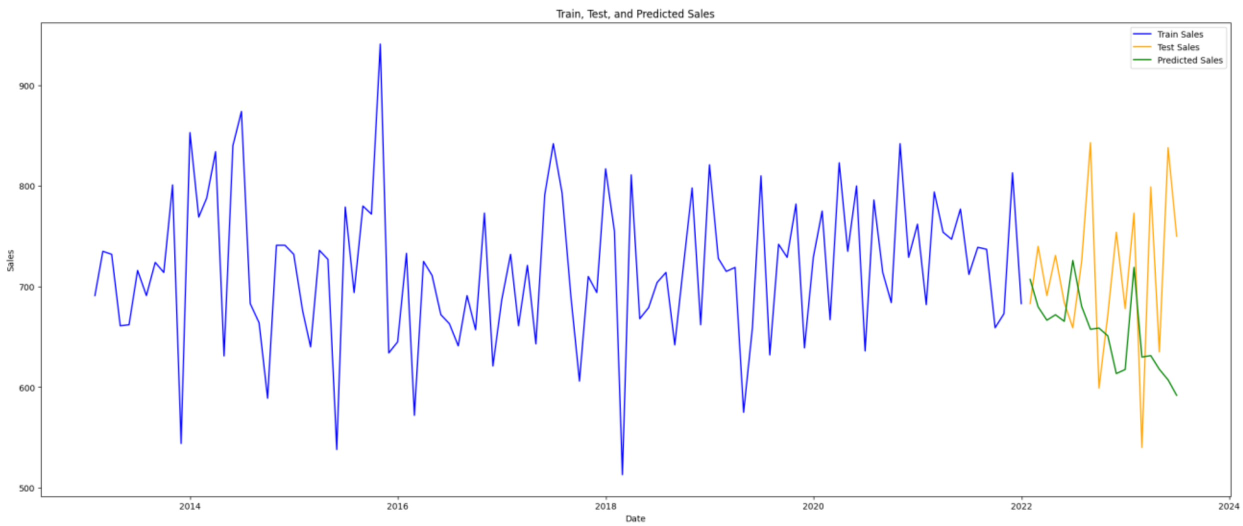


Figure 3. 32. A-ELM Forecast for Item Four.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	12794	11792.533437	1001.466563	82.377125	103.972722	11.161634

Figure 3. 33. A-ELM Sales & Metrics for Item Four.

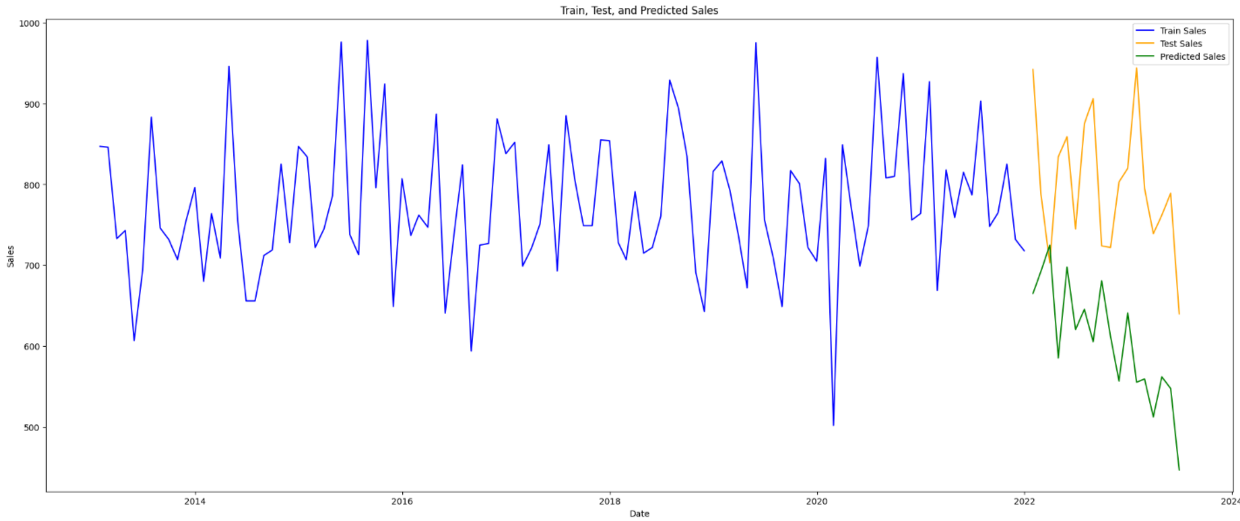


Figure 3. 34. A-ELM Forecast for Item Five.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	14390	10912.578243	3477.421757	195.621399	215.012054	23.962901

Figure 3. 35. A-ELM Sales & Metrics for Item Five.

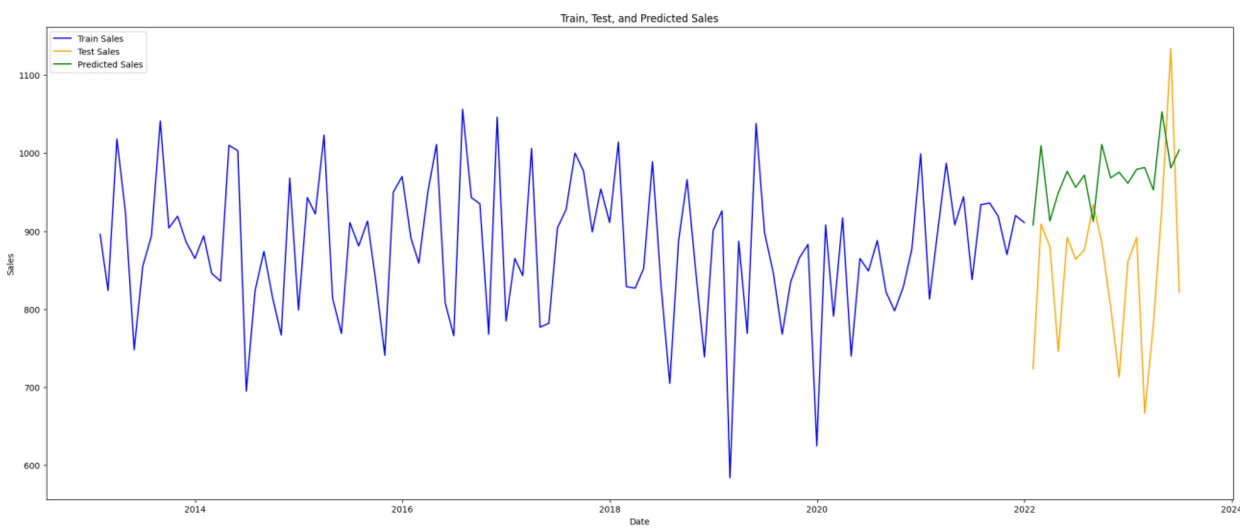


Figure 3. 36. A-ELM Forecast for Item Six.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	15316	17463.492111	-2147.492111	138.701079	156.423073	17.301701

Figure 3. 37. A-ELM Sales & Metrics for Item Six.

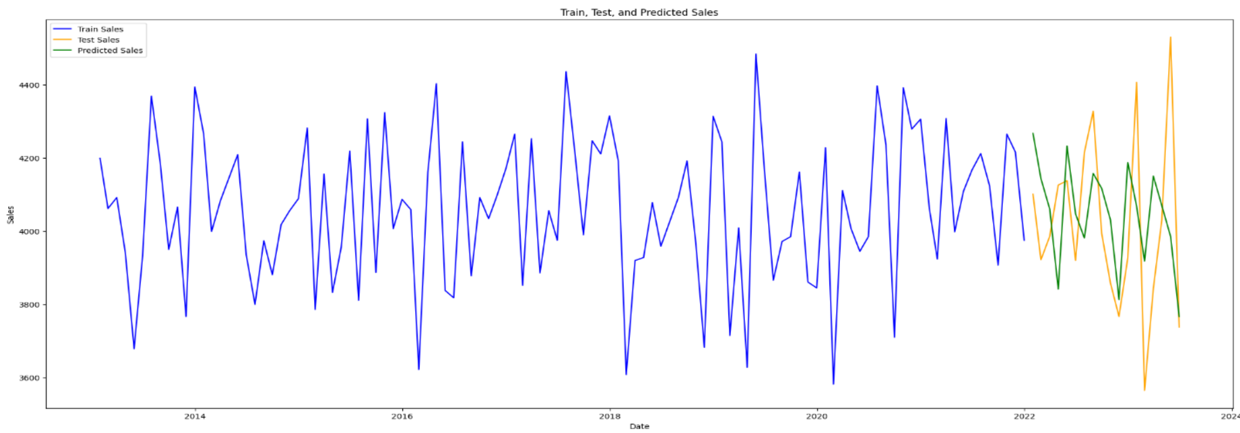


Figure 3. 38. A-ELM Forecast for Store Items.

	total_sales	total_pred_sales	SARIMA_overall_error	MAE	RMSE	MAPE
0	72396	72839.622008	-443.622008	199.27612	238.081304	4.90292

Figure 3. 39. A-ELM Sales & Metrics for Store Items.

III. Holt-Winters Exponential Smoothing model results:

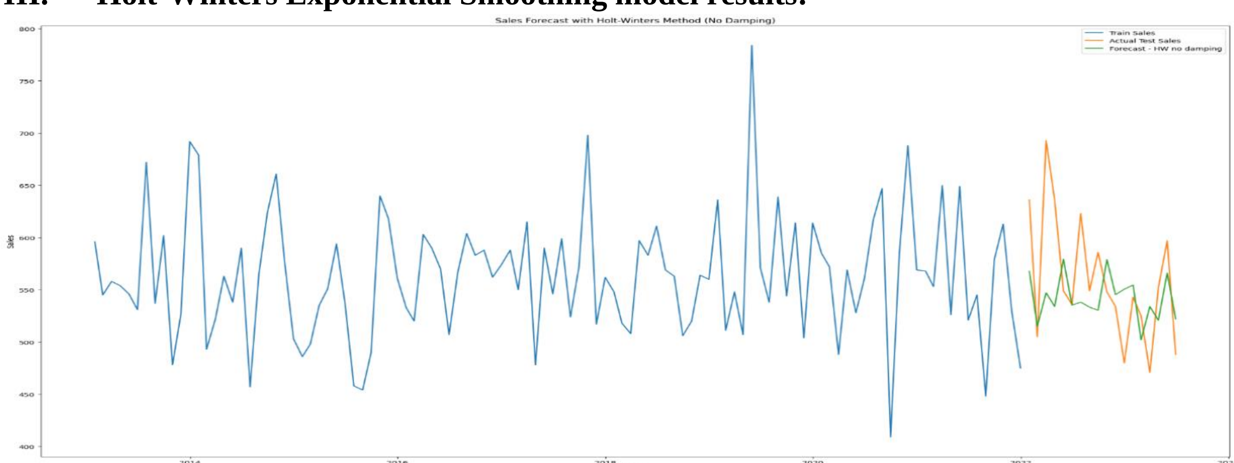


Figure 3. 40. ES-ELM Forecast for Item One (without Dumping).

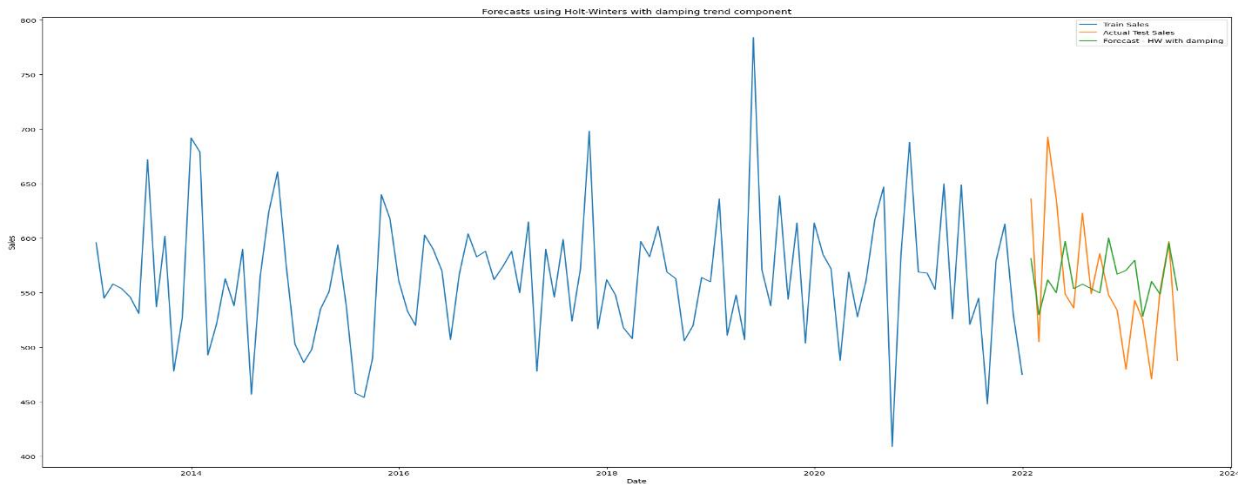


Figure 3. 41. ES-ELM Forecast for Item One (with Dumping).

```
Results for Holt-Winters Model Without Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	10051	9755.225293	295.774707	45.555732	58.535455	7.867703

```
Results for Holt-Winters Model With Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	10051	10138.905228	-87.905228	46.831468	58.659223	8.362666

Figure 3. 42. ES-ELM Sales & Metrics for Item One.

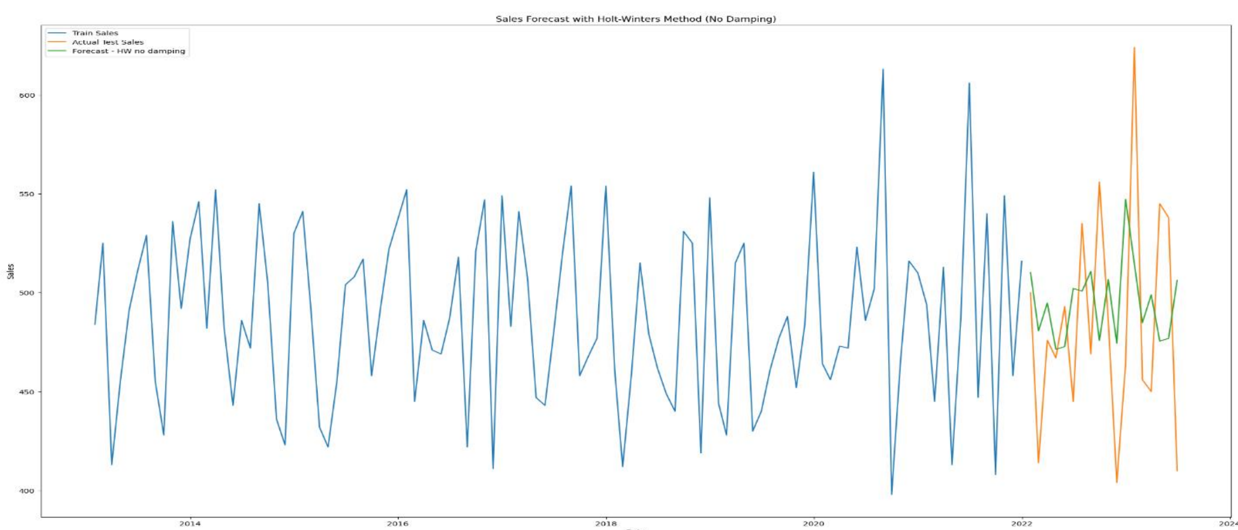


Figure 3. 43. ES-ELM Forecast for Item Two (without Dumping).

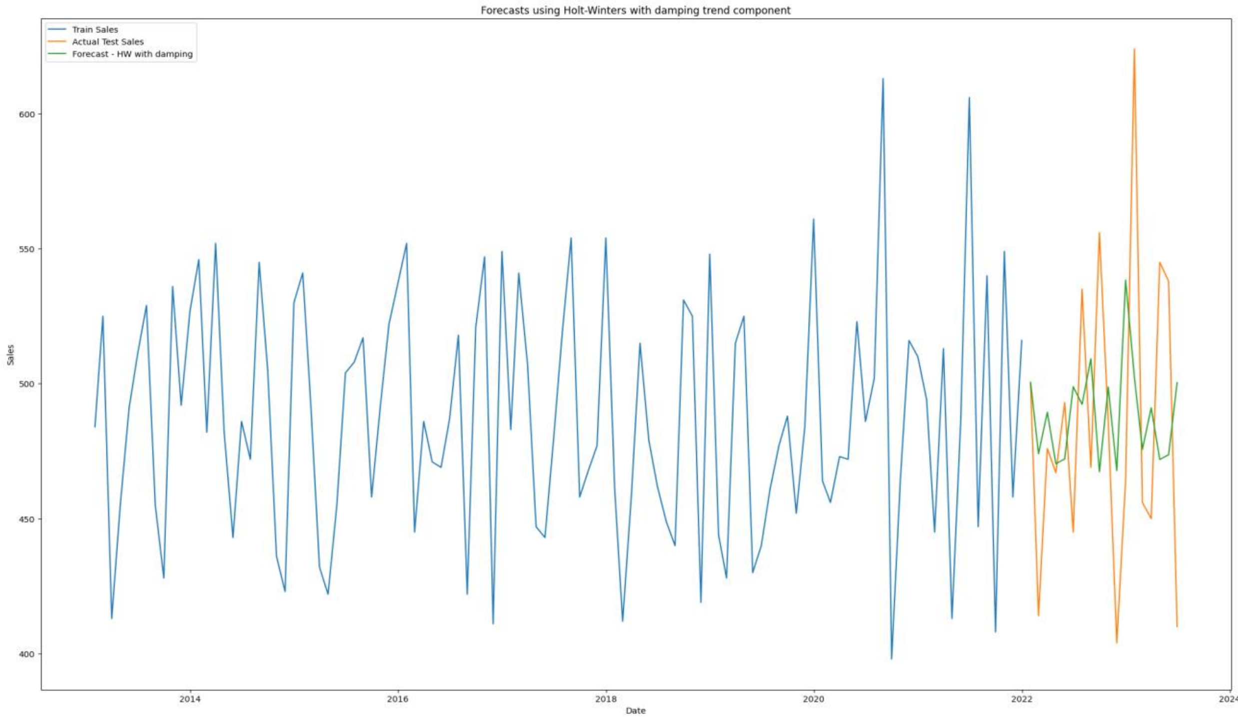


Figure 3. 44. ES-ELM Forecast for Item Two (with Dumping).

```
Results for Holt-Winters Model Without Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	8732	8904.364157	-172.364157	51.238901	59.319073	10.649887

```
Results for Holt-Winters Model With Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	8732	8794.239124	-62.239124	49.15619	59.035779	10.099173

Figure 3. 45. ES-ELM Sales & Metrics for Item Two.

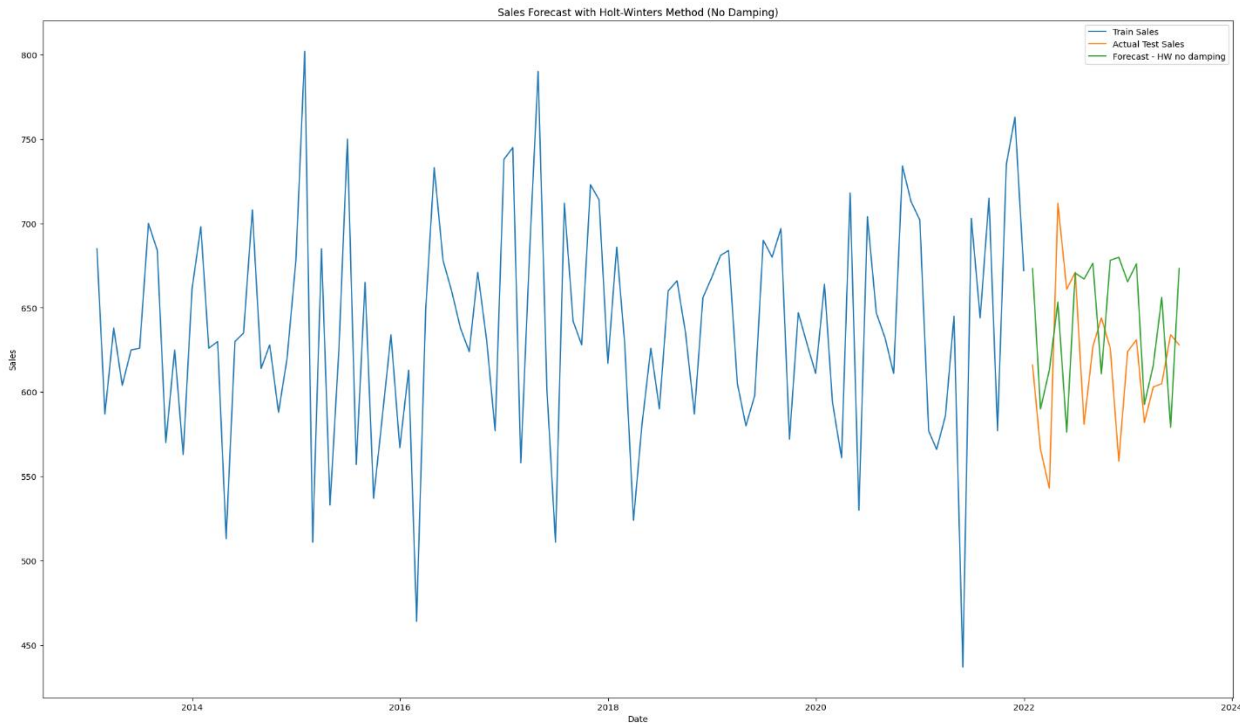


Figure 3. 46. ES-ELM Forecast for Item Three (without Dumping).

