



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING FACULTY OF ENGINEERING

SEMESTER 2 2022/2023

MCS D 1143 SUPPLY CHAIN ANALYTIC

PROJECT REPORT

Superstore Demand Forecasting

No.	Name	Matric No.
1.	Hannah Najwa binti Yahaya	MEC 211007
2.	Mehdi Mouaiz Eddine Smail	MCS221010
3.	Ahmed Hashim Taha Salim	MCS211041

PROJECT BACKGROUND

Supply chain management plays a critical role in today's business environment, enabling companies to effectively plan, source, produce, and deliver goods and services to customers. With increasing market dynamics and customer expectations, organizations are constantly seeking ways to optimize their supply chain operations. One of the key factors in achieving this optimization is accurate demand forecasting and efficient inventory management.

Accurate demand forecasting allows organizations to align their production and procurement activities with the anticipated customer demand, reducing stockouts, minimizing excess inventory, and ensuring timely fulfillment of orders. It helps in streamlining the supply chain, optimizing resource allocation, and improving overall operational efficiency. On the other hand, inventory management aims to strike a balance between meeting customer demands and minimizing inventory holding costs. By monitoring inventory levels, turnover rates, and lead times, organizations can avoid stockouts, reduce carrying costs, and optimize their working capital.

To achieve effective demand forecasting and inventory management, organizations are increasingly turning to advanced analytics techniques. By leveraging historical sales data, organizations can analyze patterns, trends, and seasonality to make accurate predictions about future demand. This allows them to plan production, adjust inventory levels, and optimize the supply chain network.

The dataset provided for this project, the "Superstore Sales" dataset by Rohit Sahoo, is a valuable resource for conducting supply chain analytics. It contains historical sales data for a retail company, encompassing various product categories, store locations, and sales volumes over time. This dataset serves as a rich source of information to develop demand forecasting models and perform inventory analysis, for more details refer to the Data Source section.

By utilizing this dataset, we aim to apply analytics techniques to generate insights and forecasts that will empower organizations to make informed supply chain decisions. The project will focus on analyzing the dataset to develop accurate demand forecasts, understand inventory dynamics, and identify opportunities for improvement within the supply chain. Additionally, descriptive analytics will provide a comprehensive understanding of the dataset, enabling us to explore its structure, uncover patterns, and gain insights into customer behavior.

Through this project we aim to leverage the "Superstore Sales" dataset to perform demand forecasting, and descriptive analytics on its categories. The results obtained from this analysis will guide organizations in making strategic and tactical decisions to optimize their supply chain operations, reduce costs, and maximize profitability.

OBJECTIVES

As a sales manager of ABC Superstore, it is important to forecast the future sales of the products so that a proper estimation of inventory level and demand planning can be done. Demand forecasting is important in making strategic and tactical decisions to optimize the company operations, reduce costs and maximize profitability. In order to achieve the company's goals, the following objectives are being carried out:

1. To provide an analysis on the historical time series dataset by examining the order date and sales data, such as observing the trend and seasonality in the datasets.
2. To perform a demand planning analytics using static method and adaptive method for time series forecasting.
3. To choose the best forecasting method for the dataset by comparing the performance error on various approaches.

DATA SOURCE

The Superstore Sales Dataset is obtained from Kaggle, an open-source platform that provides a large number of datasets and problems for users to work on, and it also provides tools for users to submit and evaluate their models.

1. Data collection

The Superstore Sales Dataset is a retail dataset from a global superstore spanning four years. It has a total of 9800 records of sales-related data with 18 attributes related to products, customers and location, which can be described in Table 1 below:

Table 1: Description of Superstore data

Columns	Descriptions
Row ID	A unique identifier for each row in the dataset
Order ID	A unique identifier for each order placed
Order Date	The date when the order was placed
Ship Date	The date when the order was shipped
Ship Mode	The shipping mode used for the order (e.g., Standard Class, Second Class).
Customer ID	A unique identifier for each customer.
Customer Name	The name of the customer who placed the order
Segment	The customer segment to which the customer belongs (e.g., consumer, corporate).
Country	The country where the order was placed
City	The city where the order was placed
State	The state where the order was placed
Postal Code	The postal code of the order's location
Region	The geographical region of the order's location
Product ID	A unique identifier for each product

Category	The category to which the product belongs (e.g., office supplies, furniture and technology)
Sub_Category	The sub-category to which the product belongs (e.g., chairs, paper).
Product Name	The name of the product.
Sales	The sales value of the order.

2. Data Preprocessing

Data preprocessing is a crucial step in any data analysis project. In this step, the data is cleaned, transformed, and organized to ensure its quality and suitability for further analysis. The following tasks have been performed as part of the data preprocessing step:

1. Removing Unwanted Columns:

Unwanted columns that are not relevant to the analysis have been removed based on the project objectives and analytics methods identified. For example, columns like "Order ID" or "Customer Name" may not provide significant insights for the specific analysis tasks such as demand forecasting, inventory analysis, or network consolidation.

2. Handling Missing Values

A check was performed for missing values in the dataset. If any columns contain missing values.

3. Organizing the dataset

To facilitate time-series analysis, the "Order Date" column was set as the index of the dataset. This indexing enables easy access and analysis of time-series data, which is often important for tasks such as forecasting or trend analysis.

3. Exploratory Data Analysis

The dataset is large and includes various characteristics that can be used to analyze different attributes such as categories, subcategories, regions, countries, and customer segments. However, we have identified that dividing the dataset based on product categories would be beneficial. Thus,

the dataset was split based on three categories: office supplies, furniture, and technology. This categorization allows for easier analysis and comparison within specific product categories while integrating the sales data from the same date together. By looking at the Figure 1 below we can see the average monthly sales for the three categories in detail.

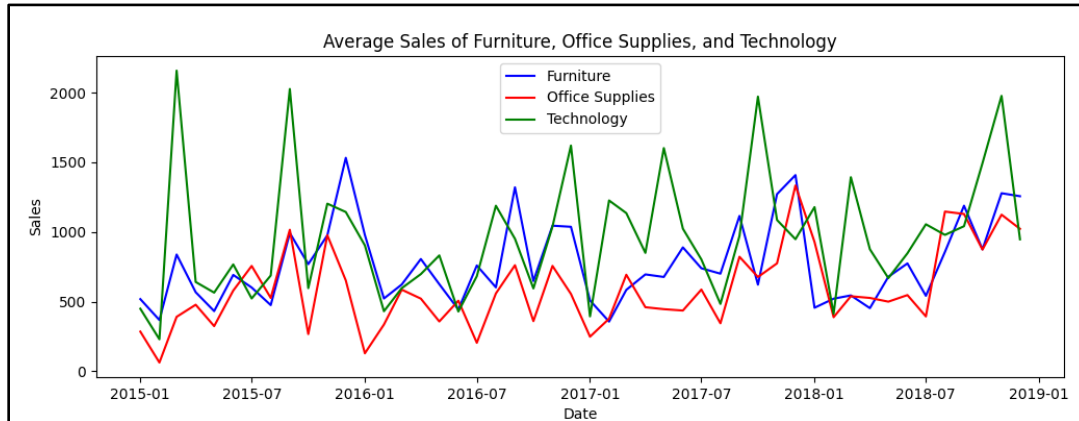


Figure 1: Average Sales for Furniture, Office Supplies and technology

The following Figures 2,3 and 4 display the collected sales data over a four-year period, represented on a daily, monthly, quarterly basis. The curve in the figure demonstrates variations in the sales trend, showcasing instances where the product sales increase or decrease. The different time intervals provide a comprehensive view of how the sales fluctuate over time.