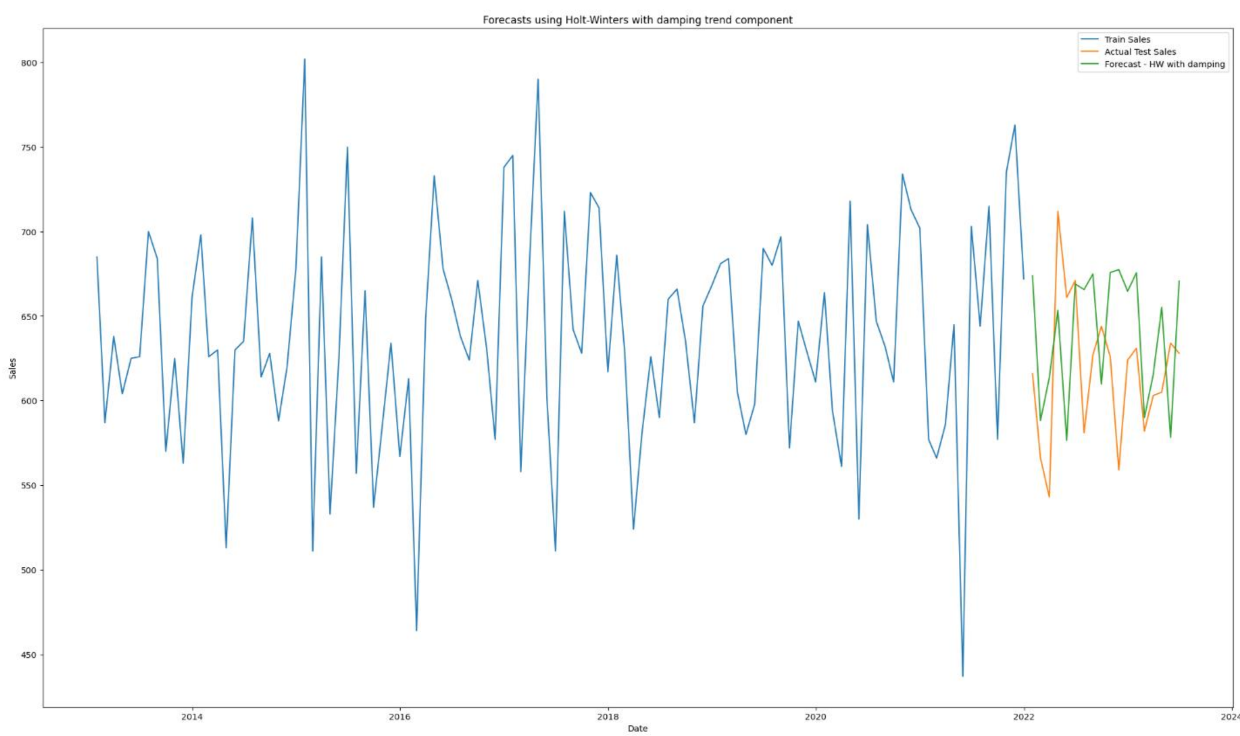


Figure 3. 47. ES-ELM Forecast for Item Three (with Dumping).

Results for Holt-Winters Model Without Damping:

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	11113	11547.258474	-434.258474	49.899657	57.435192	8.190813

Results for Holt-Winters Model With Damping:

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	11113	11528.358823	-415.358823	49.191046	56.71215	8.070248

Figure 3. 48. ES-ELM Sales & Metrics for Item Three.

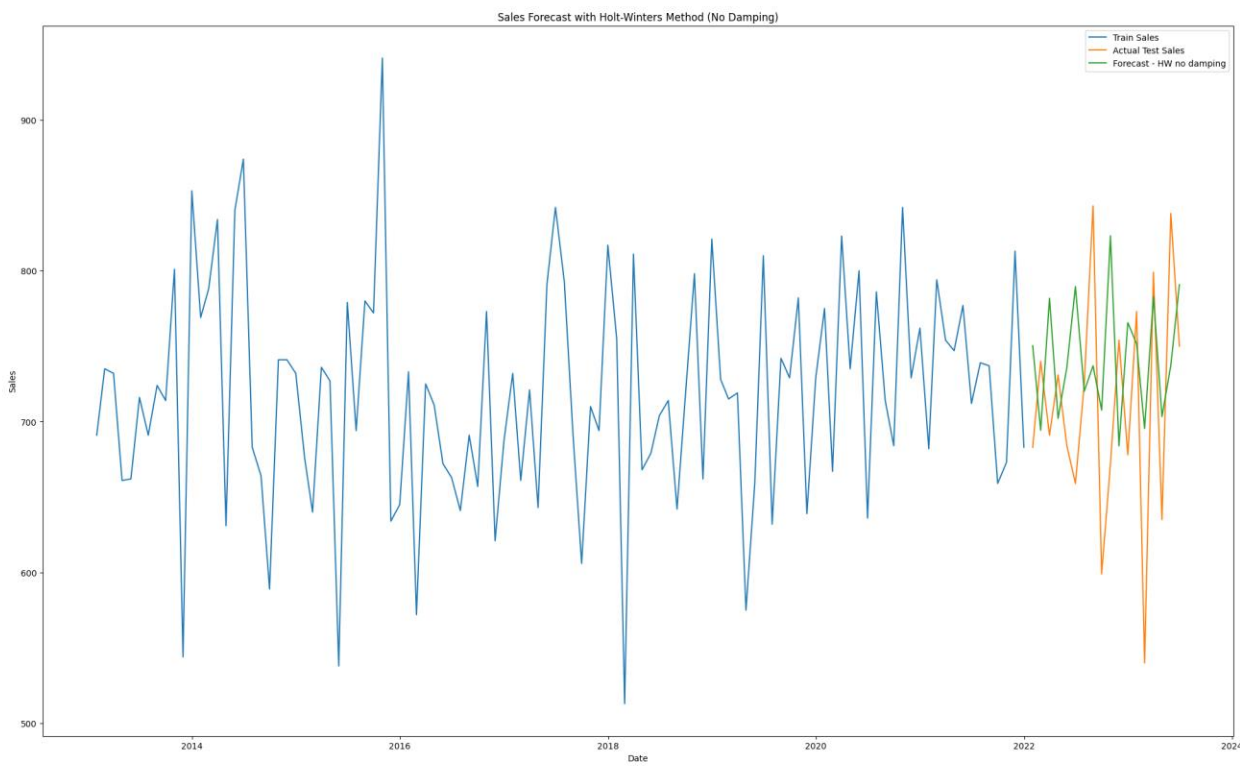


Figure 3. 49. ES-ELM Forecast for Item Four (without Dumping).

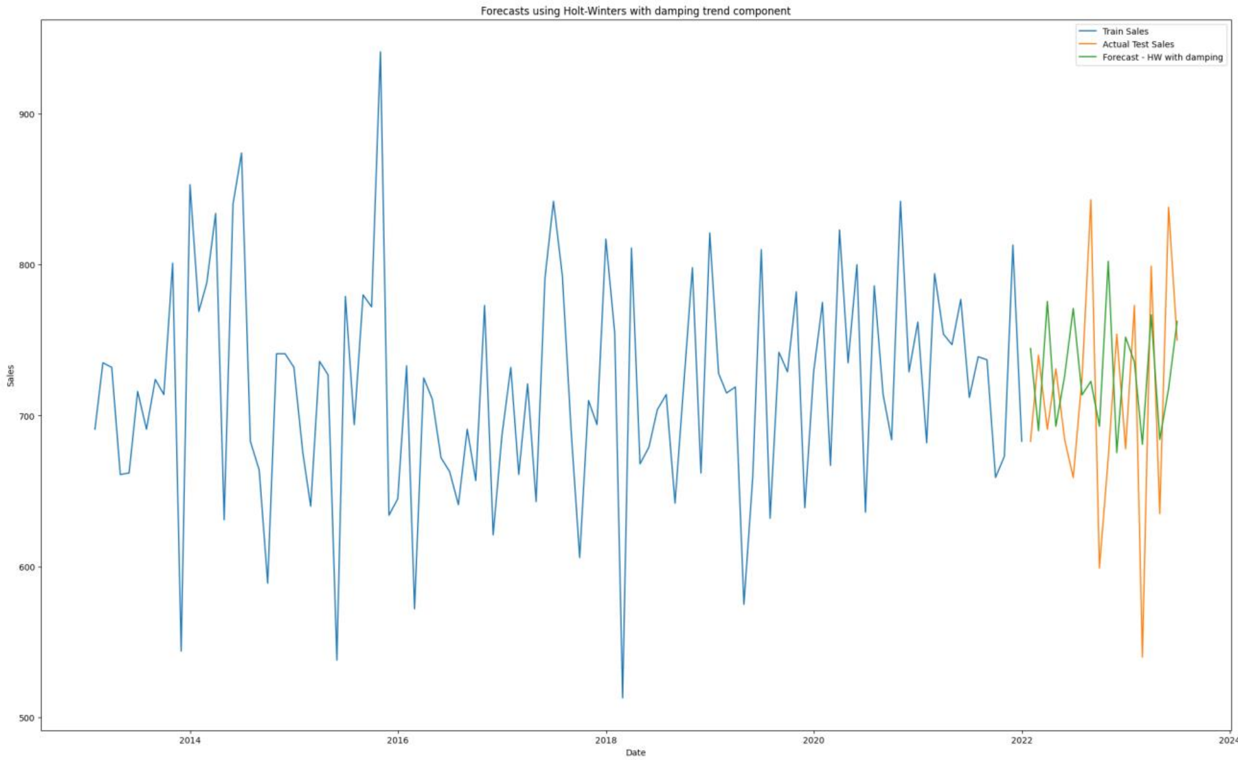


Figure 3. 50. ES-ELM Sales & Metrics for Item Four (with Dumping).

Results for Holt-Winters Model Without Damping:

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	12794	13352.613518	-558.613518	74.804405	86.703808	11.023745

Results for Holt-Winters Model With Damping:

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	12794	13107.394161	-313.394161	71.640678	81.855364	10.410467

Figure 3. 51. ES-ELM Sales & Metrics for Item Four.

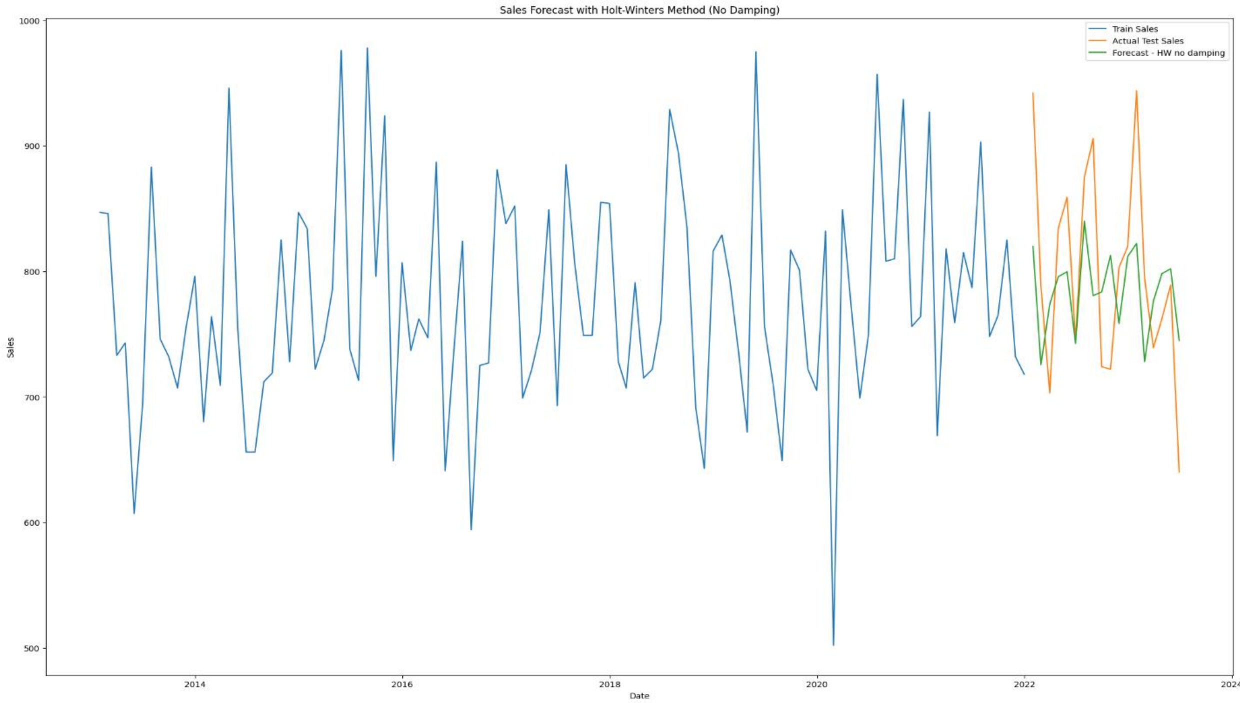


Figure 3. 52. ES-ELM Forecast for Item Five (without Dumping).

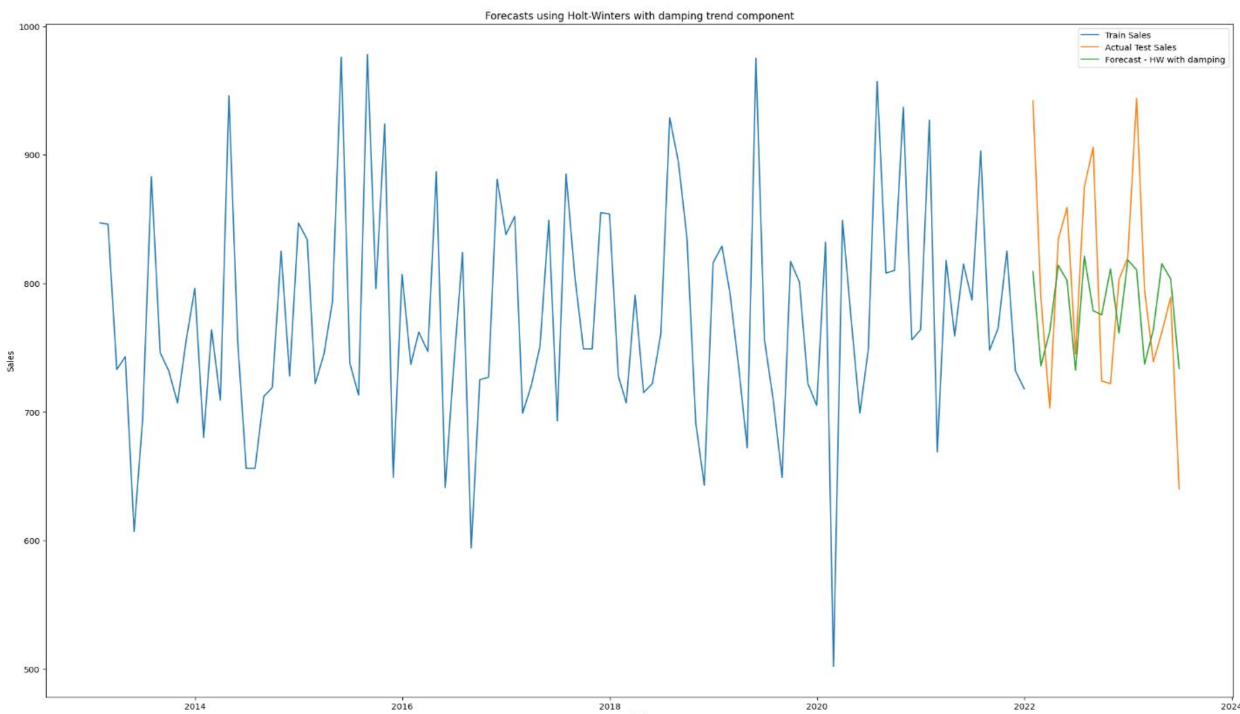


Figure 3. 53. ES-ELM Forecast for Item Five (with Dumping).

```
Results for Holt-Winters Model Without Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	14390	14115.423264	274.576736	61.014513	71.763426	7.612625

```
Results for Holt-Winters Model With Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	14390	14086.004132	303.995868	59.848144	71.794554	7.40381

Figure 3. 54. ES-ELM Sales & Metrics for Item Five.

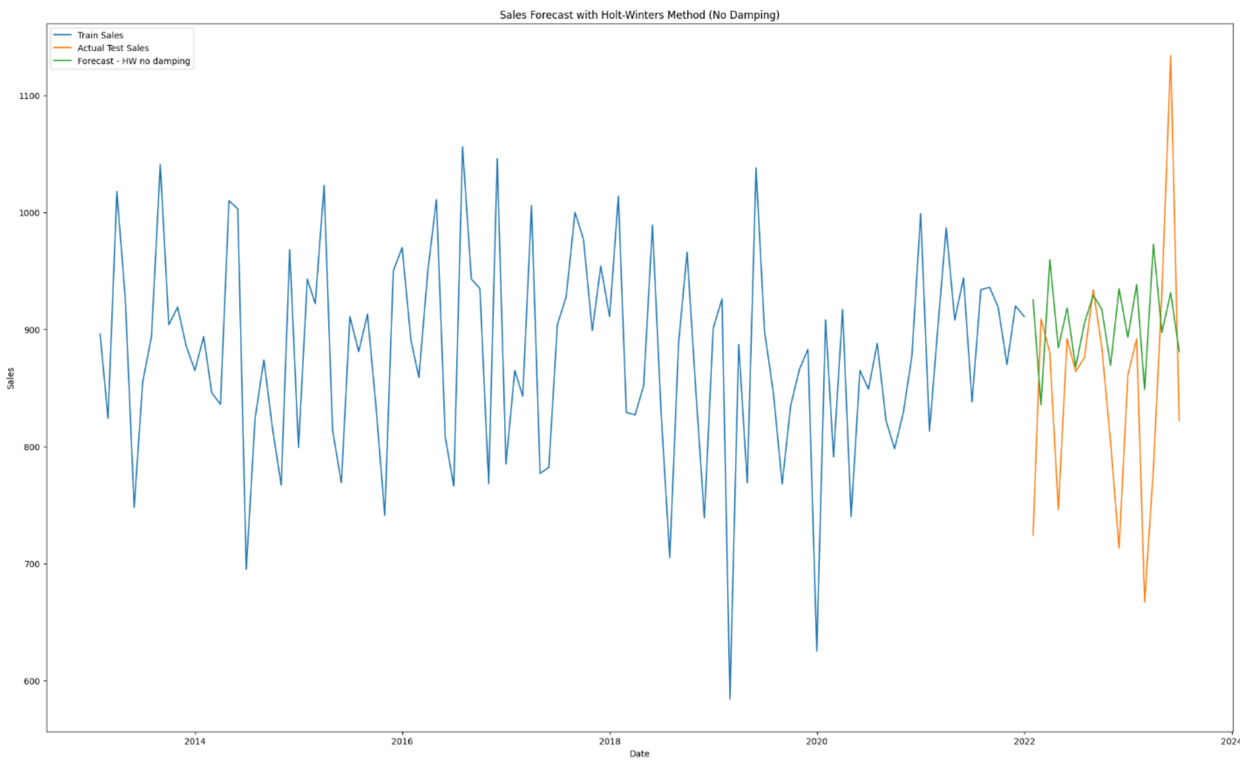


Figure 3. 55. ES-ELM Forecast for Item Six (without Dumping).