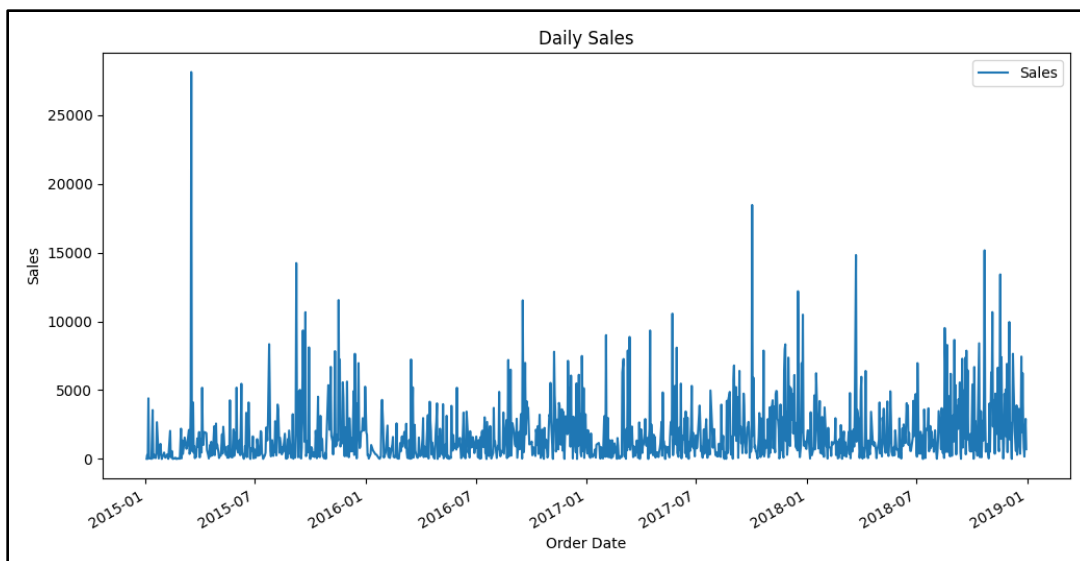


Figure 2: Average daily Superstore Sales

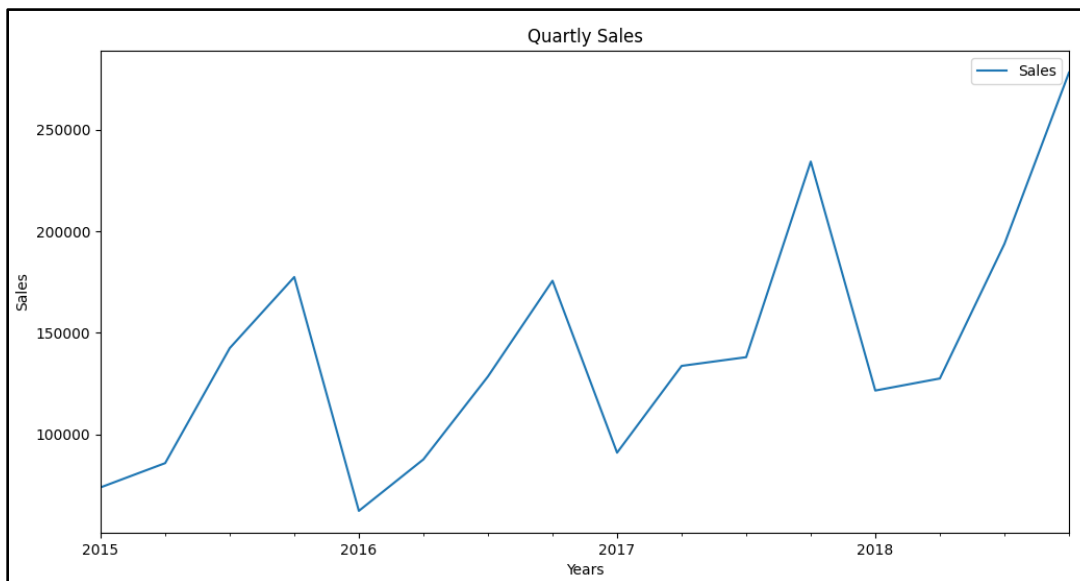


Figure 3: Average Quarter Superstore Sales

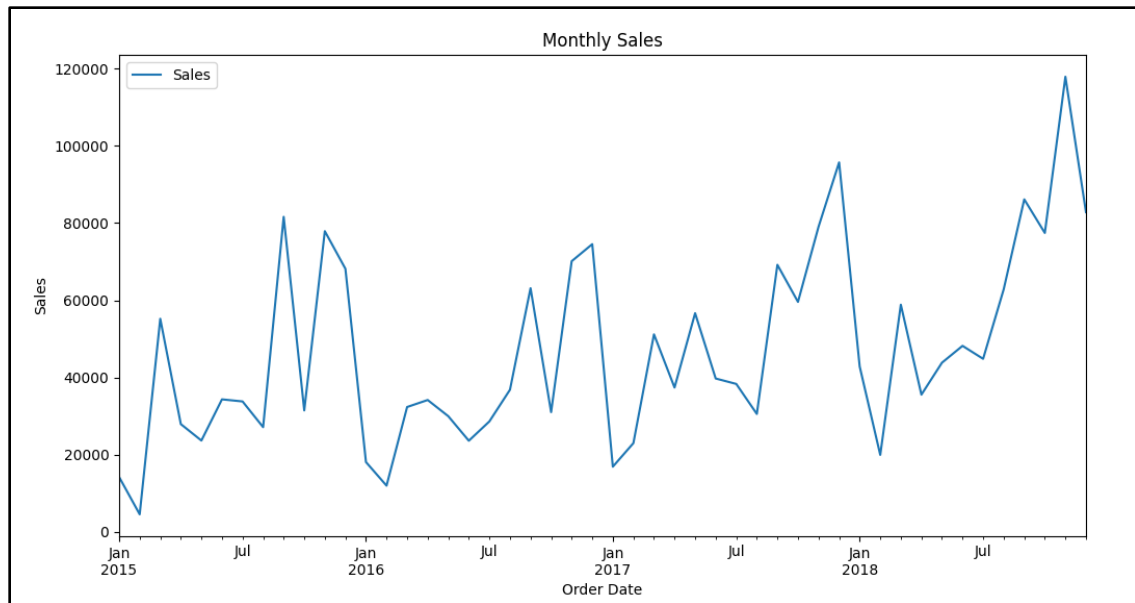


Figure 4: Monthly Superstore Sales

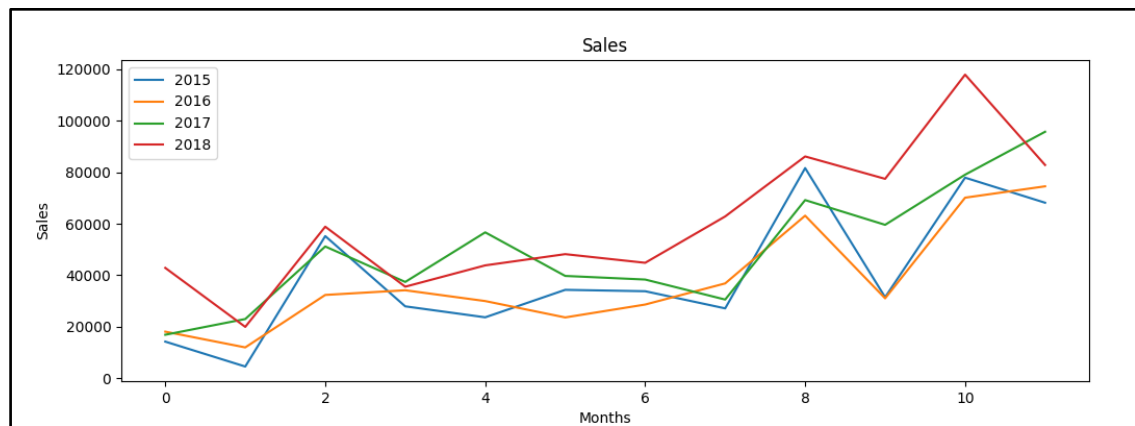


Figure 5: Each year Superstore sales

ANALYTICS METHOD

Time series forecasting is a well-known problem in data science and statistics. We will analyze and forecast the time series data of the superstore dataset specifically for demand forecasting. In our case, we consider the Sales values as the demand observed over different time periods. The Figure 6 below illustrates the Flow Diagram for Time Series Forecasting in this project.

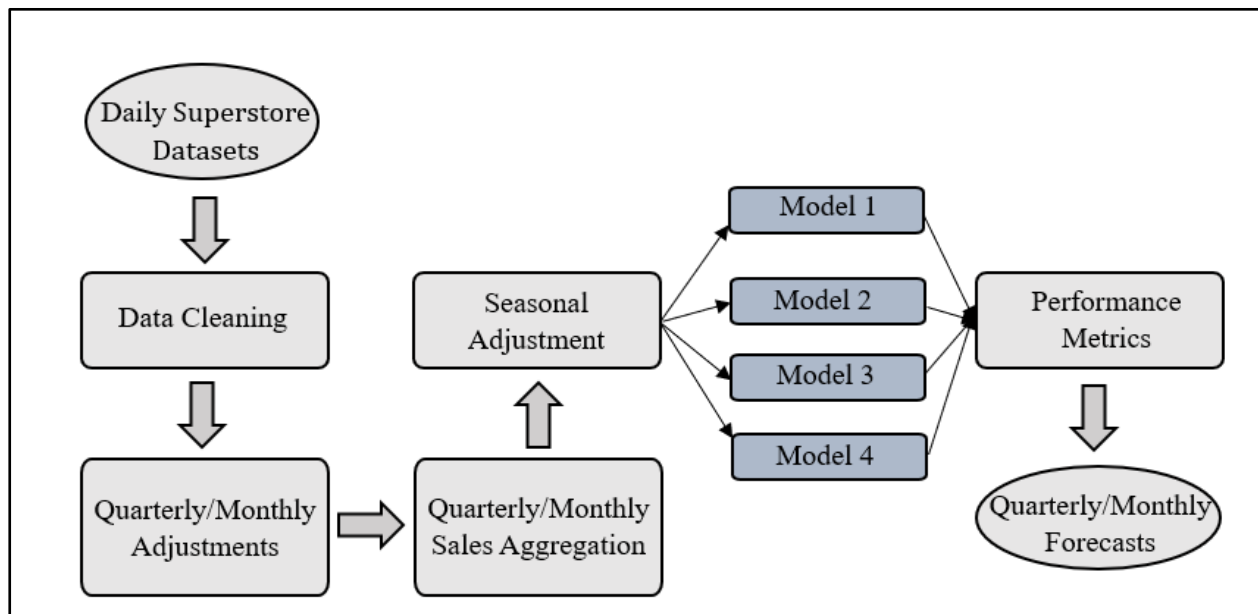


Figure 6: Flow Diagram for Time Series Forecasting

While it is possible to perform time series analysis in Excel, we will explore various methods using Python for our analysis and forecasting purposes.

Demand Forecasting

Demand forecasting helps organizations by providing managers valuable insights for their businesses, allowing them to make informed decisions regarding pricing, growth plans, and market potential. Demand forecasting is crucial for sales planning because it helps businesses foresee and get prepared for future demand for their goods and services. By accurately forecasting demand, businesses may more effectively plan their production schedules and manage inventory levels, which will save storage costs and optimize the use of resources.

Accurate demand forecasting is also essential for financial planning and budgeting since it enables companies to estimate their prospective income and profit, which enables them to make wise decisions about their pricing strategies and expenditure plans. Demand forecasting can also offer useful insights for the sales and marketing teams, as it can assist those in charge in establishing

realistic sales goals, creating a successful marketing campaign, and identifying market segments with the most potential for growth.

From Superstore dataset, demand forecasting can be done by analyzing historical data to forecast future sales patterns while assessing whether seasonality and trends are present. By examining the order date and sales data, the seasonality of a sales demand can be determined, which is helpful in optimizing production, inventory and marketing strategies accordingly. Furthermore, the dataset also allows users to recognize the long-term trends in the demand, allowing them to forecast the future demand levels and make strategic decisions. Lastly, segmenting the dataset based on factors such as product categories enables users to gain insight into different demand patterns. It assists in tailoring the forecasting and planning strategies into product categories, allowing a more accurate prediction and targeted decision making.

Static Method

In this section, our main objective is to analyze and forecast the average monthly sales for three distinct business categories: Office Supplies, Furniture, and Technologies. Each category represents a significant segment of our business, and understanding their sales patterns is crucial for strategic decision-making. By employing a static method approach, we will delve into historical sales data, identify trends, and make informed predictions for each category individually. Through this analysis, we aim to gain insights into the performance of each category and provide valuable information for future business planning.

Methodology:

1. Data Collection and Preprocessing:

- We have gathered comprehensive sales data for each category, ensuring a substantial time span to capture relevant patterns and trends.
- The collected data has been preprocessed, including cleaning, handling missing values, and organizing it in chronological order.
- Essential columns, such as order dates and sales figures, have been extracted for further analysis.

2. Office Supplies Category Analysis:

- Given that the Office Supplies category leads in terms of sales, we will begin our analysis by focusing on this category.
- We will examine the historical sales trends, seasonal patterns, and overall performance of the Office Supplies category.

3. Furniture Category Analysis:

- Moving on to the Furniture category, we will conduct a detailed analysis of its sales data.
 - We will explore its sales patterns, identify any potential seasonality or trends, and gain insights into its performance within the market.
4. Technologies Category Analysis:
 - Similarly, we will analyze the sales data for the Technologies category.
 - By studying the historical sales trends and patterns specific to this category, we will uncover valuable information about its performance and market dynamics.
 5. Average Quarterly Sales Calculation:
 - For each category, we will calculate the average Monthly sales to gain a high-level understanding of their performance over time.
 - By resampling the data on a quarterly basis, we can identify the average sales for each quarter and assess any seasonal or long-term trends.
 6. Visualization and Insights:
 - Throughout the analysis, we will create visualizations to showcase the sales trends and patterns for each category.
 - These visual representations will provide valuable insights into historical performance, seasonal fluctuations, and potential growth opportunities.
 7. Forecasting:
 - Using the static method approach, we will employ suitable forecasting techniques for each category individually.
 - Time series analysis, regression models, or other appropriate methods will be utilized to forecast the future average quarterly sales for each category.
 8. Evaluation and Comparison:
 - To evaluate the accuracy and reliability of our forecasting models, we will compare the forecasted results with the actual sales data for each category.
 - Evaluation metrics specific to forecasting, such as mean squared error or mean absolute percentage error, will be used to assess the performance of the static method approach.

Through this project, we aim to analyze and forecast the average monthly sales for three significant business categories: Office Supplies, Furniture, and Technologies. By delving into historical sales data, identifying trends, and making predictions using the static method approach, we can gain valuable insights into the performance of each category.

These insights will aid in strategic decision-making, market analysis, and resource allocation within our business. By understanding the unique sales patterns and dynamics of each category, we can optimize our operations, identify growth opportunities, and enhance overall business performance.

Adaptive Method

Adaptive forecasting model is a time series forecasting method that iteratively updates and optimizes forecasts using historical data until new data becomes available. The model is being utilized in demand planning as the model understands that demand patterns can change over time, necessitating an iterative forecasting model to capture the changes.

In the adaptive forecasting model, the initial forecast is generated using historical data and the selected forecasting model, such as moving average, simple exponential smoothing, Holt's Model or Winter's model. The model is being continuously adjusted to integrate the most recent data as new data points are observed.

Adaptive forecasting models have advantages against static methods in a way the model is iterative and the ability to capture trends and seasonalities in the data. Adaptive forecasting models enable businesses to capture the demand pattern, in this case, the sales pattern whether there is trend and seasonality occurs, allowing for flexibility and responsiveness in demand planning. The model can offer a more accurate forecast for efficient demand planning and be beneficial for relevant departments in aligning their goals and strategies for the business. In this report, the adaptive forecasting model methodology are as follows:

1. Data Preprocessing

- The monthly and quarterly average sales dataset extracted is being split into training and testing datasets.
- For quarterly average sales, *quartly_sales.shape* indicates that the size of the dataset is 16. Hence, a dataset with the size of 12 is being split into a training set and a dataset with the size of 4 is being split into the testing set. This indicated that the forecasting will be for 4 quarterly periods which also can indicate one year forecasting.
- For monthly average sales, *monthly_sales.shape* indicates that the size of the dataset is 48, so the dataset is being split into 42 months for the training set and 6 months for testing set, which indicates the forecasting.

2. Forecasting & analysis using various adaptive forecasting model methods

- The split datasets are now being used to forecast using the Simple Moving Average method for monthly forecasting and a graph is plotted to observe the pattern with the training set, the testing set and the simple moving average forecast.
- The forecasting error, RMSE and MAPE are being calculated and plotted into a dataframe.
- The same step is being repeated for quarterly sales average forecasting.
- The three steps aforementioned are now being repeated with Simple Exponential Smoothing method, Holt's Model, which only accounts for the trend and Winter's Model with Multiplication, which accounts for the trend and seasonality in the dataset.

3. Forecasting product average monthly sales

- From the forecasting error, the best method for forecasting is being selected to forecast the product monthly sales average according to its categories.
- The first product category, *y_furniture* datasets, is being split into training and testing sets where 42 datasets are for training sets and 6 datasets are for testing sets.
- Then, we can forecast using Winter's Model with Multiplication method and a graph is plotted to observe the pattern in the product sales.
- The forecasting error, RMSE and MAPE can be calculated and plotted into a dataframe.
- The same steps are being repeated for the remaining product categories and now we have the 6 months forecast for the products monthly sales average.

Performance Metrics

Three matrices are used to measure the performance accuracy of the proposed models.