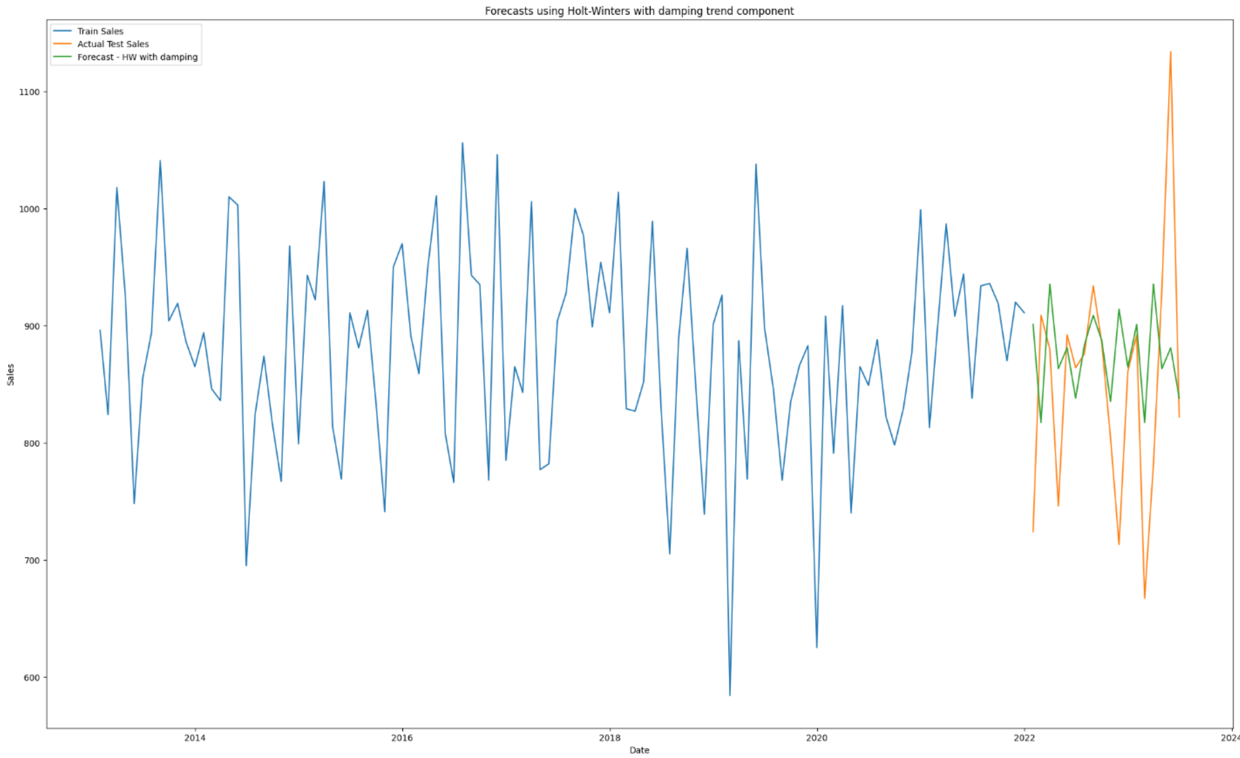


Figure 3. 56. ES-ELM Forecast for Item Six (with Dumping).

```
Results for Holt-Winters Model Without Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	15316	16309.312141	-993.312141	90.251188	116.939997	11.282414

```
Results for Holt-Winters Model With Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	15316	15765.820887	-449.820887	77.838703	109.297916	9.561075

Figure 3. 57. ES-ELM Sales & Metrics for Item Six.

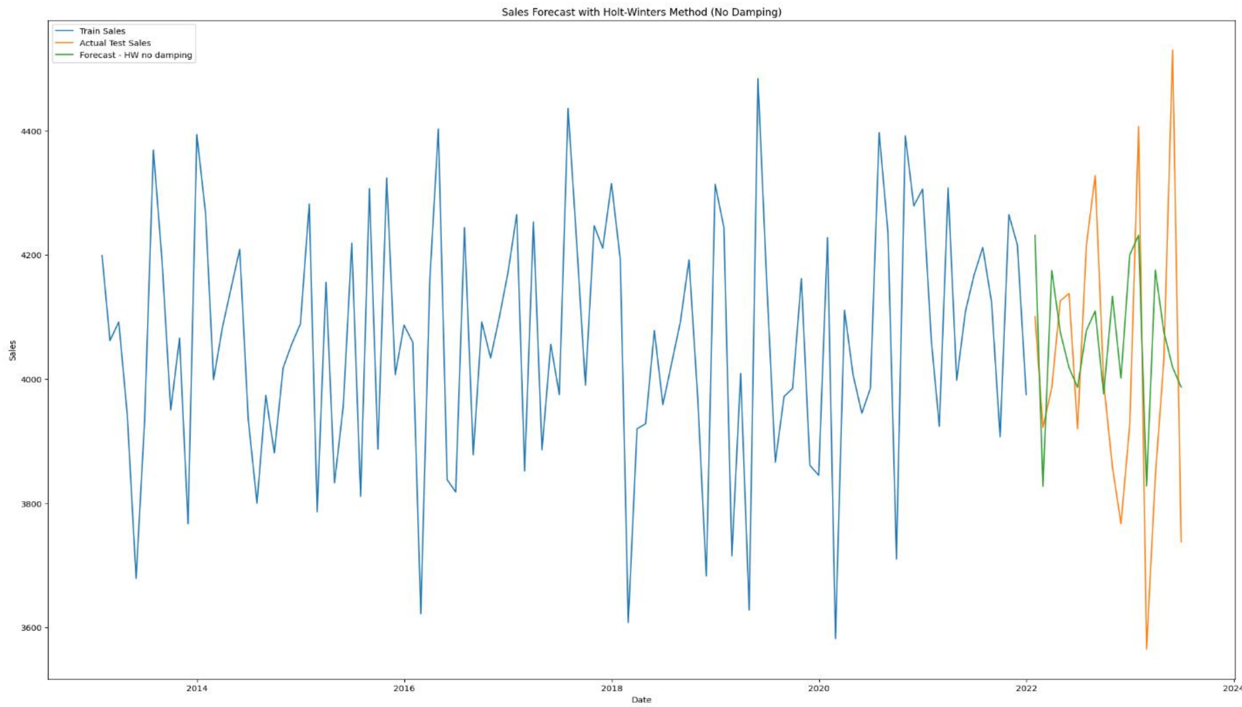


Figure 3. 58. ES-ELM Forecast for Store Items (without Dumping).

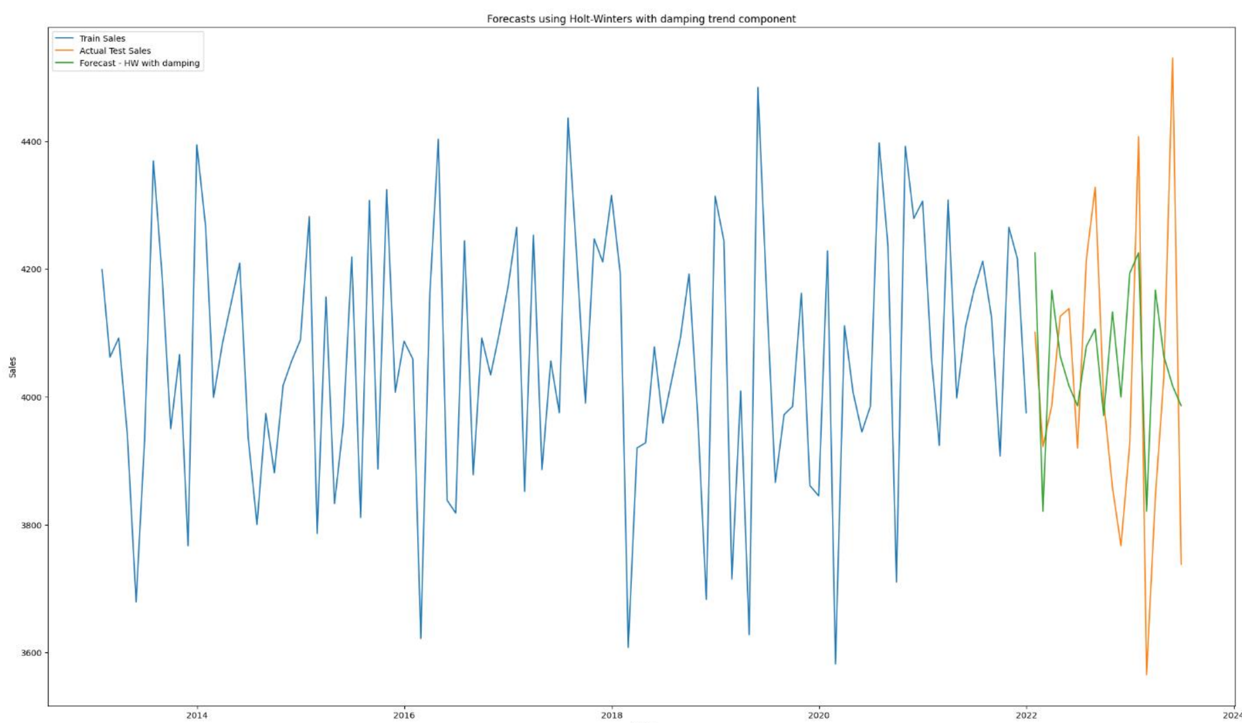


Figure 3. 59. ES-ELM Forecast for Store Items (with Dumping).

```
Results for Holt-Winters Model Without Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	72396	73130.769822	-734.769822	187.913858	222.455342	4.681375