- Mean Square Error (MSE) for weekly, monthly, or yearly forecast.
- Root Mean Square Error (RMSE) for the forecast.
- Mean Absolute Percentage Error (MAPE) for quantifying overall accuracy.

RESULTS & DISCUSSION

Static Method

1. Office Supplies Category

The analysis for the Office Supplies category was performed using the static method approach with OLS (Ordinary Least Squares) regression. The model obtained a score of 0.597025250632492, indicating a moderate level of fit between the predicted values and the actual sales data. However, the R-squared value of 0.023 suggests that only a small portion (2.3%) of the variance in the average Monthly sales can be explained by the model.

Upon examining the OLS regression results, the coefficient for the "Deseasonalized_data" variable is 0.0081, with a p-value of 0.304. The coefficient signifies the change in the average quarterly sales for every unit change in the deseasonalized data. However, the p-value suggests that this coefficient is not statistically significant at the conventional significance level ($\alpha = 0.05$). Hence, the model's ability to accurately predict the average monthly sales based on the deseasonalized data is limited.

The evaluation of the model's performance includes the analysis of the final table shown in Figure 7 below, which provides the seasonal factor (SF) for each period. The SF represents the ratio of the observed sales data to the deseasonalized data, indicating the seasonal variation present in the Office Supplies category. The SF values range from 0.150231 to 2.176449, suggesting varying degrees of seasonal impact across different periods.

Out[422]:

	Period_t	Sales	Deseasonalized_data	Dt	SF
0	1	285.008235	0.000000	412.920364	0.690226
1	2	63.042588	0.000000	419.638740	0.150231
2	3	391.020682	0.000000	426.357116	0.917120
3	4	477.777130	0.000000	433.075491	1.103219
4	5	324.165455	0.000000	439.793867	0.737085
5	6	579.199500	0.000000	446.512243	1.297164
6	7	756.060400	519.679569	453.230619	1.668158
7	8	526.441714	524.540066	459.948995	1.144565
8	9	1015.677704	544.007240	466.667370	2.176449
9	10	267.078815	553.946790	473.385746	0.564189
10	11	976.414667	557.167472	480.104122	2.033756

Figure 7: Office supply seasonal factor table

When interpreting the SF values, periods with SFs less than 1 indicate lower sales than the deseasonalized average, whereas SFs greater than 1 indicate higher sales. In the Office Supplies

category, periods 2, 10, 12, 18, 30, 37, 38, 39, 40, and 42 have SFs less than 1, indicating lower sales compared to the average. On the other hand, periods 9, 11, 16, 17, 20, 21, 26, 32, 33, 34, 35, 36, 43, 44, 45, 46, and 48 have SFs greater than 1, suggesting higher sales than the average.

Overall, the analysis of the Office Supplies category reveals that the static method approach with OLS regression has limitations in accurately predicting the average quarterly sales. The R-squared value indicates that the model does not capture a significant portion of the sales variation, and the coefficient for the deseasonalized data is not statistically significant. Additionally, the varying SF values highlight the presence of seasonal patterns, with some periods exhibiting lower or higher sales compared to the deseasonalized average.

To improve the accuracy of forecasting for the Office Supplies category, alternative forecasting techniques or models can be explored.

2. Furniture Category

The analysis for the Furniture category was performed using the static method approach with OLS (Ordinary Least Squares) regression. However, the model obtained a low score of 0.0005782344925936433, indicating a weak fit between the predicted values and the actual sales data. The R-squared value of 0.000 suggests that the model explains very little (0%) of the variance in the dependent variable, which is the Period_t.

Upon examining the OLS regression results in Figure 8 Below, the coefficient for the "Deseasonalized_data" variable is 4.313e-05, with a p-value of 0.994. The coefficient represents the change in the dependent variable for every unit change in the deseasonalized data. However, the p-value indicates that this coefficient is not statistically significant at the conventional

significance level ($\alpha = 0.05$). This means that there is no significant relationship between the deseasonalized data and the Period_t variable in the Furniture category.

Figure 8: Furniture OLS regression results

The evaluation of the model's performance includes the analysis of the final table, which provides

OLS Regression Results						
Dep. Variable:	Period_t	R-squared:	0.000			
Model:	OLS	Adj. R-squared:	-0.022			
Method:	Least Squares	F-statistic:	5.051e-05			
Date:	Thu, 22 Jun 2023	Prob (F-statistic):	0.994			
Time:	14:46:23	Log-Likelihood:	-194.28			
No. Observations:	48	AIC:	392.6			
Df Residuals:	46	BIC:	396.3			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	24.4749	4.078	6.001	0.000	16.266	32.684
Deseasonalized_data	4.313e-05	0.006	0.007	0.994	-0.012	0.012
Omnibus:	11.100		Durbin-Watson:	0.005		
Prob(Omnibus):	0.004		Jarque-Bera (JB):	2.885		
Skew:	0.002		Prob(JB):	0.236		
Kurtosis:	1.799		Cond. No.	1.34e+03		
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The condition number is large, 1.34e+03. This might indicate that there are strong multicollinearity or other numerical problems.						

the values for Period_t, Sales, Deseasonalized_data, Dt, and SF for each period. The SF represents the seasonal factor, indicating the influence of seasonal patterns on the sales data. The Dt variable represents the deseasonalized data with trend, removing the seasonal component while considering any underlying trend. Given the low R-squared value and the non-significant coefficient for the deseasonalized data, it suggests that the model does not accurately capture the patterns or factors driving the variation in the Period_t variable for the Furniture category. This implies that there may be other factors or variables not considered in the analysis that are more relevant in explaining the variations in sales.

3. Technology Category

The Technology category was analyzed using a static method approach called OLS regression (Ordinary Least Squares). The model achieved a score of 0.5576094203666919, indicating a moderate fit between the predicted values and actual sales data. However, the low R-squared value of 0.009 suggests that only a small fraction (0.9%) of the variation in average monthly sales can be explained by the model.

Examining the OLS regression results, the coefficient for the "Deseasonalized_data" variable is 0.0031, with a p-value of 0.525. This coefficient represents the impact on average quarterly sales

for each unit change in the deseasonalized data. However, the p-value indicates that the coefficient is not statistically significant at the conventional significance level ($\alpha = 0.05$). Consequently, the model's ability to accurately predict average monthly sales based on the deseasonalized data is limited. The model's performance was evaluated by analyzing the final table, which provides the seasonal factor (SF) for each period. SF represents the ratio of observed sales data to deseasonalized data, indicating the seasonal variation in the Technology category. SF values range from 0.286958 to 2.671485, signifying different degrees of seasonal impact across periods.

Interpreting the SF values in the Figure 9 below, periods with SFs below 1 indicate lower sales compared to the deseasonalized average, while SFs above 1 suggest higher sales. In the Technology category, periods 1, 2, 6, 7, 9, 10, 11, 12, 13, 14, 18, 21, 22, 27, 28, 30, 31, 32, 33, 37, 40, 41, 43, 44, 45, and 47 have SFs below 1, indicating lower sales. Conversely, periods 3, 4, 5, 8, 15, 16, 17, 19, 20, 23, 24, 25, 26, 29, 34, 35, 36, 38, 39, 42, 46, and 48 have SFs above 1, suggesting higher sales.

Out [466]:

	Period_t	Sales	Deseasonalized_data	Dt	SF
0	1	449.041429	0.000000	793.999250	0.565544
1	2	229.787143	0.000000	800.769712	0.286958
2	3	2157.331600	0.000000	807.540173	2.671485
3	4	640.938857	0.000000	814.310634	0.787094
4	5	564.698588	0.000000	821.081095	0.687750
5	6	766.905909	0.000000	827.851556	0.926381
6	7	522.618933	934.539057	834.622017	0.626174
7	8	687.465385	961.869885	841.392478	0.817057
8	9	2025.583200	905.232592	848.162940	2.388201
9	10	596.900900	842.565572	854.933401	0.698184
10	11	1202.936960	856.047443	861.703862	1.395998
11	12	1142.956889	853.091725	868.474323	1.316051
12	13	903.647200	845.668029	875.244784	1.032451
13	14	431.121250	873.180395	882.015245	0.488791
14	15	596.702471	849.257779	888.785706	0.671368
15	16	697.559500	804.386463	895.556167	0.778912
16	17	831.642857	797.422333	902.326629	0.921665
17	18	429.024400	810.410935	909.097090	0.471924
18	19	682.331733	809.049494	915.867551	0.745011
19	20	1188.049385	820.941777	922.638012	1.287666
20	21	950.856400	876.492518	929.408473	1.023077
21	22	594.716111	905.268461	936.178934	0.635259
22	23	1037.982652	943.699641	942.949395	1.100783

Figure 9: Technology seasonal factor table

In conclusion, the analysis of the Technology category highlights limitations in accurately predicting average monthly sales using the static method approach with OLS regression. The low R-squared value indicates that the model fails to capture a significant portion of sales variation, and the coefficient for deseasonalized data lacks statistical significance. Moreover, the varying SF values indicate the presence of seasonal patterns, with certain periods experiencing lower or higher sales compared to the deseasonalized average.