```
Results for Holt-Winters Model With Damping:
```

	total_sales	total_pred_sales	holt_winters_overall_error	MAE	RMSE	MAPE
0	72396	73038.93656	-642.93656	186.980323	221.017739	4.654306

Figure 3. 60. ES-ELM Sales & Metrics for Store Items.

Summary:

Embarking on the journey of sales forecasting involves a meticulously crafted roadmap, starting with the critical step of model selection. This guide, comprising multiple strategic steps, navigates through the diverse array of forecasting models to pinpoint the one best suited for the task at hand. Dataset selection is the linchpin, where careful consideration of relevance, completeness, and alignment with business goals ensures the foundation for robust predictions. Transitioning into the realm of dataset visualization, plots become the storytellers, unraveling patterns and insights within the data, and guiding subsequent modeling decisions. The model-building phase unfolds as a systematic guide, encompassing steps from preprocessing to parameter tuning, crafting a model capable of capturing the intricacies of the business environment. The climax of this journey lies in presenting the model results, where plots vividly showcase the forecasting prowess of the chosen models, offering a tangible representation of their performance against defined objectives. This iterative process underscores the symbiotic relationship between each phase, culminating in a refined and accurate sales forecasting model.

Chapter Four

Practical Application

the Practical Section shows the journey of refining sales forecasting, two pivotal elements took center stage: Model Modification and Dataset Manipulation. In the realm of Model Modification, the focus was on a strategic evolution, involving updates, changes, and alterations to simplify, enhance efficiency, and optimize the model's overall performance. Simultaneously, Dataset Manipulation played a crucial role, where a meticulous process of splitting the dataset into individual files and manipulating sales dates at varying intervals aimed to observe the model's dynamic response. This deliberate tinkering sought to uncover insights into how alterations in the dataset structure impact the model's adaptability and, consequently, the resulting forecast outcomes. The following narrative delves into these transformative processes, providing a comprehensive understanding of the deliberate modifications made to both model and dataset.

In this chapter, we will be looking at:

1. **Model Modification.**
2. **Dataset Manipulation.**
3. **Models Results.**
4. **Summary.**

Model modification:

I. Comparison table for the entry-level model and primary model made by Linear regression:

Linear regression model		
Comparison	**Entry-level model**	**Primary model**
Model Architecture and Data Preprocessing	▪ Utilizes a linear regression framework on any store item. ▪ specifically filtered for store 1. ▪ Focuses on feature engineering with date-related attributes, lag features, and rolling statistics.	▪ Employs a comprehensive approach to any store item. ▪ filtering for store 1. ▪ Incorporates advanced analyses such as ACF/PACF plots, seasonal decomposition, and correlation matrices, alongside feature engineering.
Feature Selection	▪ Applies SelectKBest for feature selection. ▪ emphasizing simplicity and interpretability in model design.	▪ Leverages SelectKBest for feature selection. ▪ incorporating a broader set of features and analyses. ▪ suitable for capturing nuanced time-series dynamics.
Model Training and Evaluation	▪ Trains the model using linear regression.	▪ Utilizes linear regression for training. ▪ additional evaluation metrics and visualizations, showcasing a more comprehensive understanding of time-series patterns.

	▪ accompanied by a rigorous evaluation using metrics like MAE, RMSE, MAPE, and Accuracy. ▪ Visualizations provide insights into model performance.	
Model Insights and Applicability	▪ Focused on simplicity and streamlined interpretability. ▪ suitable for scenarios where a straightforward approach suffices.	▪ Offers a more intricate solution. ▪ integrate advanced time-series analyses for a nuanced comprehension of underlying patterns. ▪ particularly valuable when detailed insights into time-series dynamics are crucial.
summary	▪ while the **entry-level model** prioritizes simplicity and interpretability. ▪ **The primary model** presents a more detailed and comprehensive approach, integrating advanced time-series analyses for a nuanced understanding of sales forecasting dynamics. ▪ The choice between the two models depends on the specific requirements and characteristics of the sales forecasting task at hand.	

Table 3. 1. Linear regressions Entry-level & Primary Models Comparison.

II. Comparison table for the entry-level model and primary model made by ARIMA:

ARIMA Model		
Comparison	**Entry-level model**	**Primary model**
Model Architecture and Data Preprocessing	▪ Data Loading: Reads data from CSV file and formats the date. ▪ Indexing: Sets the date as the index for time-series analysis. ▪ Feature Engineering: Creates additional date-related features (year, month, day, weekday). ▪ Train-Test Split: Segregates data into training and testing sets. ▪ Modeling: Utilizes SARIMAX with order= (1, 1, 0) and seasonal_order= (1, 1, 0, 7). ▪ Forecasting: Generates predictions for the test dataset. ▪ Evaluation: Calculates metrics such as MAE, RMSE, R2, and Adjusted R2. ▪ Visualization: Displays a time-series plot with forecast and error magnitude. ▪ Summary Statistics: Presents total actual and predicted sales.	▪ Data Loading and Filtering: Reads data from CSV file and filters for store 1. ▪ Indexing and Feature Creation: Sets date as the index and adds date-related features. ▪ Stationarity Test: Conducts Augmented Dickey-Fuller Test for stationarity. ▪ Model Selection and Fitting: Fits ARIMA (6, 1, 1) and SARIMAX models. ▪ Forecasting: Generates predictions for the test dataset. ▪ Evaluation: Computes MAE, RMSE, and MAPE metrics. ▪ Visualization: Displays plots of train, test, predicted sales, and prediction errors. ▪ Summary Statistics: Aggregates total actual and predicted sales, along with error metrics.
Feature Selection	▪ No explicit feature selection; ▪ SARIMAX handles feature importance implicitly.	▪ No specific feature selection methodology was mentioned.

Model Training and Evaluation	▪ Training Method: SARIMAX with specified orders. ▪ Evaluation Metrics: MAE, RMSE, R2, Adjusted R2, and Accuracy.	▪ Training Methods: ARIMA (6, 1, 1) and SARIMAX with specified orders. ▪ Evaluation Metrics: MAE, RMSE, MAPE.
Model Insights and Applicability	▪ Focused on SARIMAX time-series modeling with a detailed statistical summary. ▪ Suited for scenarios requiring in-depth time-series analysis and accurate sales forecasting.	▪ Integrates ARIMA and SARIMAX models, emphasizing stationarity testing. ▪ Suitable for applications where both classical time-series and advanced modeling approaches are valuable.
summary	▪ while both models exhibit a commitment to rigorous time-series analysis. ▪ entry-level model emphasizes SARIMAX modeling with a detailed statistical summary, making it apt for scenarios requiring a nuanced understanding of time-series dynamics. ▪ In contrast, the primary model combines ARIMA and SARIMAX models, highlighting the importance of stationarity testing and providing valuable insights for applications that benefit from a hybrid modeling approach. ▪ The choice between the two models depends on the specific requirements and characteristics of the sales forecasting task at hand.	

Table 3. 2. ARIMA Entry-level & Primary Models Comparison.

III. Comparison table for the entry-level model and primary model made by the Exponential smoothing model:

Holt Winter's triple Exponential Smoothing model		
comparison	**Entry-level model**	**Primary model**
Data Preprocessing	▪ Reads data from CSV file. ▪ Converts the 'date' column to a datetime format. ▪ Sets the 'date' column as the index. ▪ Filters data for 'store' equal to 1. ▪ Splits the data into training and testing sets.	▪ Reads data from CSV file. ▪ Filters data for 'store' equal to 1. ▪ Converts the 'date' column to a datetime format. ▪ Sets the 'date' column as the index. ▪ Splits the data into training and testing sets.
Seasonal decomposition	——	▪ Uses seasonal decomposition to visualize the trend, seasonal, and residual components of the training data.
Model training	▪ Uses the Holt-Winters method with additive trend and additive seasonal components. ▪ Fits two models: one without damping and one with damping.	▪ Uses the Holt-Winters method with additive trend and additive seasonal components. ▪ Fits two models: one without damping and one with damping.
Visualization	▪ Plots actual test sales along with forecasts for both models.	▪ Plots train sales, actual test sales, and forecasts for both models. ▪ Plots the forecast errors for both models.

	▪ Plots the error magnitude for both models.	
Evaluation metrics	▪ Defines custom metrics functions for Mean Squared Logarithmic Error (MSLE), Mean Percentage Error (MPE), and Adjusted R-squared. ▪ Calculates and displays metrics such as MAE, RMSE, R2, and Accuracy for both models. ▪ Prints the total actual sales and total predicted sales for both models.	▪ Defines custom metrics functions for Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). ▪ Calculates and displays metrics for both models. ▪ Prints the total actual sales, total predicted sales, and overall error for both models.
summary	▪ Both models follow a similar approach using the Holt-Winters method but differ in terms of data preprocessing, visualization, and custom metrics. ▪ The entry-level model focuses on plotting error magnitude and uses R2 as an accuracy metric, ▪ The primary model has a detailed seasonal decomposition plot and uses MAPE as an additional evaluation metric. ▪ The choice between the two models may depend on specific preferences and requirements for forecasting and model evaluation.	

Table 3. 3. Exponential Smoothing Entry-level & Primary Models Comparison.

Dataset manipulation:

A deliberate approach was taken to enhance the granularity and test the adaptability of the sales forecasting model. The dataset underwent a meticulous process of fragmentation, with individual files dedicated to each item and a comprehensive file amalgamating all data. The goal was to gauge the model's response to nuanced changes, such as manipulating the sales date from daily to monthly intervals. This intentional tweaking aimed to unravel the model's sensitivity to variations in data structure and temporal granularity, providing valuable insights into its robustness and responsiveness to diverse scenarios.

Model results:

I. Linear regression results:

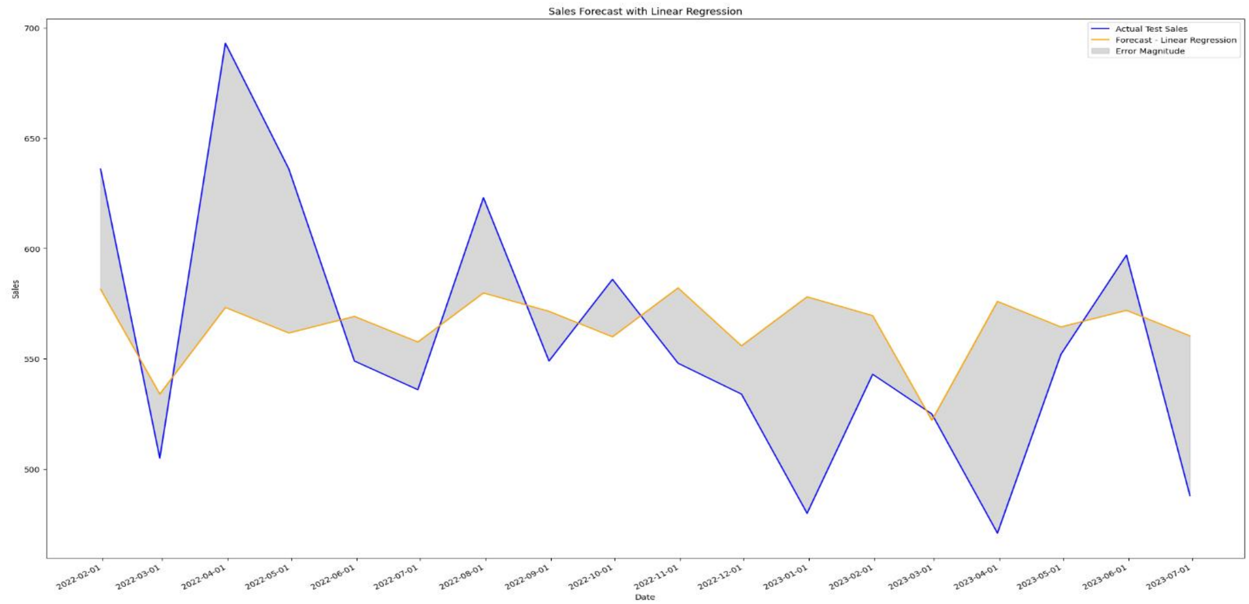


Figure 4. 1. LR-PM Forecast for Item One.

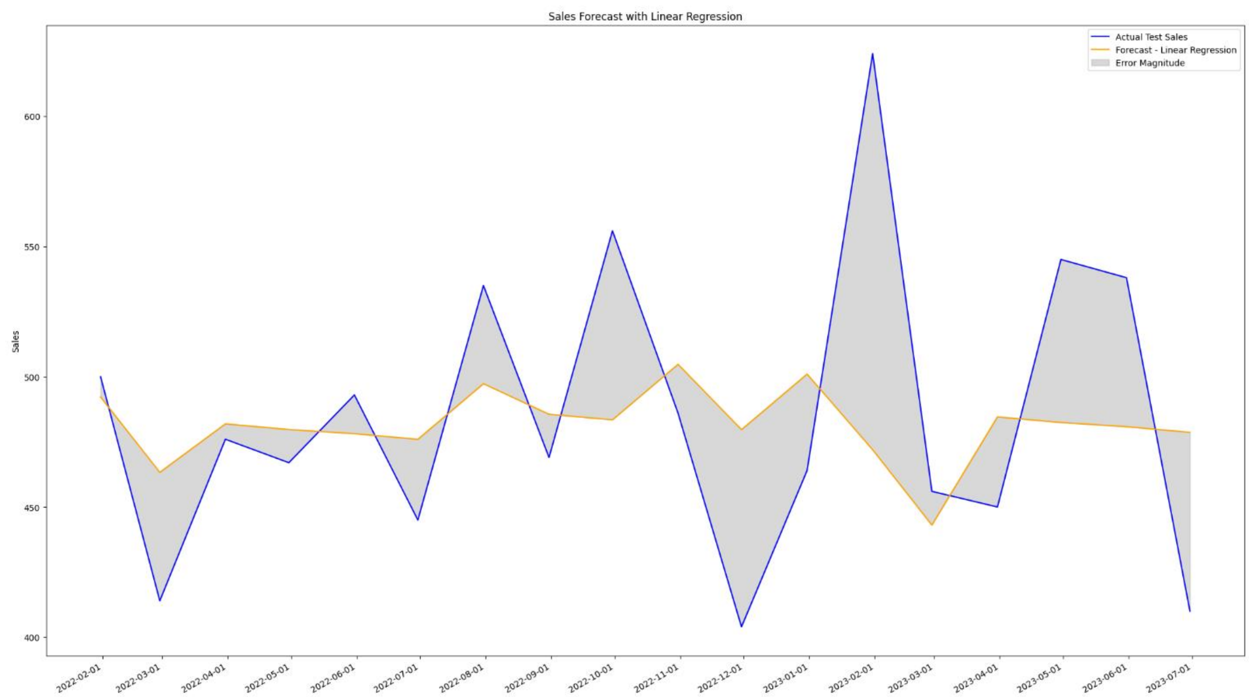


Figure 4. 2. LR-PM Forecast for Item Two.

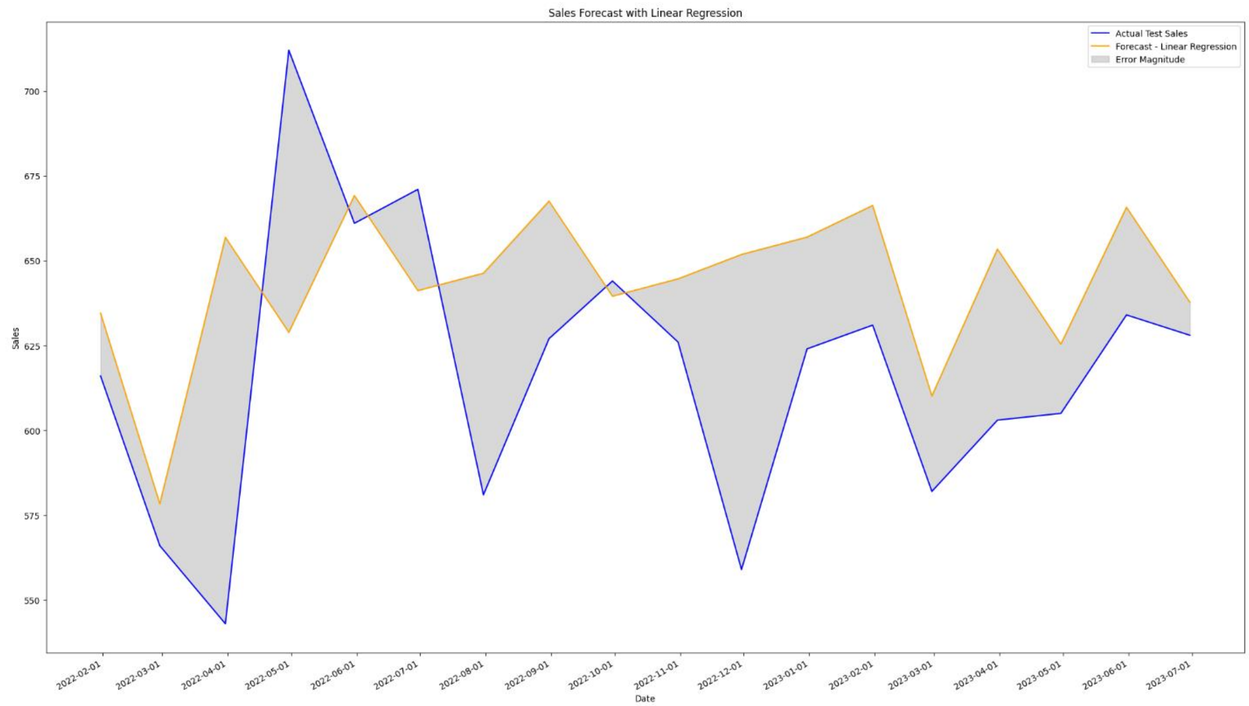


Figure 4. 3. LR-PM Forecast for Item Three.

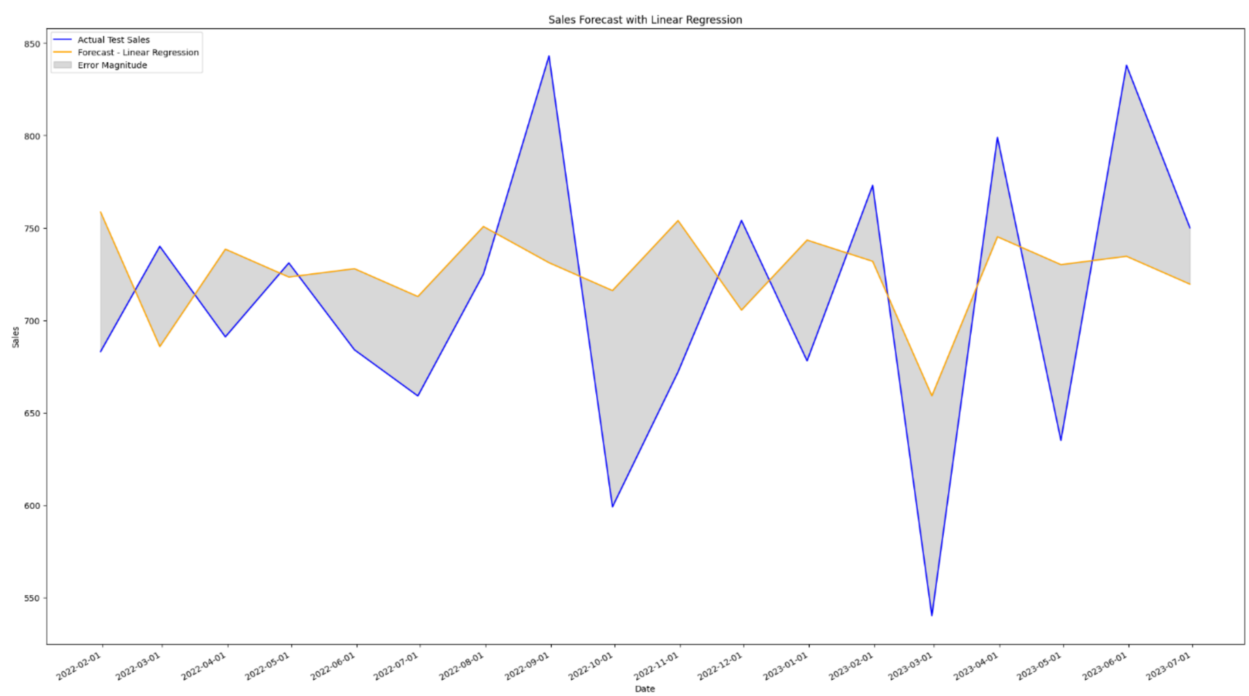


Figure 4. 4. LR-PM Forecast for Item Four.

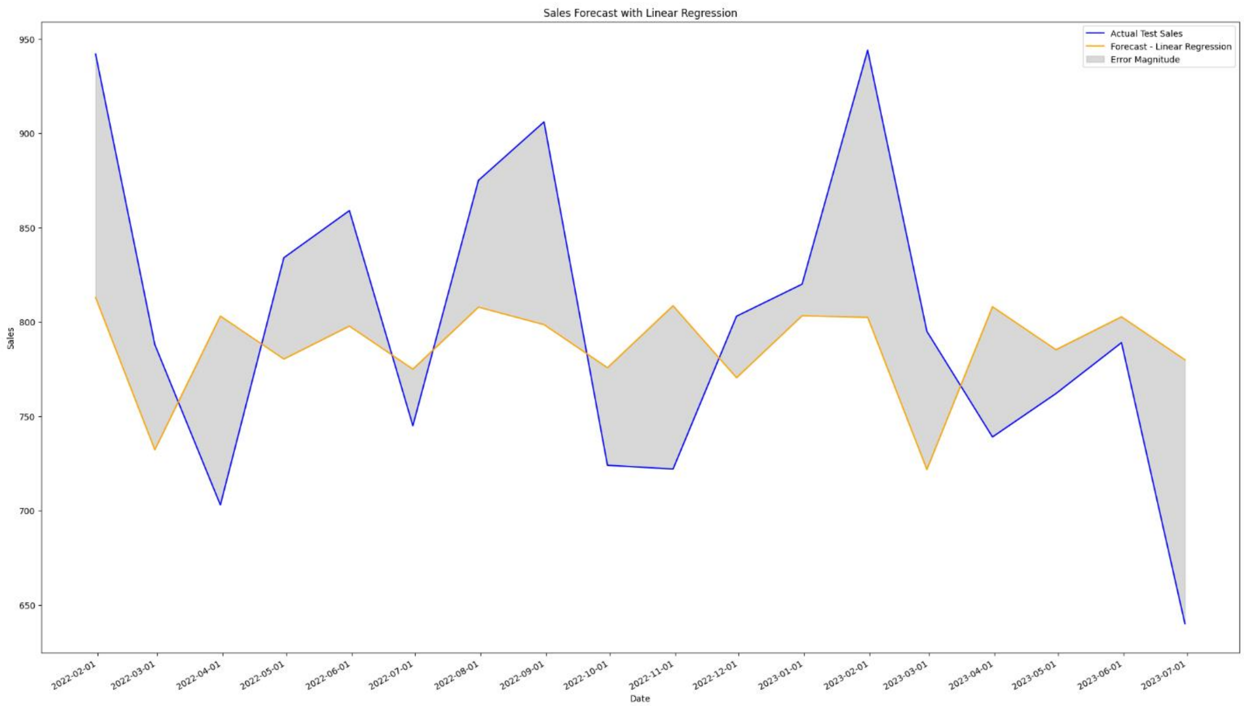


Figure 4. 5. LR-PM Forecast for Item Five.

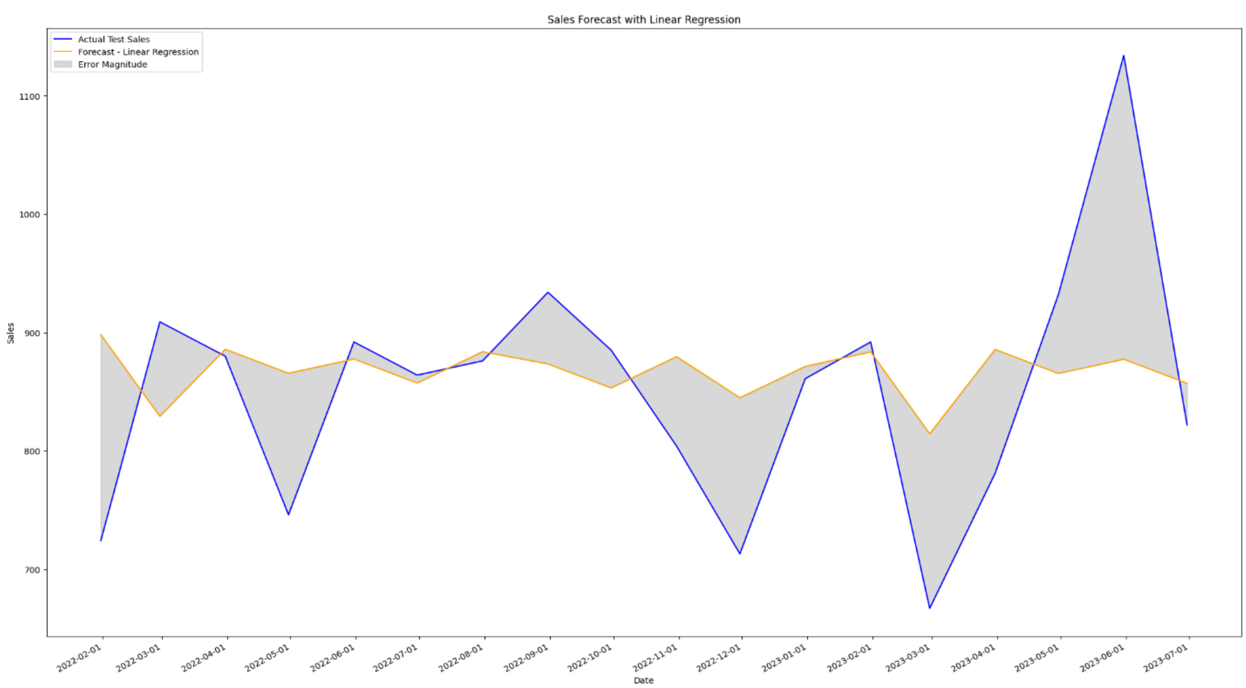


Figure 4. 6. LR-PM Forecast for Item Six.

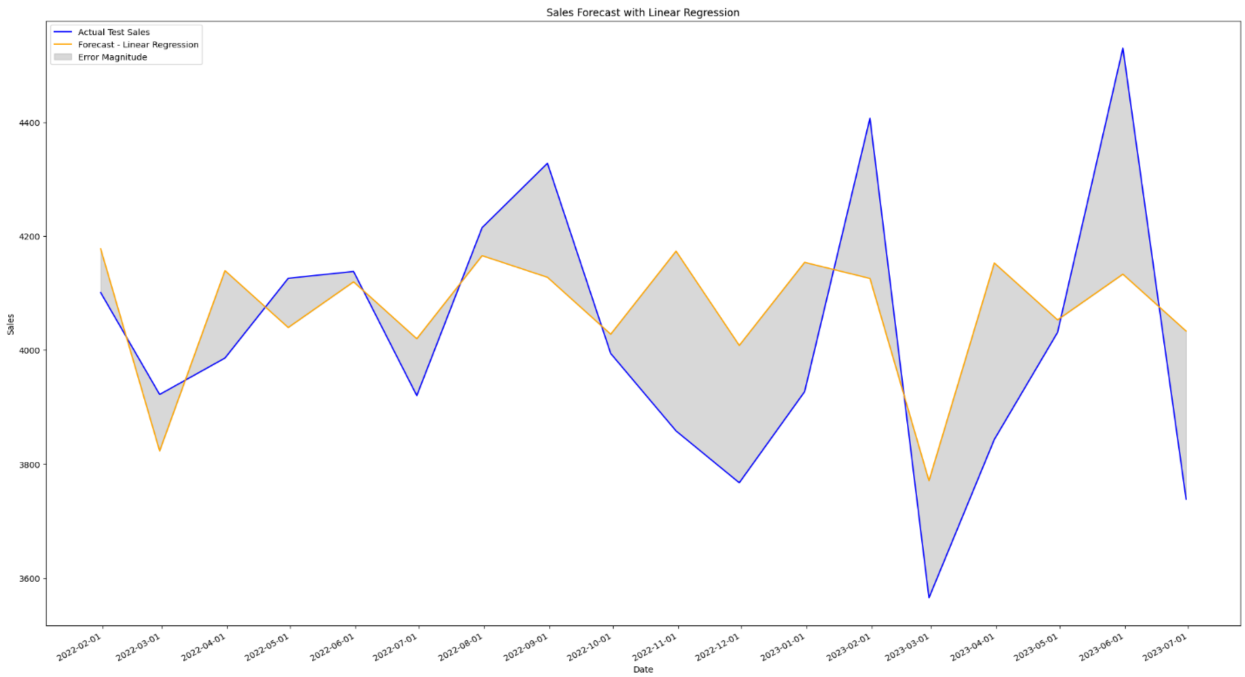


Figure 4. 7. LR-PM Forecast for Store Items.

	Metric	Value
0	MAE	44.981949
1	RMSE	56.092928
2	MAPE	8.126070
3	Accuracy	7.085278

Figure 4. 8. LR-PM Metrics for Item One.

	Metric	Value
0	MAE	42.660614
1	RMSE	55.106307
2	MAPE	8.631555
3	Accuracy	1.611151

Figure 4. 9. LR-PM Metrics for Item Two.

	Metric	Value
0	MAE	38.678579
1	RMSE	49.100483
2	MAPE	6.415502
3	Accuracy	-45.904880

Figure 4. 10. LR-PM Metrics for Item Three.

	Metric	Value
0	MAE	65.336911
1	RMSE	72.814165
2	MAPE	9.500125
3	Accuracy	8.880370

Figure 4. 11. LR-PM Metrics for Item Four.

	Metric	Value
0	MAE	69.615618
1	RMSE	80.074302
2	MAPE	8.756294
3	Accuracy	2.315994

Figure 4. 12. LR-PM Metrics for Item Five.

	Metric	Value
0	MAE	74.309535
1	RMSE	100.992613
2	MAPE	9.028653
3	Accuracy	3.959324

Figure 4. 13. LR-PM Metrics for Item Six.

	Metric	Value
0	MAE	172.898211
1	RMSE	207.185579
2	MAPE	4.320232
3	Accuracy	23.705739

Figure 4. 14. LR-PM Metrics for Store Items.

```
Total Actual Sales: 10051.00
Total Predicted Sales: 10169.33
```

Figure 4. 15. LR-PM Sales for Item One.

```
Total Actual Sales: 8732.00
Total Predicted Sales: 8664.62
```

Figure 4. 16. LR-PM Sales for Item Two.

```
Total Actual Sales: 11113.00
Total Predicted Sales: 11574.26
```

Figure 4. 17. LR-PM Sales for Item Three.

```
Total Actual Sales: 12794.00
Total Predicted Sales: 13067.91
```

Figure 4. 18. LR-PM Sales for Item Four.

```
Total Actual Sales: 14390.00
Total Predicted Sales: 14165.60
```

Figure 4. 19. LR-PM Sales for Item Five.

```
Total Actual Sales: 15316.00
Total Predicted Sales: 15603.57
```

Figure 4. 20. LR-PM Sales for Item Six.

```
Total Actual Sales: 72396.00
Total Predicted Sales: 73245.29
```

Figure 4. 21. LR-PM Sales for Store Items.

II. ARIMA results:

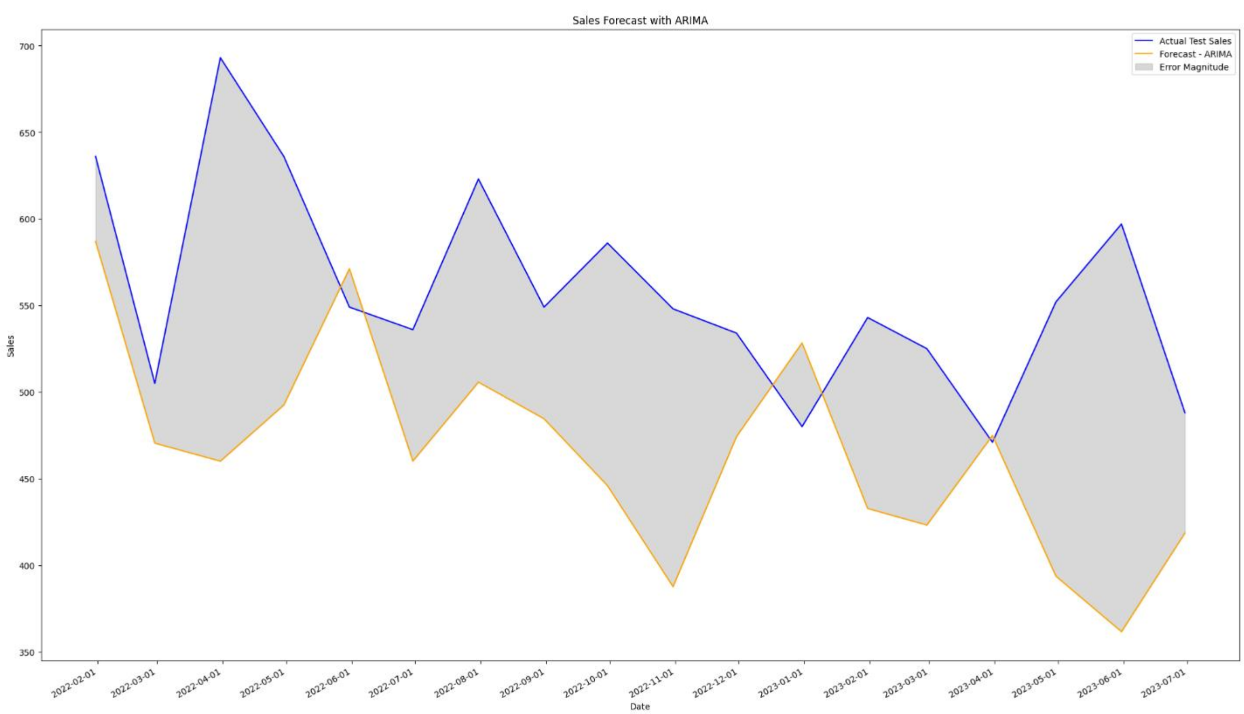


Figure 4. 22. A-PM Forecast for Item One.

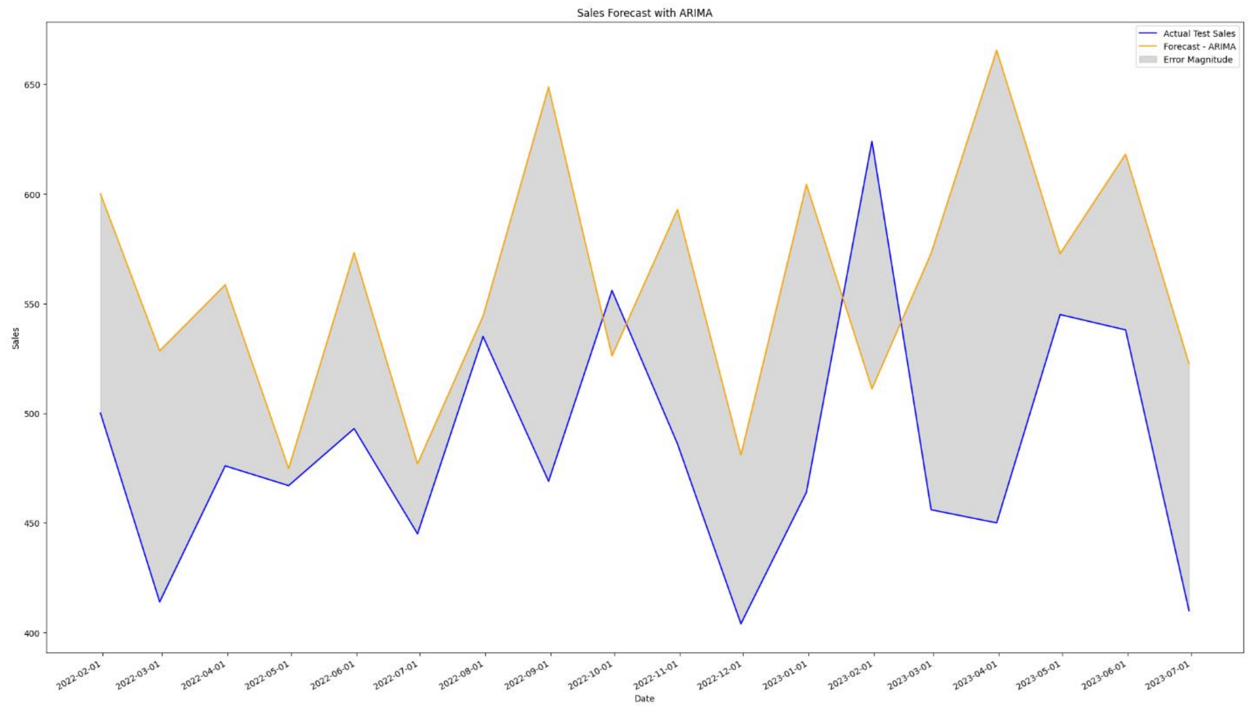


Figure 4. 23. A-PM Forecast for Item Two.

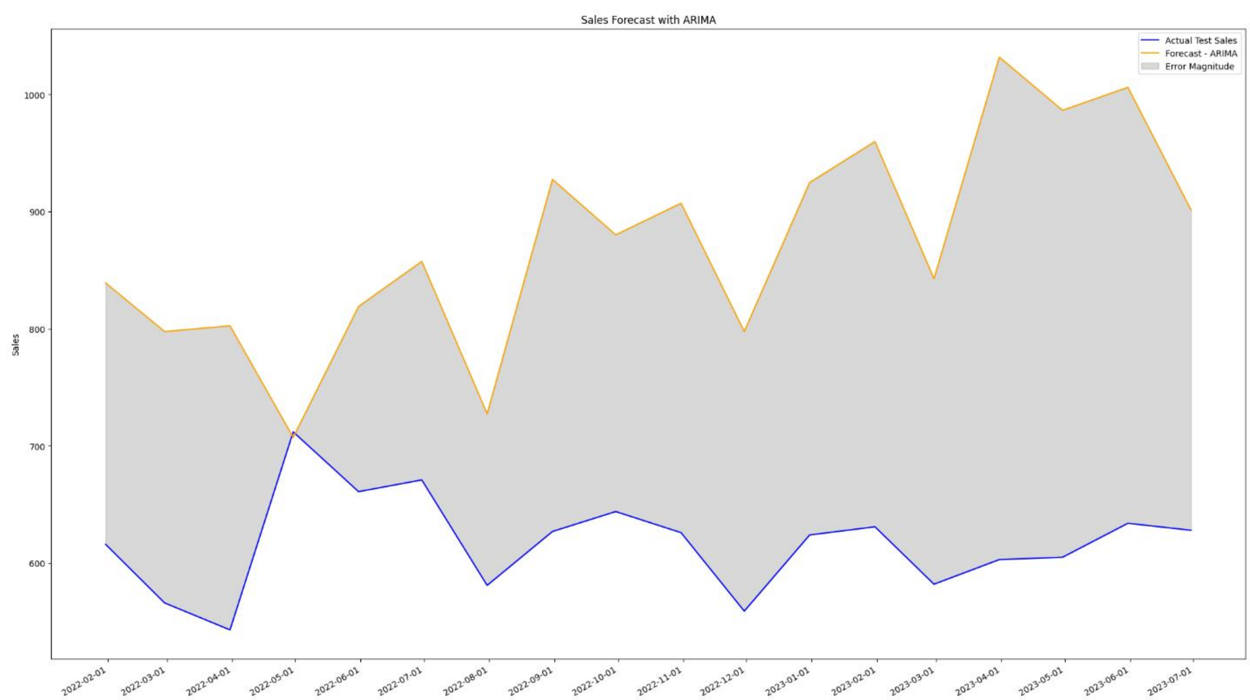


Figure 4. 24. A-PM Forecast for Item Three.

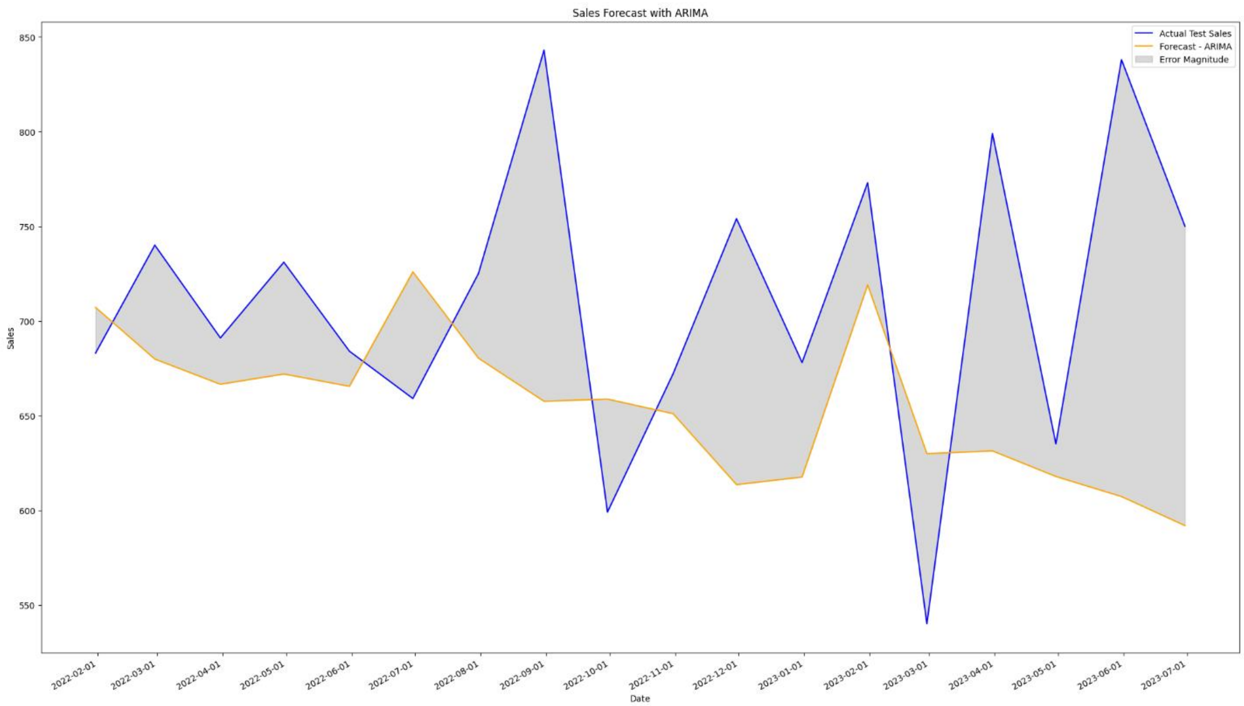


Figure 4. 25. A-PM Forecast for Item Four.

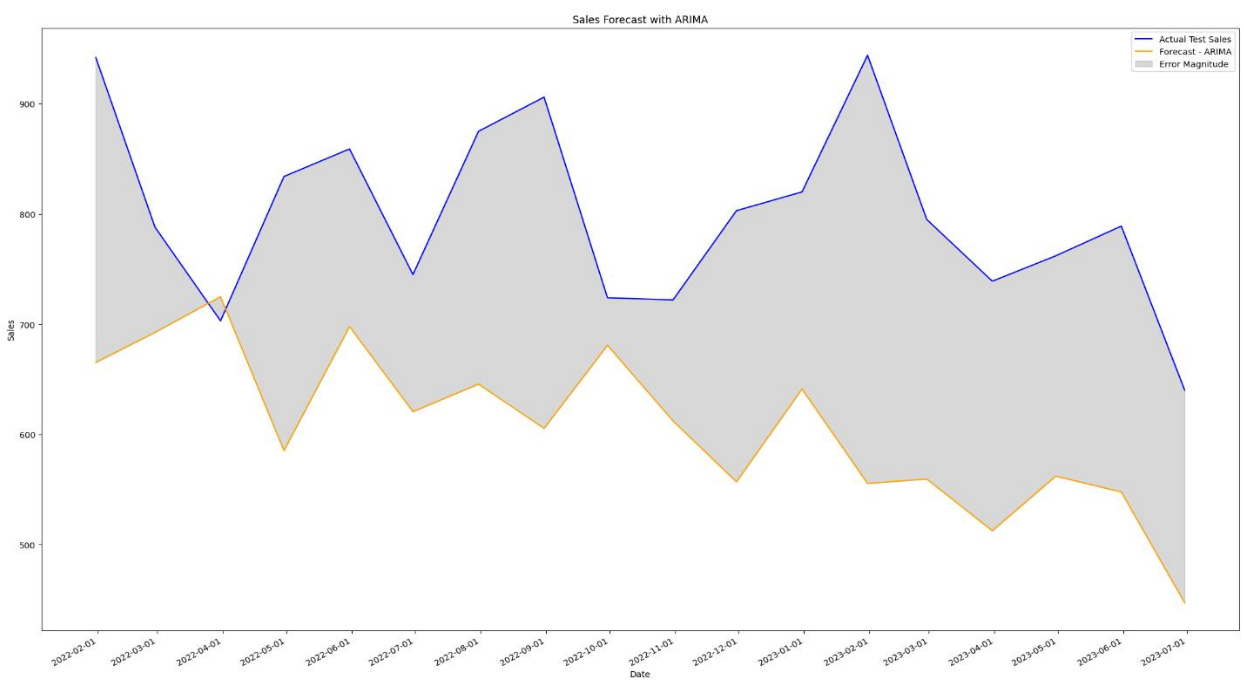


Figure 4. 26. A-PM Forecast for Item Five.

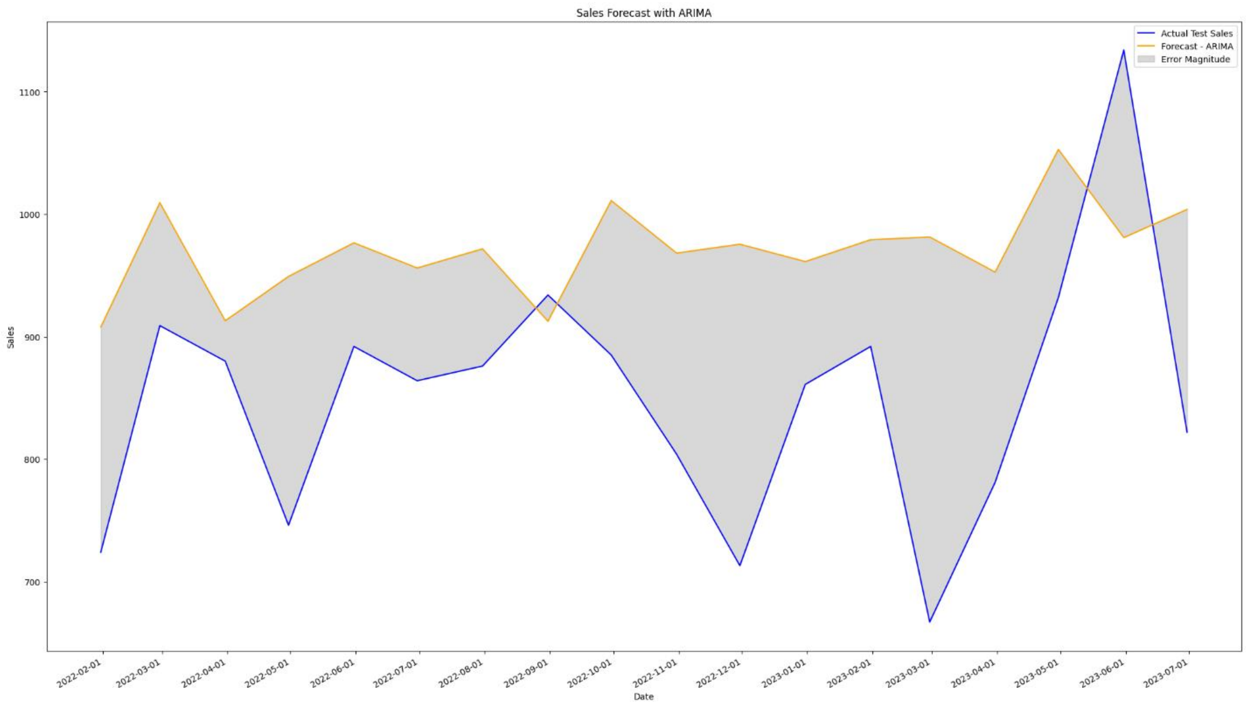


Figure 4. 27. A-PM Forecast for Item Six.

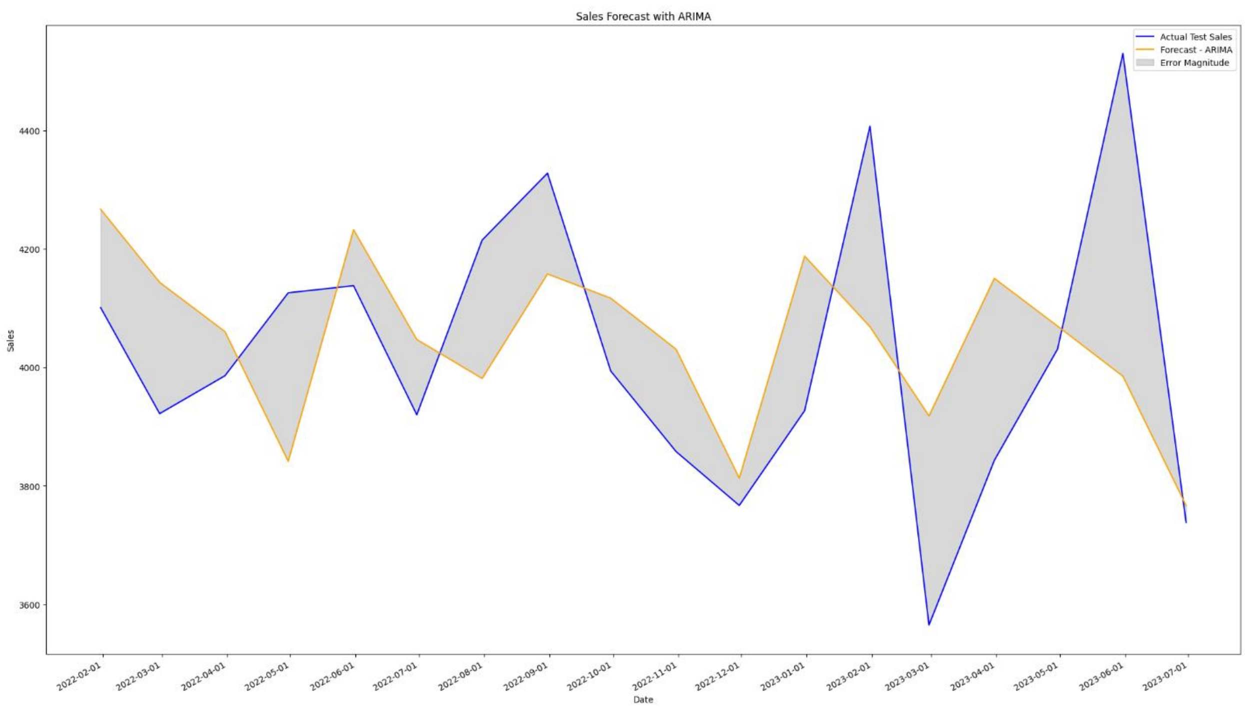


Figure 4. 28. A-PM Forecast for Store Items.

	Metric	Value
0	MAE	101.538058
1	RMSE	120.586068
2	R2	-3.294005
3	Adjusted R2	-3.562381
4	Accuracy	-329.400537

Figure 4. 29. A-PM Metrics for Item One.

	Metric	Value
0	MAE	90.358858
1	RMSE	105.572899
2	R2	-2.611172
3	Adjusted R2	-2.836870
4	Accuracy	-261.117194

Figure 4. 30. A-PM Metrics for Item Two.

	Metric	Value
0	MAE	256.259453
1	RMSE	273.183768
2	R2	-44.165629
3	Adjusted R2	-46.988481
4	Accuracy	-4416.562875

Figure 4. 31. A-PM Metrics for Item Three.

	Metric	Value
0	MAE	82.377125
1	RMSE	103.972722
2	R2	-0.857886
3	Adjusted R2	-0.974004
4	Accuracy	-85.788589

Figure 4. 32. A-PM Metrics for Item Four.

	Metric	Value
0	MAE	195.621399
1	RMSE	215.012054
2	R2	-6.043082
3	Adjusted R2	-6.483275
4	Accuracy	-604.308225

Figure 4. 33. A-PM Metrics for Item Five.

	Metric	Value
0	MAE	138.701079
1	RMSE	156.423073
2	R2	-1.303974
3	Adjusted R2	-1.447973
4	Accuracy	-130.397422

Figure 4. 34. A-PM Metrics for Item Six.

	Metric	Value
0	MAE	199.276120
1	RMSE	238.081304
2	R2	-0.007450
3	Adjusted R2	-0.070415
4	Accuracy	-0.744979

Figure 4. 35. A-PM Metrics for Store Items.