Adaptive Method

Monthly and Quarterly Sales Analysis

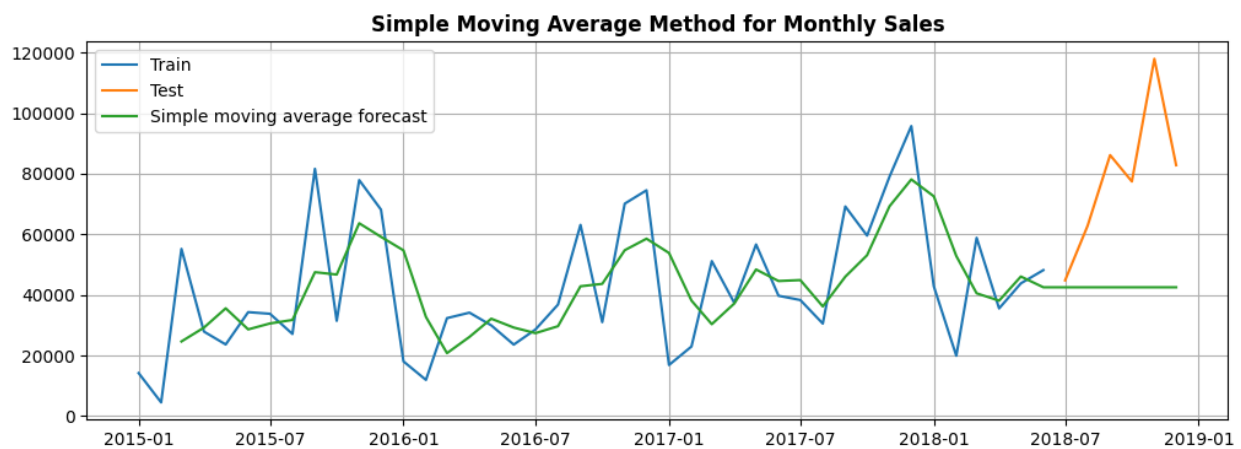


Figure 10: Graph of Simple Moving Average Model for Monthly Sales

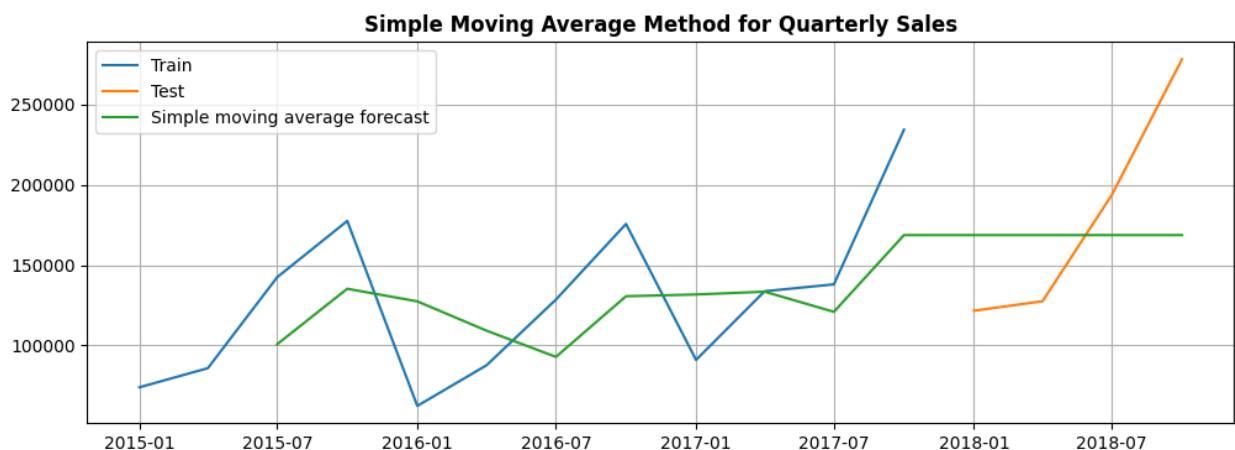


Figure 11: Graph of Simple Moving Average Model for Monthly Sales

Figure 10 and Figure 11 shows the plot for adaptive forecasting using the simple moving average model for monthly sales and quarterly sales. From the graph, it can be seen that the sales dataset shows an upwards trend and seasonality occurs. The model takes the rolling window at a period of 3 months of historical data to predict the future values. However, it can be seen that the model can only predict the training sets and not account the testing sets. This makes sense since the disadvantage of a simple moving average model is that the model does not account for future events and only can be used for past datasets. The model also does not capture the trend and seasonality occurred, hence, cannot be used for future sales forecasting methods.

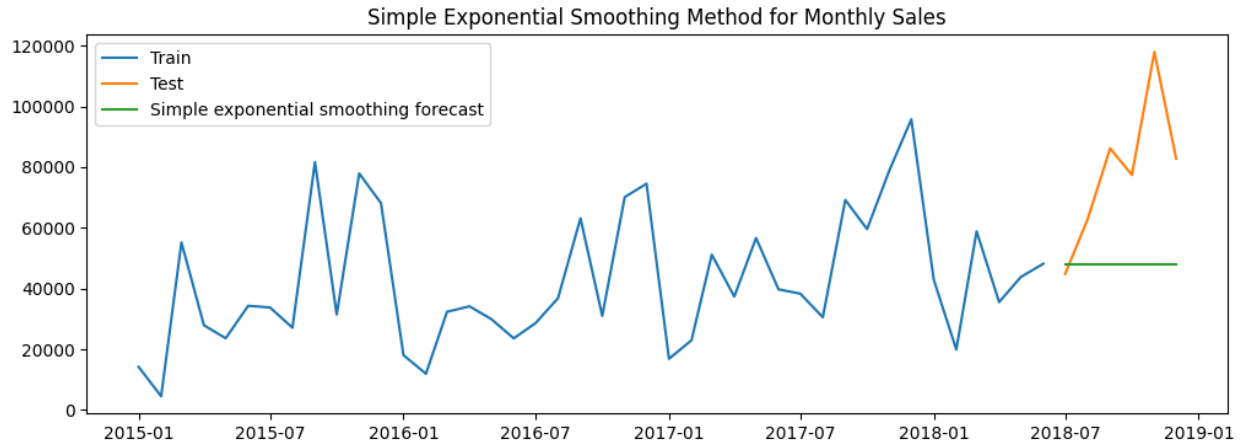


Figure 12: Graph of Simple Exponential Model for Monthly Sales

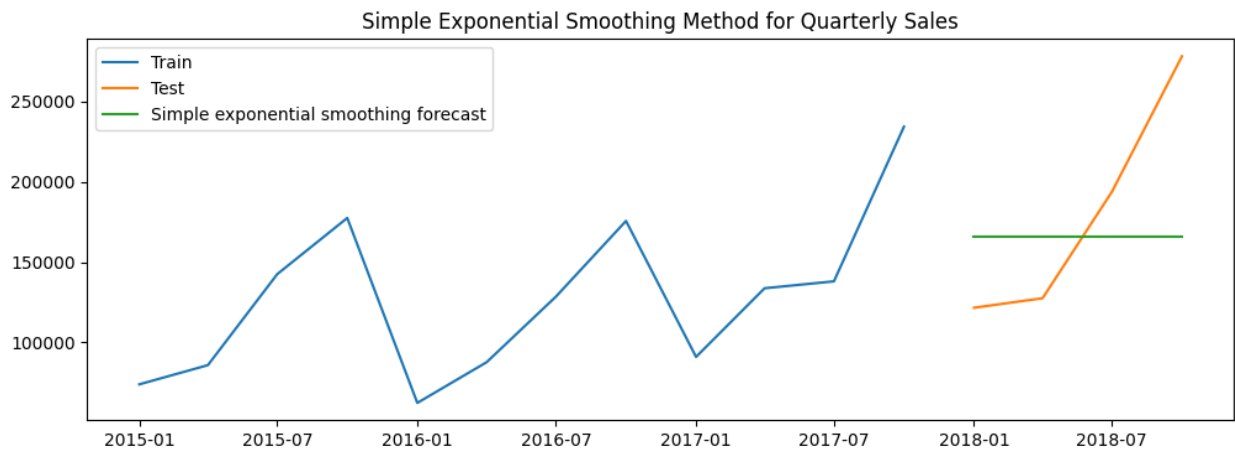


Figure 13: Graph of Simple Exponential Model for Quarterly Sales

The graph for simple exponential smoothing forecast for monthly and quarterly sales is being plotted as shown in Figure 12 and Figure 13. It can be seen that the simple exponential smoothing does not provide a good forecasting result as the forecasting values remain the same. According to Hydman and Athanasopoulos, a simple exponential smoothing forecast can give the same value as the model has a “flat” forecast function. This indicates that the forecast obtains the same value equal to the last level component (Hydman and Athanasopoulos, 2018). This is because the model works the best with time series data that has no trend or seasonal component. Since the time series dataset contains both trend and seasonal components, the model does not capture the components thus resulting in the flat result. Therefore, the model is not suitable for future sales forecasting since we have to account for both trend and seasonality.