```
Total Actual Sales: 10051.00
Total Predicted Sales: 8372.19
```

Figure 4. 36. A-PM Sales for Item One.

```
Total Actual Sales: 8732.00
Total Predicted Sales: 10073.03
```

Figure 4. 37. A-PM Sales for Item Two.

```
Total Actual Sales: 11113.00
Total Predicted Sales: 15716.18
```

Figure 4. 38. A-PM Sales for Item Three.

```
Total Actual Sales: 12794.00
Total Predicted Sales: 11792.53
```

Figure 4. 39. A-PM Sales for Item Four.

```
Total Actual Sales: 14390.00
Total Predicted Sales: 10912.58
```

Figure 4. 40. A-PM Sales for Item Five.

```
Total Actual Sales: 15316.00
Total Predicted Sales: 17463.49
```

Figure 4. 41. A-PM Sales for Item Six.

```
Total Actual Sales: 72396.00
Total Predicted Sales: 72839.62
```

Figure 4. 42. A-PM Sales for Store Items.

III. Exponential smoothing results:

Figure 4. 43. ES-PM Forecast for Item One (without Dumping).

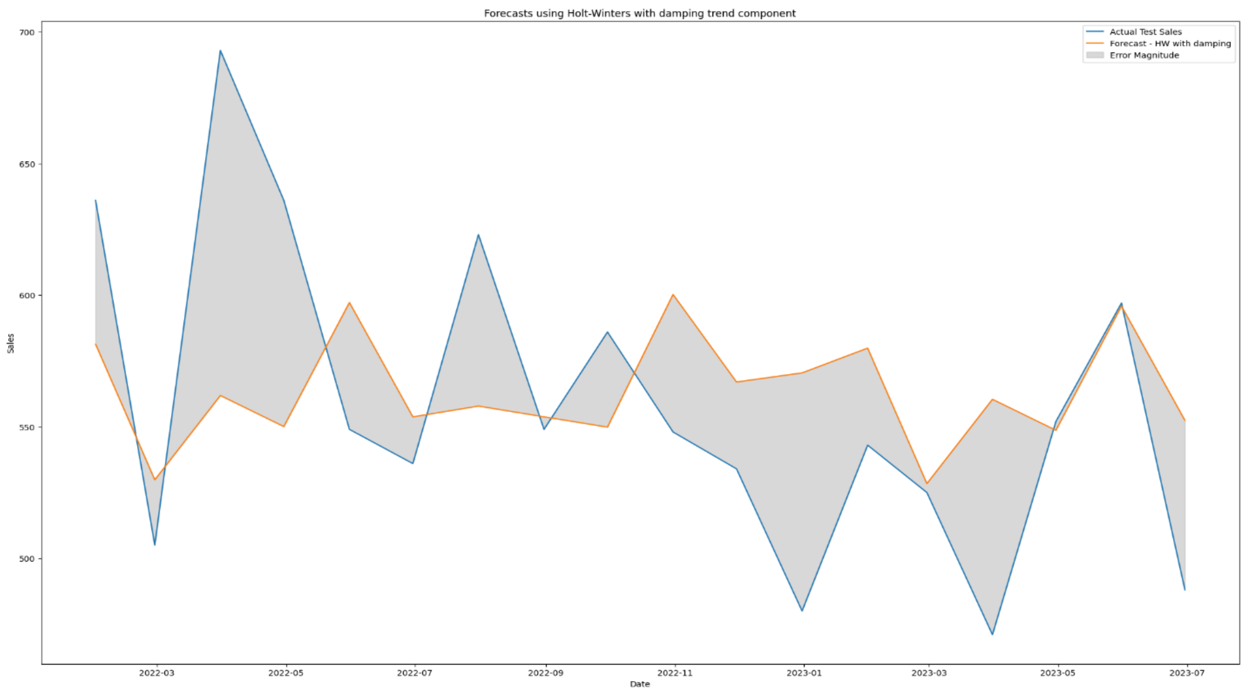


Figure 4. 44. ES-PM Forecast for Item One (with Dumping).

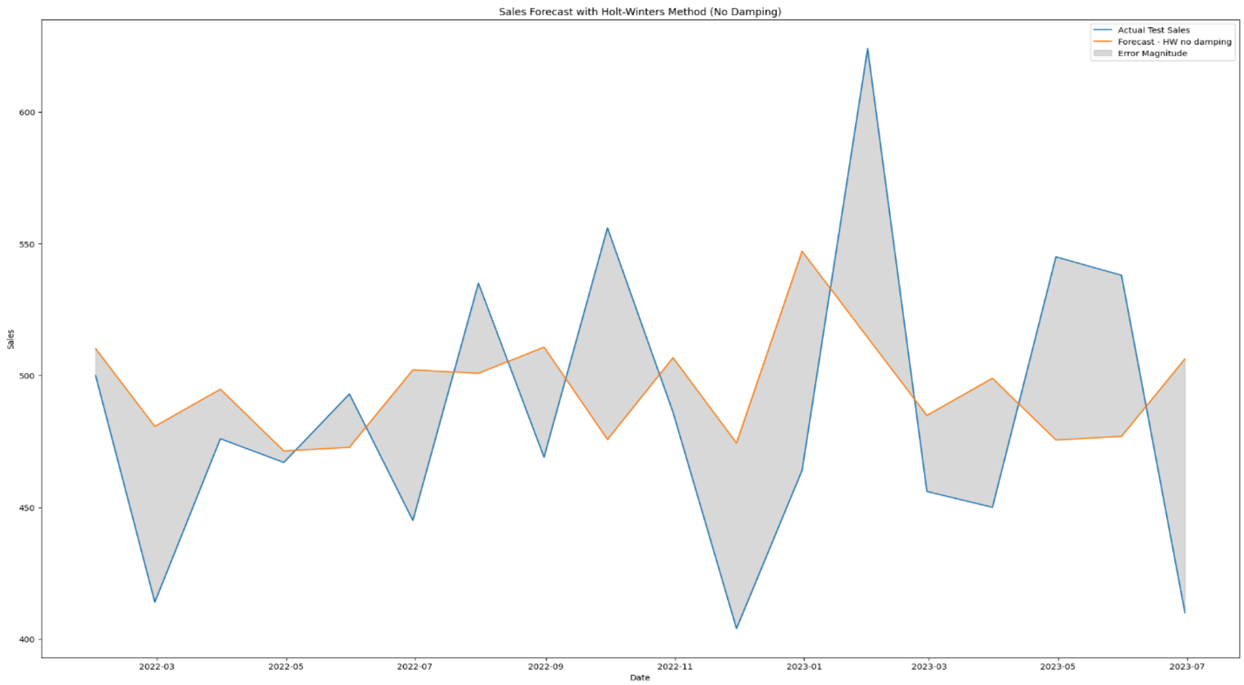


Figure 4. 45. ES-PM Forecast for Item Two (without Dumping).

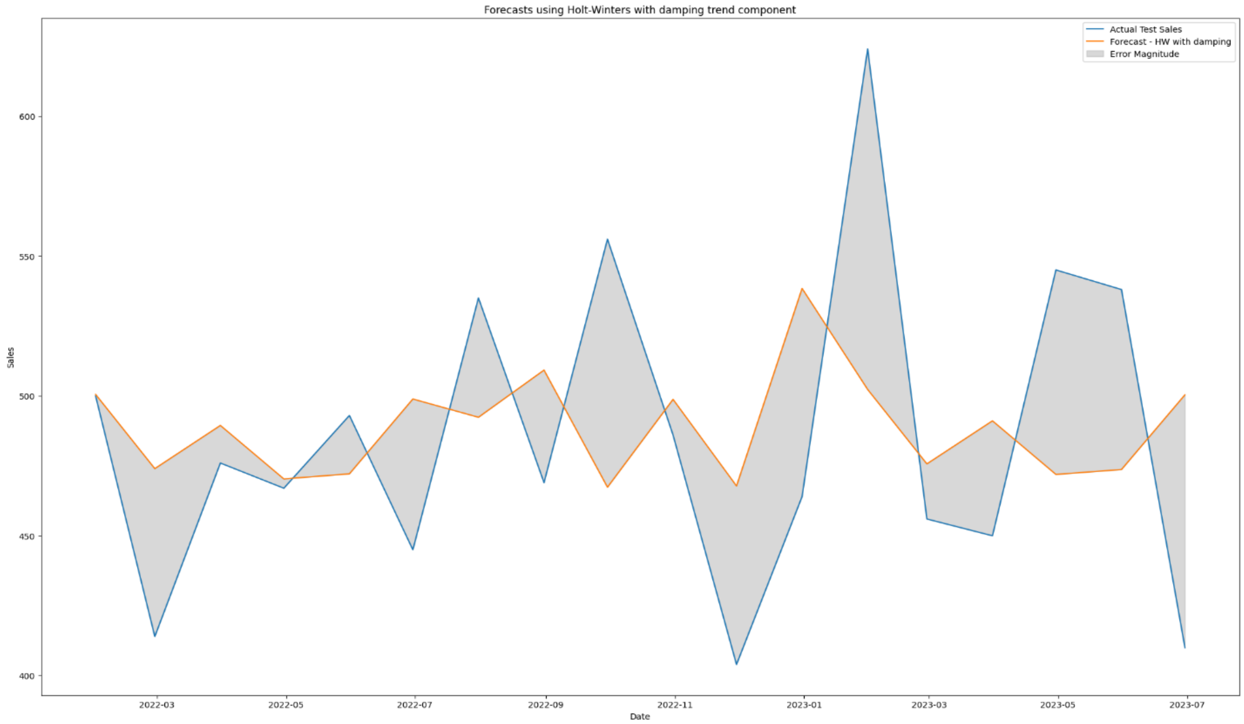


Figure 4. 46. ES-PM Forecast for Item Two (with Dumping).

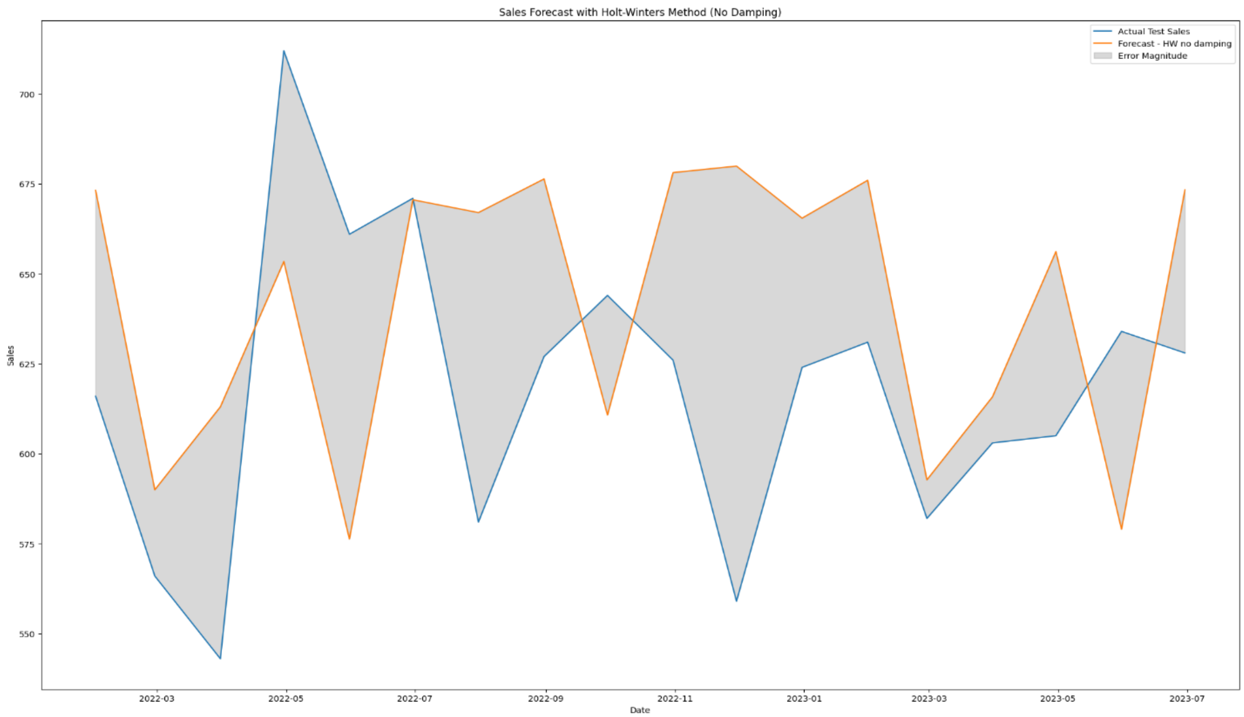


Figure 4. 47. ES-PM Forecast for Item Three (without Dumping).

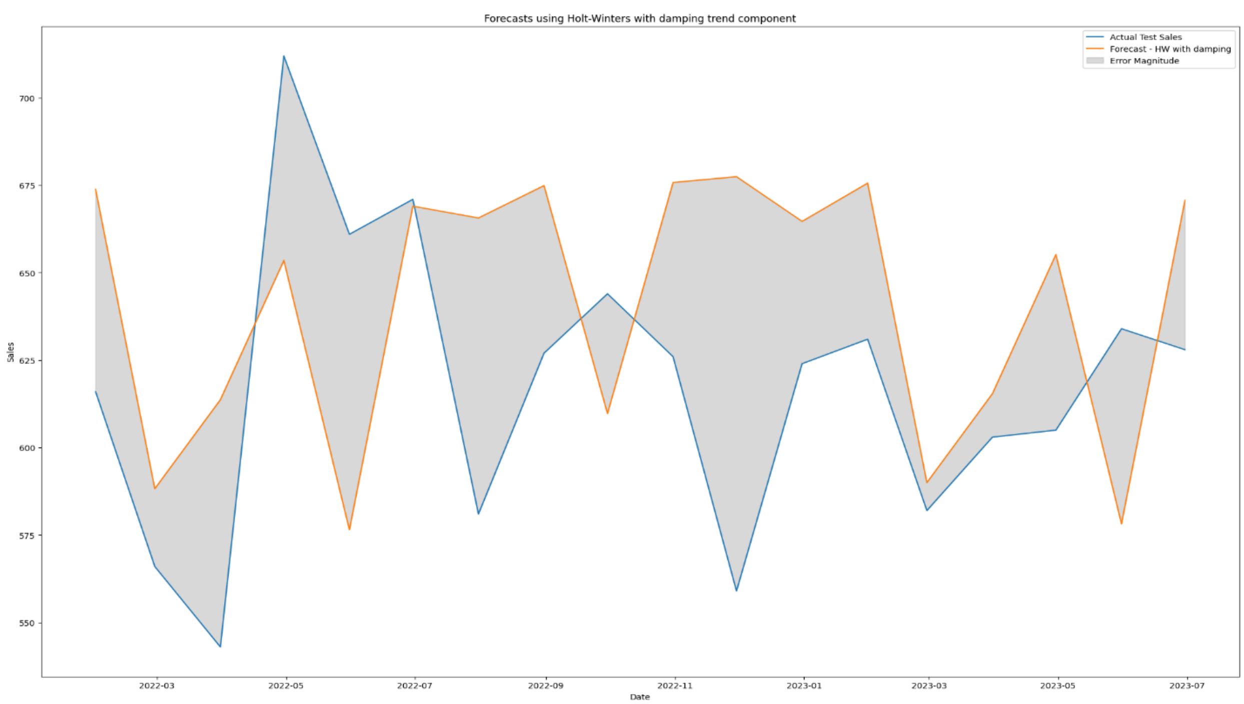


Figure 4. 48. ES-PM Forecast for Item Three (with Dumping).

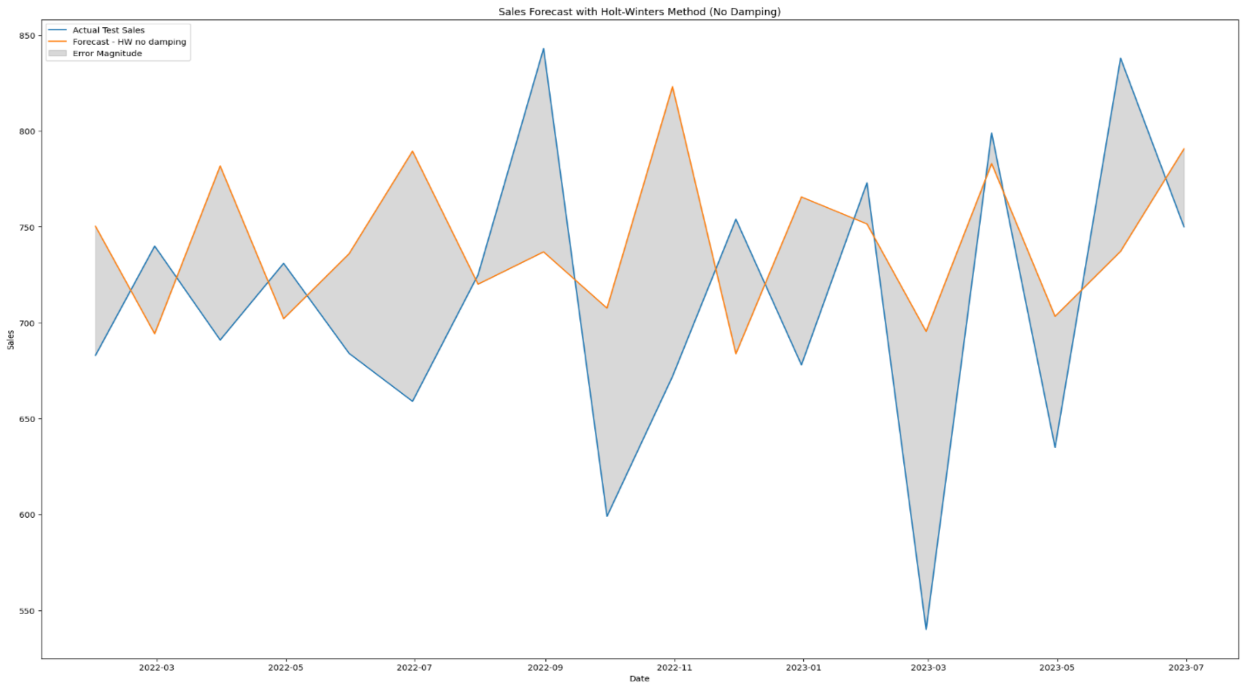


Figure 4. 49. ES-PM Forecast for Item Four (without Dumping).

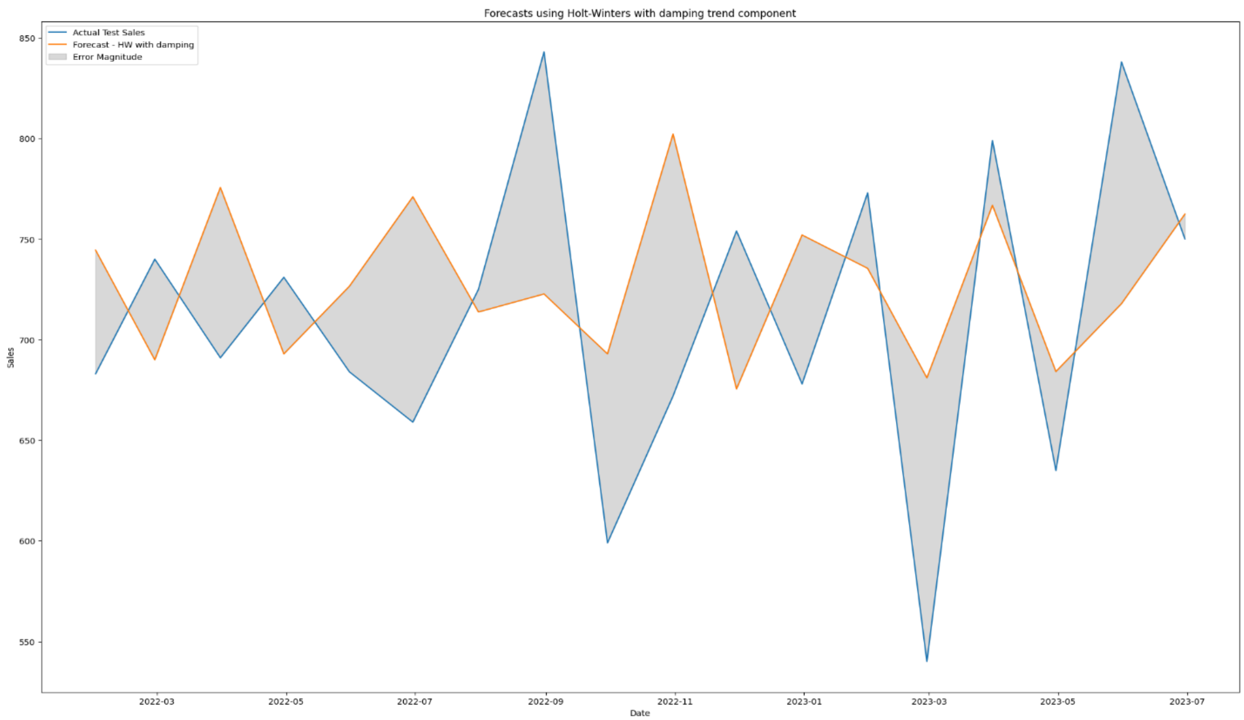


Figure 4. 50. ES-PM Forecast for Item Four (with Dumping).

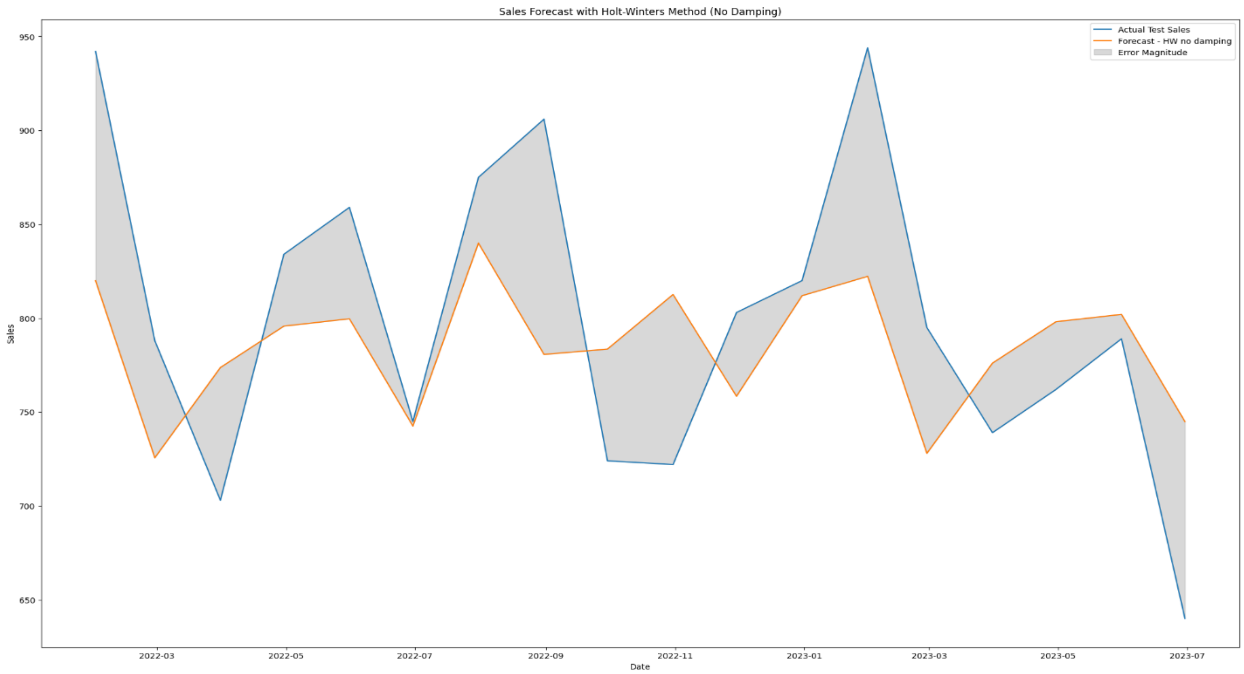


Figure 4. 51. ES-PM Forecast for Item Five (without Dumping).

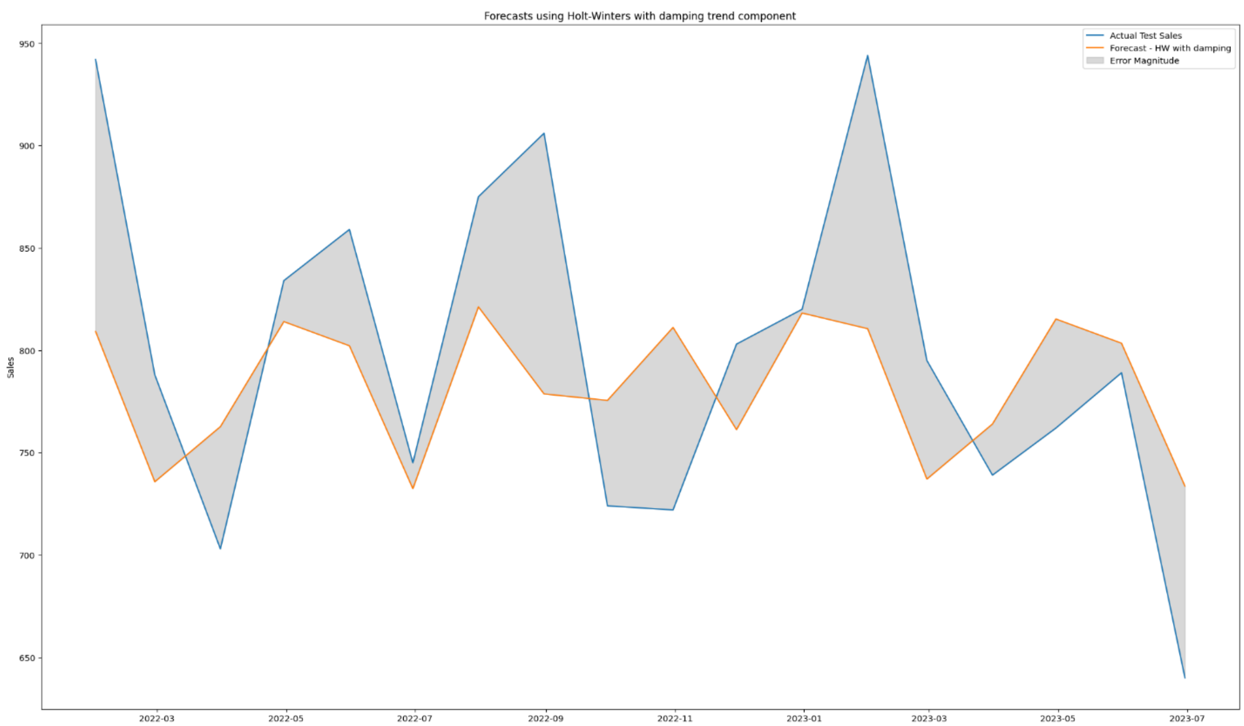


Figure 4. 52. ES-PM Forecast for Item Five (with Dumping).

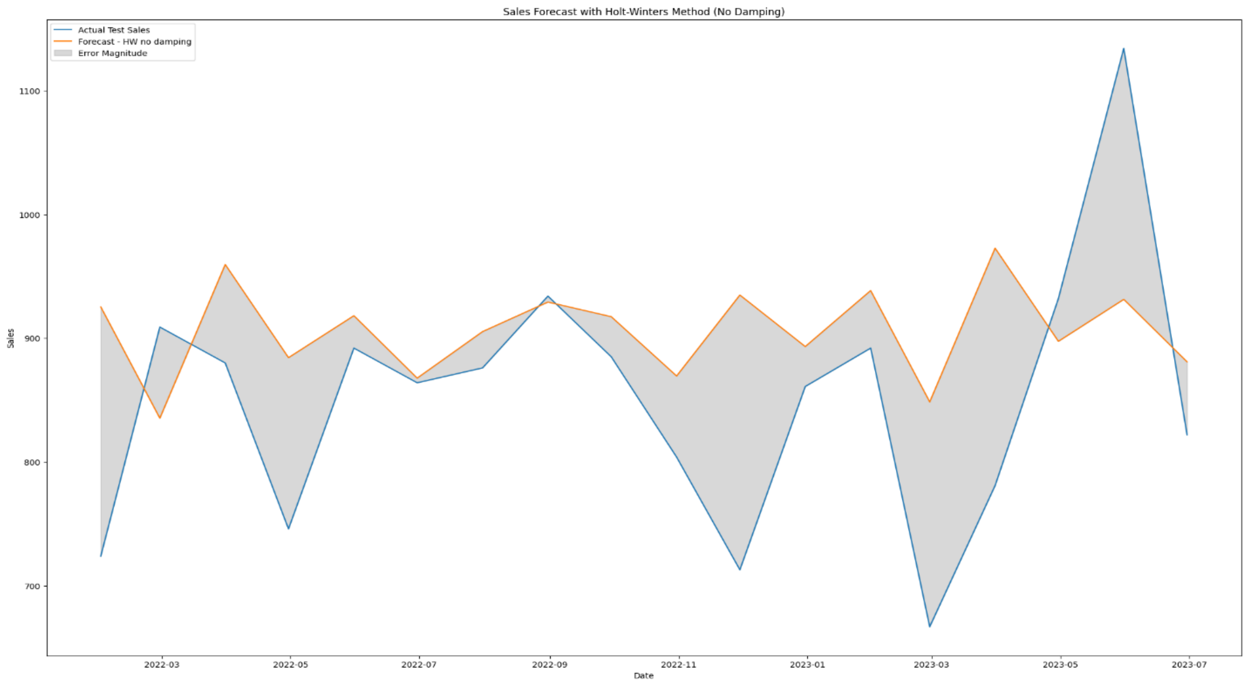