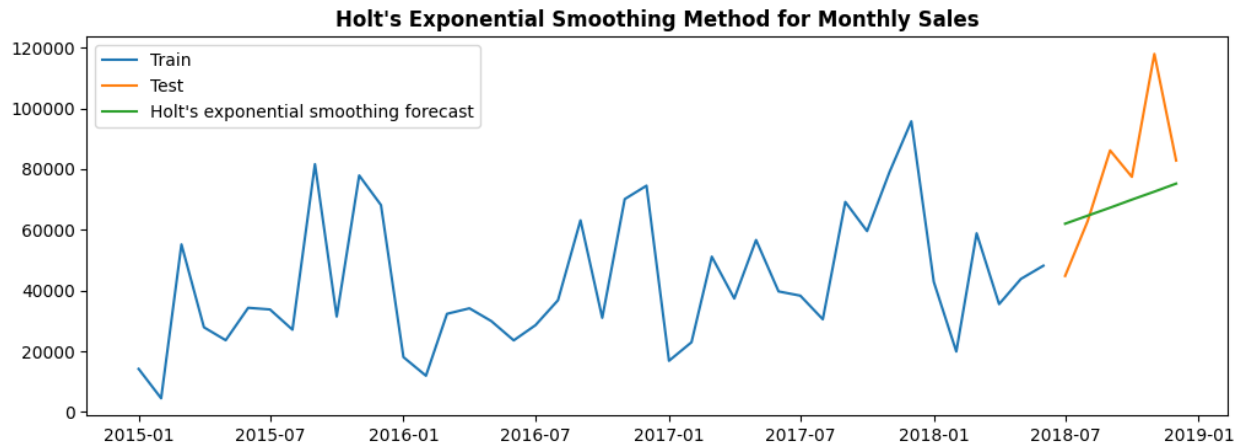


Figure 14: Graph of Holt's Model for Monthly Sales

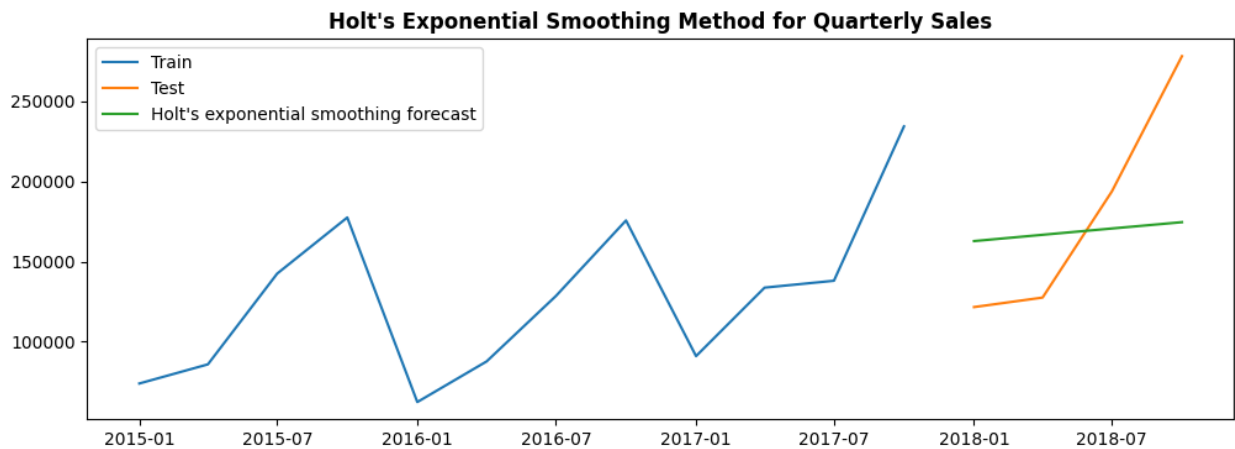


Figure 15: Graph of Holt's Model for Quarterly Sales

Figure 14 and Figure 15 show the forecasting result for Holt's Model for monthly and quarterly sales. Similarly to the simple exponential smoothing model, the model only accounts for the trend component but not the seasonality component of the time series dataset. This is because Holt's Model allows the forecasting time series dataset with trend components. The forecast function can be obtained by running a linear regression as a function of the trend and level. Since the model only captures trend components, the long-term forecasts can be inaccurate as it does not account for seasonality. However, the model can be used for short term forecasts where the seasonality component is absent. Therefore, the model is not suitable for forecasting the long-term sales for the company.

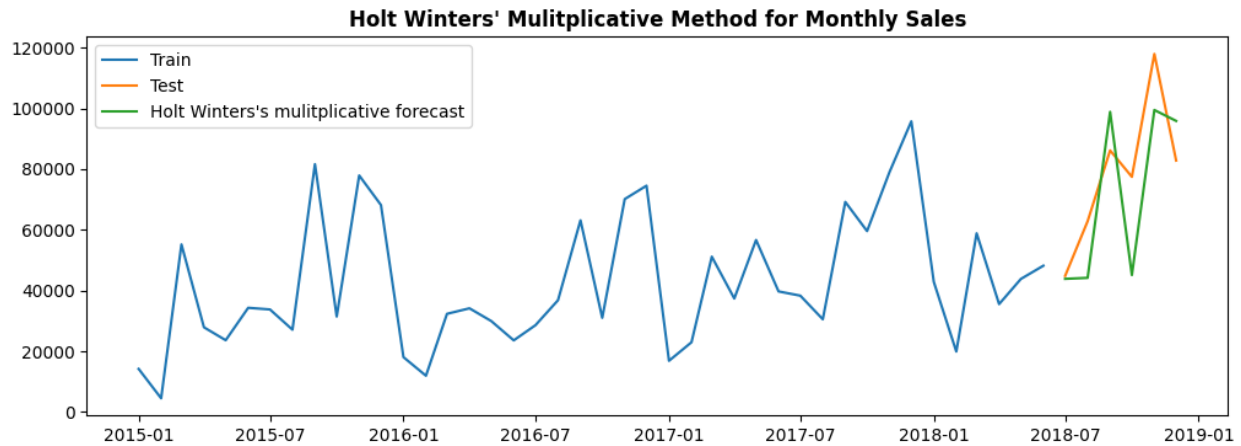


Figure 16: Graph of Winter's Model for Monthly Sales

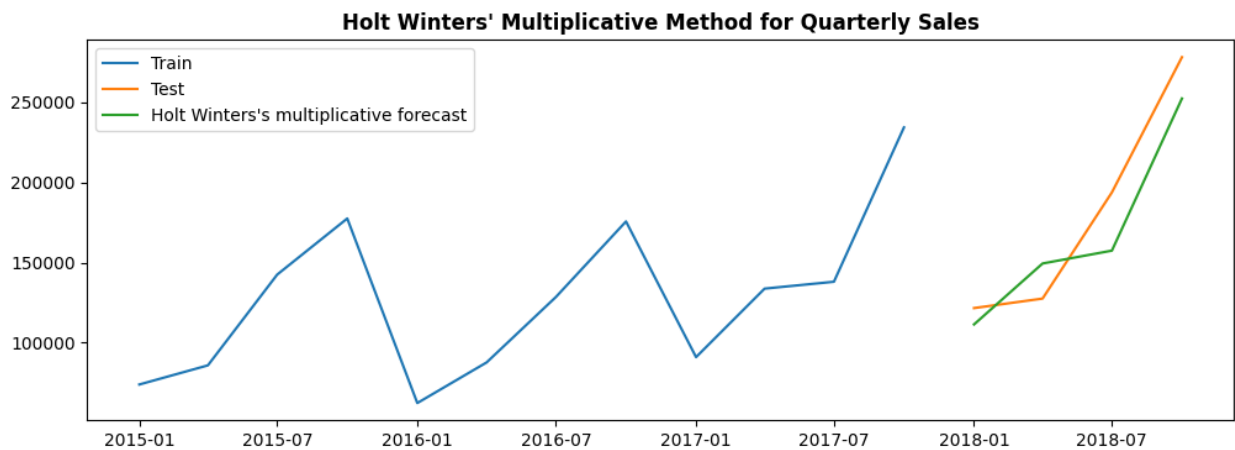


Figure 17: Graph of Winter's Model for Quarterly Sales

The last forecasting method used in this project is Winter's Model or popularly known as Holt's Winter's Multiplicative Model. The figure above shows the graph of the monthly and quarterly sales forecast using the said model. From the graph, it can be seen that the model managed to capture both trend and seasonality present in the time series dataset. This is beneficial in predicting the future sales as the business exhibits seasonality and trend in their sales. In the Winter's Multiplicative Model, the parameters alpha, beta and gamma are present, representing the smoothing factor. Alpha represents the level smoothing factor, beta represents the trend smoothing factor and gamma represents the seasonality smoothing factor. The parameters are being updated to the past values to make predictions for the future time periods. Using this method, the smoothing process removes noise and emphasizes the trend and seasonality of the data, making it easier to make forecasts and obtain valuable insights. Figure 18 and Figure 19 shows the summary of the fitted model obtained, and it can be seen that the smoothing parameters are being constantly optimized to produce the best forecasting. Hence, this model is the most suitable method to forecast the future sales for the company.

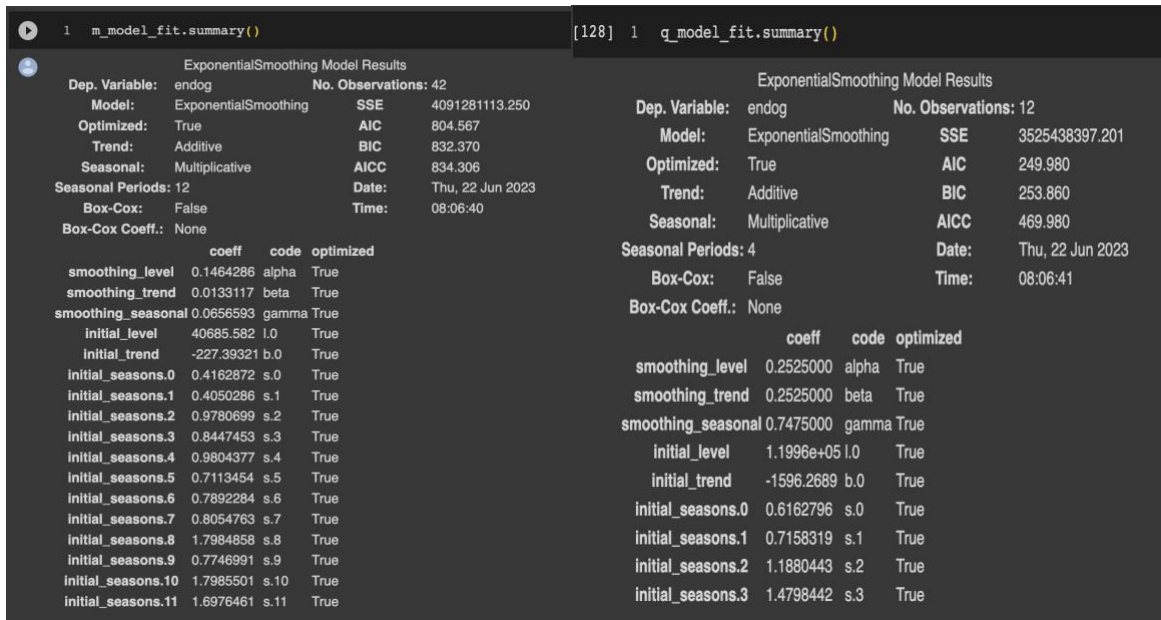


Figure 18: Summary of Winter's Model for Monthly Sales

Figure 19: Summary of Winter's Model for Quarterly Sales

Forecasting Error for Monthly and Quarterly Sales using Various Models

Table 2: Summary of Forecasting Error for Monthly and Quarterly Sales

	Method	RMSE	MAPE
Monthly	Simple Moving Average	42533.42	40.97
	Simple Exponential Smoothing	37881.75	35.64
	Holt's Model	21699.61	20.08
	Winter's Multiplication Model	18558.43	19.95
Quarterly	Simple Moving Average	64282.42	30.83
	Simple Exponential Smoothing	64905.95	30.27
	Holt's Model	60207.04	28.44
	Winter's Multiplication Model	25342.93	13.39

Based on Table 2, the forecasting errors of the models are being calculated and the RMSE and MAPE values are being recorded. RMSE, referring to Root Mean Squared Error, is the square root of mean squared error between the predicted and actual values. RMSE is better when the value is lower as higher RMSE value means there is a large deviation in the forecast. On the other hand, Mean Absolute Percentage Error (MAPE) is the mean of all absolute percentages between the predicted and actual values. Lower MAPE value indicates that the model performance is better hence more accurate forecasting.

From the table, it can be seen that Winter’s Multiplication model has the lowest RMSE and MAPE value, showing the least error in both monthly and quarterly sales forecasting compared to the other models. This is aligned with the time series data as both trend and seasonality components exist, hence the model provides the best forecast. Therefore, in forecasting the future sales for Company ABC, Winter’s Multiplication Model is being selected.

Product Monthly Sales Forecasting by Product Categories

The adaptive forecasting model selected for the sales forecasting for Company ABC products is Winter’s Multiplication Model. By applying the same methodology explained in the Analytics Method section, the monthly sales forecasting for three product categories, Furniture, Office Supplies and Technology is being done and a graph of the forecasting result is being plotted as shown in figure below. The forecasting error for the three product categories also is being calculated and as shown in Table 3.