Figure 4. 53. ES-PM Forecast for Item Six (without Dumping).

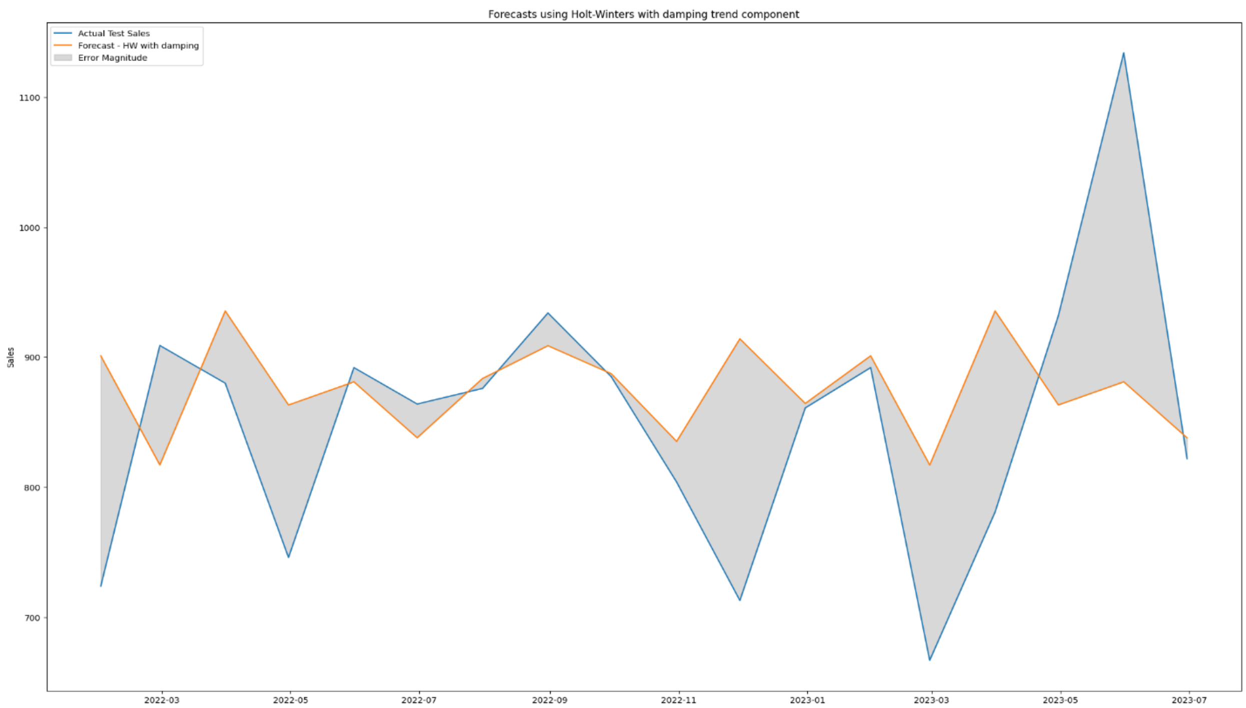


Figure 4. 54. ES-PM Forecast for Item Six (with Dumping).

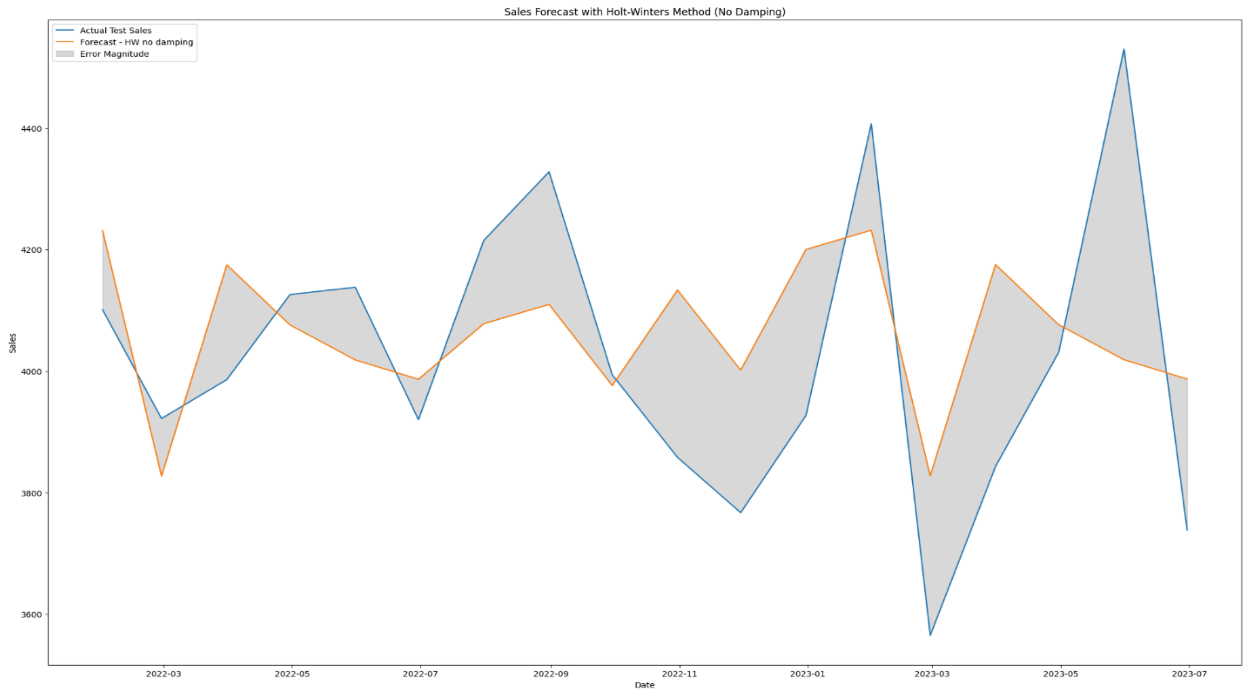


Figure 4. 55. ES-PM Forecast for Store Items (without Dumping).

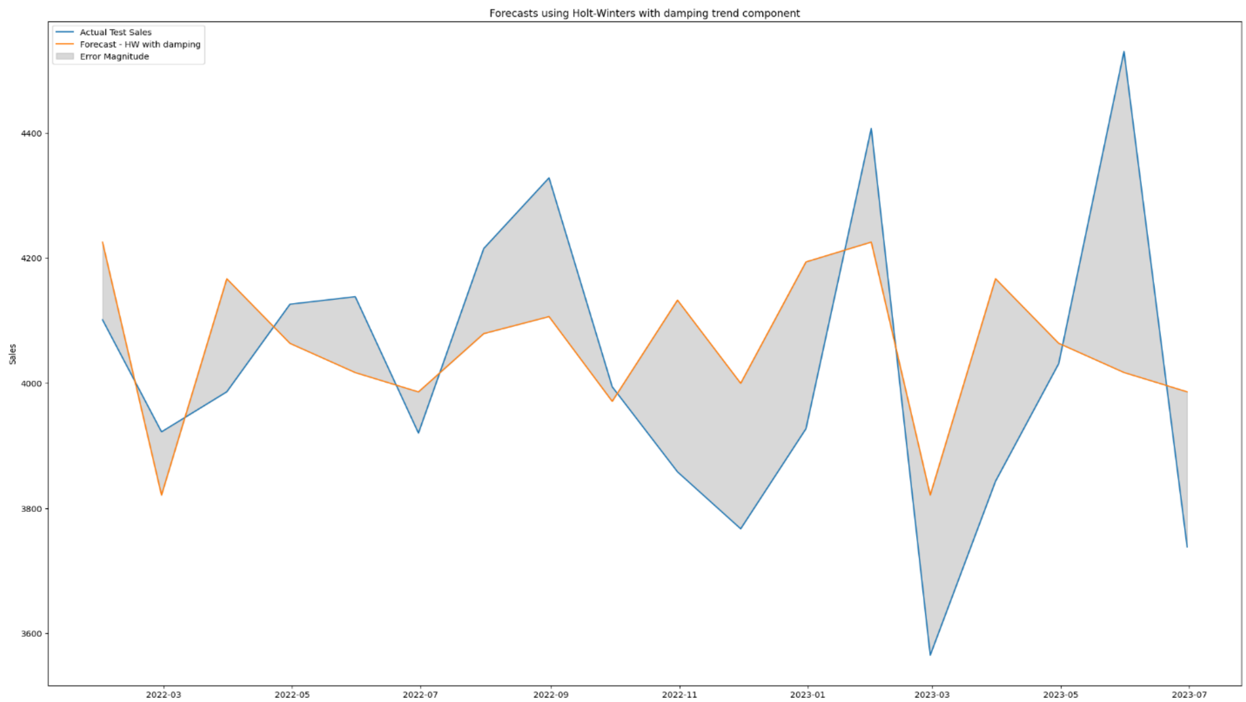


Figure 4. 56. ES-PM Forecast for Store Items (with Dumping).

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	45.555732	46.831468
1	RMSE	58.535455	58.659223
2	R2	-0.011827	-0.016110
3	Accuracy	-1.182710	-1.611046

```
Total Actual Sales: 10051.00
Total Predicted Sales (No Damping): 9755.23
Total Predicted Sales (With Damping): 10138.91
```

Figure 4. 57. ES-PM Sales & Metrics for Item One.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	51.238901	49.156190
1	RMSE	59.319073	59.035779
2	R2	-0.140071	-0.129208
3	Accuracy	-14.007124	-12.920779

```
Total Actual Sales: 8732.00
Total Predicted Sales (No Damping): 8904.36
Total Predicted Sales (With Damping): 8794.24
```

Figure 4. 58. ES-PM Sales & Metrics for Item Two.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	49.899657	49.191046
1	RMSE	57.435192	56.712150
2	R2	-0.996432	-0.946482
3	Accuracy	-99.643162	-94.648249

```
Total Actual Sales: 11113.00
Total Predicted Sales (No Damping): 11547.26
Total Predicted Sales (With Damping): 11528.36
```

Figure 4. 59. ES-PM Sales & Metrics for Item Three.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	74.804405	71.640678
1	RMSE	86.703808	81.855364
2	R2	-0.291982	-0.151528
3	Accuracy	-29.198226	-15.152798

```
Total Actual Sales: 12794.00
Total Predicted Sales (No Damping): 13352.61
Total Predicted Sales (With Damping): 13107.39
```

Figure 4. 60. ES-PM Sales & Metrics for Item Four.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	61.014513	59.848144
1	RMSE	71.763426	71.794554
2	R2	0.215409	0.214728
3	Accuracy	21.540872	21.472793