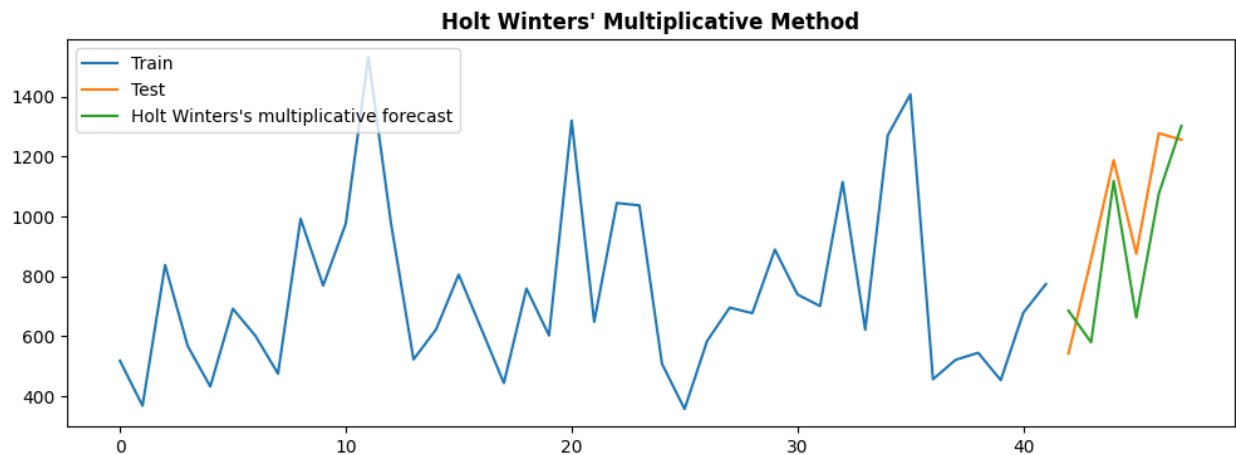


Figure 20: Forecasting plot for the monthly sales forecast of furniture category

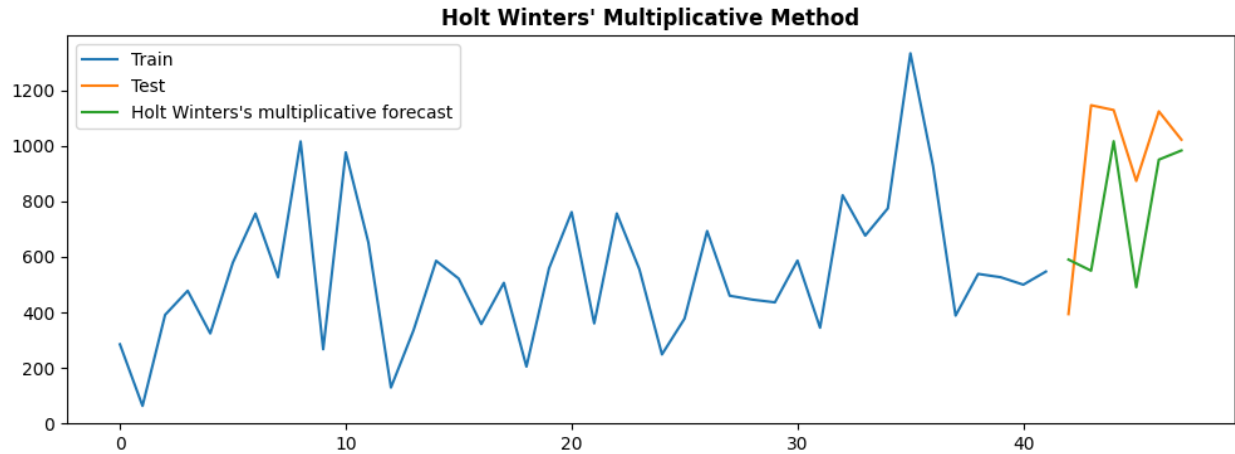


Figure 21: Forecasting plot for the monthly sales forecast of office supplies category

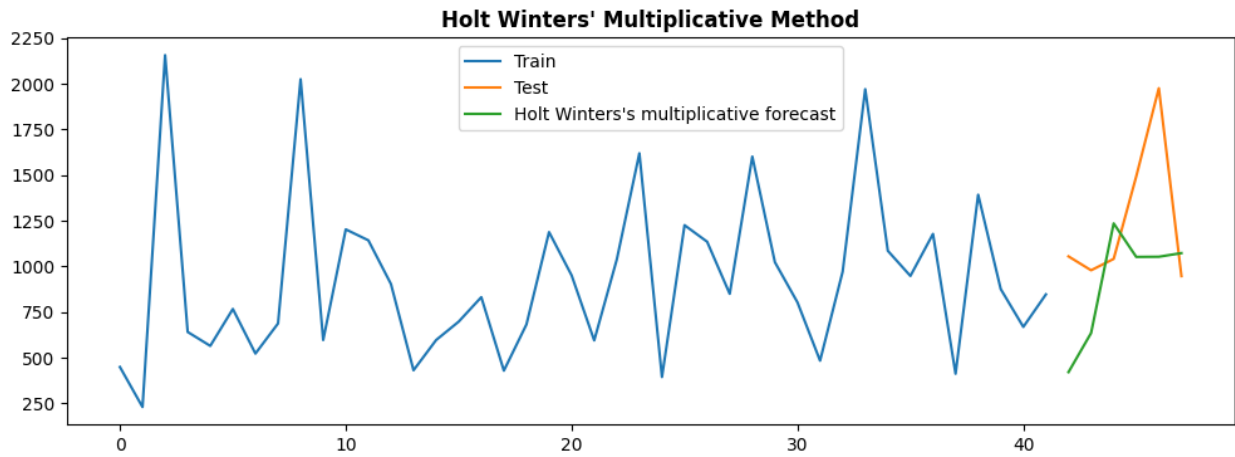


Figure 22: Forecasting plot for the monthly sales forecast of technology category

Table 3: Forecasting Error for Product Categories

Product Categories	RMSE	MAPE
Furniture	178.10	18.08
Office Supplies	312.14	29.15
Technologies	519.82	33.91

Although in the previous section it was shown that the quarterly average forecasting gives the best forecasting result, this section focuses on the monthly average forecasting. This is because the monthly average forecasting gives a better insight on the sales planning for the three product categories. By breaking down the product categories into three different categories, it can be seen that the furniture category has the largest seasonality while the sales in office supplies and technology category shows an upward trend. Monthly forecasting results in a more detailed

prediction to capture the short-term insights. It enables a better understanding of the seasonal effects, such as the monthly trends and holiday seasons and the company can adjust their sales strategy and production levels accordingly. Monthly forecasting also allows the company to have a better insight in allocating resources more efficiently, such as allocating more production, inventory and workforce levels on busier months. For example, the sales are expected to spike during the holiday season, so more workforce, production and inventory should be prepared to accommodate the higher demand.

Another insight obtained is that when breaking down the product categories, the error becomes larger. This can be due to limited data availability that results in higher variability and larger forecasting error. In order to mitigate these limitations, the company can improve their data quality by incorporating more external factors, such as the market trend to improve the forecasting accuracy. The company can also improve their forecasting error by regularly updating and adjusting the forecasting model as soon as the new data becomes available. Alternatively, the company can employ different forecasting techniques such as Seasonal ARIMA model (SARIMA) or machine learning algorithms to improve the forecasting technique and reduce the errors.

Lastly, based on the forecasting result, there are still some limitations to the Winter's Multiplication model. Since the model utilizes the trend and seasonality from historical data, the model might assume the future pattern will remain the same as the historical data. This assumption can be false if there are significant changes or external factors that affect the dataset, thus resulting in more forecasting error. Next, the model is sensitive to initialization where the choice of initial values for level, trend and seasonal component can affect the forecasting performance. Finally, Winter's Multiplication model does not capture any outlier or irregular patterns that may deviate the results, hence leading to less accurate forecasting. As a result, the company may still need to employ more forecasting models when they want to do long-term forecasting.

CONCLUSION

In this project, we have analyzed the time-series Superstore retail dataset using demand planning analytic. We employed both static forecasting methods and adaptive forecasting methods, including models such as Simple Moving Average model, Simple Exponential Smoothing model, Holt's Model and Winter's Model to predict the store's future sales.

After cleaning the data and aggregating it into monthly average sales, we focused on three categories: Office Supplies, Furniture, and Technology. Our analysis revealed the presence of seasonality and trend in these categories. However, the static method approached with OLS regression demonstrates limitation in accurately predicting average sales. The models show weak fits, low R-squared values, and non-significant coefficients for deseasonalized data. This indicates that additional variables or factors may be necessary to improve the accuracy of sales predictions in these categories. On the other hand, in the Adaptive forecasting model, it was found that Winter's Model produces the least forecasting error due to the ability of the model to capture the trend and seasonality component in the dataset. The monthly sales forecasting for the next 6 months using Winter's Model was also done on product categories such as furniture, office supplies and technology supplies and it was observed that by breaking down the forecasting into smaller categories, there is some limitation to the model that leads to larger forecasting error.

Based on our findings, there are several adjustments that can be made to improve the forecasting and reduce forecasting error. First, the company should improve the data quality by accounting for external factors, and capturing any outlier or anomalies occurring in the dataset that might lead to larger forecasting error. Next, the company needs to regularly update their forecasting model to keep updating the new data to be trained into the forecasting model and lastly, the company can employ machine learning algorithms or other sophisticated forecasting techniques such as SARIMA model to improve the forecasting accuracy. By addressing these areas, the company can improve their forecasting accuracy, improve resource allocation and have an enhanced operational efficiency.

As a conclusion, this report emphasizes on having efficient demand planning by conducting demand analytics with appropriate methods. It is crucial for businesses to acknowledge the importance of continuously improving their demand planning and forecasting strategies. By leveraging the data-driven insights, incorporating sophisticated techniques, the forecasting error and limitation could be mitigated, leading to improved customer satisfaction, an efficient resource allocation and the company can make informed decisions in achieving their business goals.

REFERENCE

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Jha, B. K., & Pande, S. (2021, April). Time series forecasting model for supermarket sales using FB-prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 547-554). IEEE

Vincent, T. (2017). A Guide to Time Series Forecasting with Prophet in Python 3. URL: <https://www.digitalocean.com/community/tutorials/a-guide-to-time-seriesforecasting-with-prophet-in-python-3>.

Additional Remark

The Python code for this project is being done in two notebooks:

Notebook 1 includes the data preparation, EDA and adaptive forecasting models and accessible in this Google Colab notebook:

https://colab.research.google.com/drive/16_RX4vnbWDuOKgRWzUMJPPJ0pUZLOd5K?usp=sharing

Notebook 2 includes the static forecasting models and is attached in the submission file '*Hannah-Mehdi-Ahmad (Static Method).ipynb*'.