```
Total Actual Sales: 14390.00
Total Predicted Sales (No Damping): 14115.42
Total Predicted Sales (With Damping): 14086.00
```

Figure 4. 61. ES-PM Sales & Metrics for Item Five.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	90.251188	77.838703
1	RMSE	116.939997	109.297916
2	R2	-0.287663	-0.124863
3	Accuracy	-28.766279	-12.486331

```
Total Actual Sales: 15316.00
Total Predicted Sales (No Damping): 16309.31
Total Predicted Sales (With Damping): 15765.82
```

Figure 4. 62. ES-PM Sales & Metrics for Item Six.

	Metric	Value (No Damping)	Value (With Damping)
0	MAE	187.913858	186.980323
1	RMSE	222.455342	221.017739
2	R2	0.120454	0.131785
3	Accuracy	12.045411	13.178539

```
Total Actual Sales: 72396.00
Total Predicted Sales (No Damping): 73130.77
Total Predicted Sales (With Damping): 73038.94
```

Figure 4. 63. ES-PM Sales & Metrics for Store Items.

Summary:

the journey of refining sales forecasting models unfolded through strategic Model Modification and Dataset Manipulation. The Model Modification phase witnessed deliberate updates and alterations to simplify, enhance efficiency, and improve overall effectiveness. On the other front, Dataset Manipulation involves the meticulous division of the dataset into individual files and nuanced changes in sales date intervals. This intentional manipulation allowed for a profound exploration of how these alterations influenced the models' responses and, subsequently, the forecast outcomes. The results of each model encapsulated in plots, serve as insightful visual narratives, explaining the intricate dynamics of their performance. This dual-pronged approach, comprising model refinement and dataset manipulation, forms the crux of a comprehensive strategy to elevate the precision and adaptability of sales forecasting endeavors.

Chapter Five

Results

This section marks the culmination of an intricate journey through sales forecasting, where the efficacy and nuances of three prominent models are laid bare. Within these findings, a comprehensive comparison table meticulously dissects the performance of each model, offering a visual roadmap for understanding their strengths and limitations. Beyond mere numbers, the exploration delves into the metrics employed in each model, providing a deeper understanding of their predictive capabilities. Furthermore, this section ventures into the realm of innovation, hinting at the technological advancements poised to shape the next project. As we unravel the outcomes and insights, the Results section serves as a critical juncture, unveiling the transformative power of data-driven decision-making in the realm of sales forecasting.

In this chapter, we will be looking at:

1. **Model Evaluation Metrics: linear regression.**
2. **Model Evaluation Metrics: ARIMA.**
3. **Model Evaluation Metrics: Holt-Winters triple exponential smoothing.**
4. **Comparison table.**
5. **What is next?**
6. **References.**

Model Evaluation Metrics: linear regression:

In the pursuit of effective store sales forecasting, the evaluation of model performance is a pivotal step. This entails a thorough analysis of key metrics to gauge the accuracy, precision, and overall predictive capabilities of the employed model. The following section provides a comprehensive assessment using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). Through these metrics, we aim to unravel the nuanced aspects of the Linear Regression model's ability to forecast sales accurately.

Let's break down the key metrics used:

I. Mean Absolute Error (MAE):

- **Definition:** The average absolute difference between the predicted and actual values.
- **Calculation:** $MAE = n1\sum i = 1n \mid y\text{true}, i - y\text{pred}, i \mid$
- **Interpretation:** The MAE provides the average magnitude of errors, giving an idea of the model's overall accuracy.

II. Root Mean Squared Error (RMSE):

- **Definition:** The square root of the average squared differences between predicted and actual values.
- **Calculation:** $RMSE = n1\sum i=1n(y\text{true},i - y\text{pred},i)2$
- **Interpretation:** RMSE penalizes larger errors more heavily, providing a sense of the model's precision.

III. Mean Absolute Percentage Error (MAPE):

- **Definition:** The average percentage difference between predicted and actual values.
- **Calculation:** $MAPE = n1\sum i = 1n \mid\mid y_{\text{true}, i} y_{\text{true}, i} - y_{\text{pred}, i} \mid\mid \times 100$
- **Interpretation:** MAPE expresses errors as a percentage of the actual values, making it interpretable across different scales.

IV. R-squared (R^2) / Accuracy:

- **Definition:** A measure of how well the model's predictions match the actual data.
- **Calculation:** R^2 ranges from 0 to 1, where 1 indicates a perfect fit.
- **Interpretation:** R^2 represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

These metrics collectively provide a comprehensive evaluation of the Linear Regression model's performance in forecasting sales. The subsequent visualizations and plots further enhance the understanding of the model's predictive capabilities and highlight areas of improvement. The code concludes with a display of evaluation metrics and a graphical representation of the actual vs. predicted sales, providing a holistic view of the model's effectiveness.

Model Evaluation Metrics: ARIMA:

As we delve into the realm of sales forecasting, the evaluation of forecasting models stands as a crucial phase. This section meticulously examines the performance of the ARIMA (Autoregressive Integrated Moving Average) model in predicting sales. Key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), Adjusted R-squared (Adjusted R^2), and overall accuracy will be scrutinized to unveil the ARIMA model's efficacy. Additionally, a visual representation of the model's forecasting prowess is presented through comparative plots of actual sales against ARIMA-predicted sales.

Let's break down the key metrics used:

I. Mean Absolute Error (MAE):

- **Definition:** MAE represents the average absolute difference between the predicted and actual values.
- **Formula:** $MAE = n1\sum i = 1n \mid y_{\text{true}, i} - y_{\text{pred}, i} \mid$
- **Interpretation:** A lower MAE indicates better accuracy.

II. Root Mean Squared Error (RMSE):

- **Definition:** RMSE is the square root of the average of squared differences between predicted and actual values.
- **Formula:** $RMSE = n1\sum i=1n(y_{\text{true},i} - y_{\text{pred},i})2$
- **Interpretation:** RMSE gives more weight to large errors and is sensitive to outliers.

III. R-squared (R^2):

- **Definition:** R^2 measures the proportion of the variance in the dependent variable explained by the independent variables.
- **Formula:** $R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{\text{true},i} - y_{\text{pred},i})^2}{\sum_{i=1}^{n} (y_{\text{true},i} - \bar{y}_{\text{true}})^2}$
- **Interpretation:** R^2 ranges from 0 to 1, where 1 indicates a perfect fit.

IV. Adjusted R-squared (Adjusted R^2):

- **Definition:** Adjusted R^2 considers the number of predictors in the model, adjusting for the complexity.
- **Formula:** $Adjusted R^2 = 1 - \frac{(1 - R^2) \times (n-1)}{n-k-1}$
- **Interpretation:** Adjusted R^2 penalizes models with unnecessary predictors.

V. Accuracy:

- **Definition:** In this context, accuracy is a measure of how well the model predicts the sales, calculated as a percentage.
- **Formula:** $Accuracy = R^2 \times 100$
- **Interpretation:** Higher accuracy indicates a better overall fit.

Model Evaluation Metrics: Holt-Winters triple exponential smoothing:

These metrics collectively provide a comprehensive assessment of the ARIMA model's performance in predicting sales. They offer insights into different aspects of model accuracy, error, and explanatory power.

In the pursuit of enhancing sales forecasting accuracy, the application of sophisticated forecasting models becomes imperative. This analysis focuses on the evaluation of the Holt-Winters forecasting model under two distinct configurations: one without damping and the other with damping. By scrutinizing metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), and Accuracy, we aim to assess the model's predictive prowess under varying conditions. Additionally, the total actual sales and predicted sales provide a holistic view of the model's performance, offering insights into its efficacy in real-world sales forecasting scenarios.

Let's break down the key metrics used:

I. Holt-Winters Model with No Damping (pred_fit_1):

I. **Mean Absolute Error (MAE):**
   - i. **Definition:** Average absolute difference between predicted and actual values.
   - **ii. Formula:** $MAE = \frac{1}{n} \sum_{i=1}^{n} |y_{\text{true},i} - y_{\text{pred},i}|$

II. **Root Mean Squared Error (RMSE):**

i. **Definition:** Square root of the average of squared differences between predicted and actual values.

ii. **Formula:** $RMSE=\frac{1}{n}\sum_{i=1}^{n}\left(y_{\text{true},i}-y_{\text{pred},i}\right)^2$

II. R-squared (R^2):

i. **Definition:** Proportion of the variance in the dependent variable explained by the independent variables.

ii. **Formula:** $R^2=1-\frac{\sum_{i=1}^{n}\left(y_{\text{true},i}-\bar{y}_{\text{true}}\right)^2}{\sum_{i=1}^{n}\left(y_{\text{true},i}-y_{\text{pred},i}\right)^2}$

III. Accuracy:

i. **Definition:** Percentage of accuracy calculated from R-squared.

ii. **Formula:** $Adjusted R^2=1-\frac{n-k-1}{(1-R^2)\times(n-1)}$

IV. Holt-Winters Model with Damping (pred_fit_2):

Same metrics as above (MAE, RMSE, R^2, Accuracy), but for the Holt-Winters model with damping.

V. Total Actual Sales and Predicted Sales:

I. **Total Actual Sales:** The sum of actual sales from the test data.

II. **Total Predicted Sales (No Damping):** The sum of predicted sales for the Holt-Winters model with no damping.

III. **Total Predicted Sales (With Damping):** The sum of predicted sales for the Holt-Winters model with damping.

All these metrics provide a comprehensive evaluation of the performance of the Holt-Winters forecasting model under different conditions, allowing for a detailed comparison of the two scenarios (with and without damping).

Comparison table:

item	Linear regression	Arima	Exponential smoothing
One	▪ **Total Actual Sales:** 10051 ▪ **Total Predicted Sales:** 10169	▪ **Total Actual Sales:** 10051 ▪ **Total Predicted Sales:** 8372	▪ **Total Actual Sales:** 10051 ▪ **Total Predicted Sales (No Damping):** 9755 ▪ **Total Predicted Sales (With Damping):** 10138
Two	▪ **Total Actual Sales:** 8732 ▪ **Total Predicted Sales:** 8664	▪ **Total Actual Sales:** 8732 ▪ **Total Predicted Sales:** 10073	▪ **Total Actual Sales:** 8732 ▪ **Total Predicted Sales (No Damping):** 8904 ▪ **Total Predicted Sales (With Damping):** 8794
Three	▪ **Total Actual Sales:** 11113	▪ **Total Actual Sales:** 11113 ▪ **Total Predicted Sales:**	▪ **Total Actual Sales:** 11113 ▪ **Total Predicted Sales (No Damping):** 11547

	▪ **Total Predicted Sales:** 11574	15716	▪ **Total Predicted Sales (With Damping):** 11528
Four	▪ **Total Actual Sales:** 12794 ▪ **Total Predicted Sales:** 13067	▪ **Total Actual Sales:** 12794 ▪ **Total Predicted Sales:** 11792	▪ **Total Actual Sales:** 12794 ▪ **Total Predicted Sales (No Damping):** 13352 ▪ **Total Predicted Sales (With Damping):** 13107
Five	▪ **Total Actual Sales:** 14390 ▪ **Total Predicted Sales:** 14165	▪ **Total Actual Sales:** 14390 ▪ **Total Predicted Sales:** 10912	▪ **Total Actual Sales:** 14390 ▪ **Total Predicted Sales (No Damping):** 14115 ▪ **Total Predicted Sales (With Damping):** 14086
six	▪ **Total Actual Sales:** 15316 ▪ **Total Predicted Sales:** 15603	▪ **Total Actual Sales:** 15316 ▪ **Total Predicted Sales:** 17463	▪ **Total Actual Sales:** 15316 ▪ **Total Predicted Sales (No Damping):** 16309 ▪ **Total Predicted Sales (With Damping):** 15765
Store items	▪ **Total Actual Sales:** 72396 ▪ **Total Predicted Sales:** 73245	▪ **Total Actual Sales:** 72396 ▪ **Total Predicted Sales:** 72839	▪ **Total Actual Sales:** 72396 ▪ **Total Predicted Sales (No Damping):** 73130 ▪ **Total Predicted Sales (With Damping):** 73038

Table 5. 1. Comparing Models Results.

From the provided table showcasing the results of three different forecasting models (Linear Regression, ARIMA, Exponential Smoothing) across various store items, several key observations can be made:

I. Item-Level Analysis:

I. Item One:

i. **Linear Regression:** Total Predicted Sales (10169) exceeded actual sales.
ii. **ARIMA:** Total Predicted Sales (8372) fell short of actual sales.
iii. **Exponential Smoothing:** Total Predicted Sales (No Damping: 9755, With Damping: 10138) exhibit variations, with damping contributing to a slightly higher prediction.

II. Item Two:

i. **Linear Regression:** Total Predicted Sales (8664) are close to actual sales.
ii. **ARIMA:** Total Predicted Sales (10073) deviate significantly from actual sales.
iii. **Exponential Smoothing:** Both damping scenarios (No Damping: 8904, With Damping: 8794) show differences in predictions, with No Damping being slightly closer to actual sales.

III. Item Three:

i. **Linear Regression:** Total Predicted Sales (11574) exceed actual sales.
ii. **ARIMA:** Total Predicted Sales (15716) deviate significantly from actual sales.

iii. **Exponential Smoothing:** Both damping scenarios (No Damping: 11547, With Damping: 11528) demonstrate variations in predictions, with Damping being closer to actual sales.

IV. **Item Four:**

i. **Linear Regression:** Total Predicted Sales (13067) are close to actual sales.

ii. **ARIMA:** Total Predicted Sales (11792) deviate significantly from actual sales.

iii. **Exponential Smoothing:** Both damping scenarios (No Damping: 13352, With Damping: 13107) exhibit variations, with No Damping being slightly closer to actual sales.

V. **Item Five:**

i. **Linear Regression:** Total Predicted Sales (14165) are close to actual sales.

ii. **ARIMA:** Total Predicted Sales (10912) deviate significantly from actual sales.

iii. **Exponential Smoothing:** Both damping scenarios (No Damping: 14115, With Damping: 14086) show differences in predictions, with No Damping being slightly closer to actual sales.

VI. **Item Six:**

i. **Linear Regression:** Total Predicted Sales (15603) exceed actual sales.

ii. **ARIMA:** Total Predicted Sales (17463) deviate significantly from actual sales.

iii. **Exponential Smoothing:** Both damping scenarios (No Damping: 16309, With Damping: 15765) demonstrate variations, with No Damping being slightly closer to actual sales.

II. Store-Level Analysis:

I. Across all items, Linear Regression and Exponential Smoothing tend to have predictions closer to actual sales, while ARIMA shows larger variations.

II. The total actual sales and total predicted sales for each model are summarized at the store level.

These observations highlight the importance of selecting the appropriate forecasting model, as different models exhibit varying levels of accuracy in predicting sales for different items.

Item	**Linear regression**	**ARIMA**	**Exponential Smoothing**
One	10169/10051 * 100 = ~101.06%	8372/10051 * 100 = ~83.31%	No dumping: 9755/10051 * 100 = ~97.01% With dumping: 10138/10051 * 100 = ~108.65%
Two	8664/8732 * 100 = ~99.22%	10073/8732 * 100 = ~115.49%	No dumping: 8904/8732 * 100 = ~101.97% With dumping: 7894/8732 * 100 = ~100.71%
Three	11574/11113 * 100 = ~104.56%	15716/11113 * 100 = ~141.46%	No dumping: 11547/11113 * 100 = ~103.89% With dumping:11528 /11113 * 100 = ~103.34%
Four	13067/12794 * 100 = ~101.34%	11792/12794 * 100 = ~92.29%	No dumping: 13352/12794 * 100 = ~104.38%

			With dumping: 13107/12794 * 100 = ~102.42%
Five	14165/14390 * 100 = ~98.44%	10912/14390 * 100 = ~75.83%	No dumping: 14115/14390 * 100 = ~98.08% With dumping: 14086/14390 * 100 = ~97.89%
six	15603/15316 * 100 = ~101.88%	17463/15316 * 100 = ~113.18%	No dumping: 16309/15316 * 100 = ~106.47% With dumping: 15765/15316 * 100 = ~102.95%

Table 5. 2. Comparing Models Results by Percentages.

Here are some observations based on the provided percentage accuracy calculations:

I. **Linear Regression:**
   - Achieves relatively consistent performance across items.
   - Percentages vary around 100%, indicating a balanced performance.

II. **ARIMA:**
   - Shows mixed performance across items.
   - Item Three has a notably high percentage, but others vary.

III. **Exponential Smoothing:**
   - **No Damping:**
     i. Achieves competitive results across items.
     ii. Offers a good balance between over and underestimation.
   - **With Damping:**
     i. Slightly more conservative predictions.
     ii. Achieves balanced results, though some percentages are below 100%.

Considering these observations, the **Exponential Smoothing model without damping** appears to be a robust performer, providing accurate forecasts across multiple items. However, the choice of the best model should also consider other factors like model complexity, interpretability, and the specific requirements of your forecasting application. It might be beneficial to explore ensemble methods or hybrid approaches for further improvements.

What is next?

As we conclude our current exploration of sales forecasting models, our focus shifts towards a promising horizon: the implementation of deep learning techniques. This next semester, we aim to leverage deep learning's capabilities to enhance our sales forecasting accuracy, exploring models like CNNs and RNNs. We plan to integrate these models using tools such as TensorFlow and Kera's, comparing their performance against our existing strategies. This comparison will focus on metrics like accuracy, MAE, and RMSE.

We also intend to refine our current models by incorporating deep learning elements. This enhancement will consider factors that influence forecasting, including seasonal trends and promotional impacts. By doing so, we expect to not only improve our forecasting precision but also gain deeper insights into the dynamics of store sales.

This journey into deep learning represents a step forward in our analytical capabilities, promising more accurate and insightful predictions. It's a venture filled with potential challenges, but one that holds the promise of significant advancements in our forecasting methodology.

This draft encapsulates the intent to implement and compare deep learning models with your existing models, aiming for enhanced forecasting accuracy and including key factors affecting sales predictions.

References:

1. **Applied Linear Regression** by Sanford Weisberg - Linear Regression Metrics.
2. **Building Machine Learning Powered Applications** by Emmanuel Ameisen - Model Building.
3. **Data Analysis with Microsoft Excel** by Kenneth N. Berk and Patrick Carey - Google, Excel, Power BI.
4. **Data Mining: Practical Machine Learning Tools and Techniques** by Ian H. Witten, Eibe Frank, and Mark A. Hall - Dataset Selection.
5. **Data Preparation for Data Mining** by Dorian Pyle - Using and Preparing Datasets for Forecasting.
6. **Data Science for Business** by Foster Provost and Tom Fawcett - Data and Dataset.
7. **Evaluating Machine Learning Models** by Alice Zheng - Model Evaluation Metrics.
8. **Forecasting in Retail and FMCG** by Paul Goodwin - Store Sales Forecasting.
9. **Forecasting: Principles and Practice** by Rob J Hyndman and George Athanasopoulos - Holt-Winters Method.
10. **Introduction to Time Series Analysis and Forecasting** by Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci - Time Series Analysis.
11. **Machine Learning for Time Series Forecasting with Python** by Francesca Lazzeri - Machine Learning.
12. **Pattern Recognition and Machine Learning** by Christopher M. Bishop - Model Modification.
13. **Python Data Science Handbook** by Jake VanderPlas - Python, Anaconda, and Jupyter Notebook.
14. **Python Machine Learning** by Sebastian Raschka and Vahid Mirjalili - Machine Learning.
15. **Retail Sales Forecasting at Walmart** from Walmart Corporate - The Aim of Store Sales Forecasting.
16. **Sales Forecasting: A Practical Guide** by Clive Miller - Sales Forecasting Models.
17. **The Data Warehouse Toolkit** by Ralph Kimball and Margy Ross - Data Warehouse.
18. **The Elements of Statistical Learning** by Trevor Hastie, Robert Tibshirani, and Jerome Friedman - Model Selection.
19. **The Visual Display of Quantitative Information** by Edward R. Tufte - Dataset Visualization.
20. **Time Series Analysis: Forecasting and Control** by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung - ARIMA Metrics.
21. **Case studies from the MIT Sloan Management Review** - Insightful case studies on forecasting models.
22. **Demand Forecasting in Retail: Best Practices & Pitfalls by Harvard Business Review** - Pros and Cons of Store Sales Forecasting.
23. **Kaggle's own tutorials and competition case studies** - Kaggle.