HUMBER INSTITUTE OF TECHNOLOGY
AND ADVANCED LEARNING
(HUMBER COLLEGE)

TOPIC: Predictive Analysis of Product Demand for Modern Supply Chain Management

GROUP NUMBER 10:

Last Name	**First Name**
Nasim	Maha
Narain	Dakshita
Gosrani	Krupa
Jinwala	Maaz
Patel	Shivam
Valecha	Charmi

Introduction:

In today's fast-paced business environment, staying ahead of demand fluctuations is crucial for companies aiming to optimize their operations and enhance customer satisfaction. Whether in retail, manufacturing, or logistics, accurately predicting demand ensures efficient inventory management and resource allocation.

This report delves into the realm of demand prediction, addressing a common challenge faced by companies: forecasting product demand. By harnessing historical data and advanced forecasting techniques, businesses can anticipate future demand patterns, enabling proactive decision-making and resource optimization.

Business Problem:

The challenge lies in predicting future demand amidst dynamic market conditions and ever-changing consumer behavior. Inaccurate forecasts can lead to overstocking, stockouts, and missed revenue opportunities. To overcome this, our goal is to develop robust forecasting models that provide accurate insights into future demand trends.

Key Objectives:

- Data Exploration and visualization: Conduct thorough analysis of historical demand data to identify trends, patterns, and seasonality and uncover hidden insights
- Model Development: Develop and evaluate various forecasting models, including linear trend, exponential trend, polynomial trend, and seasonal models.
- Performance Evaluation: Assess the accuracy and reliability of each forecasting model using appropriate evaluation metrics.
- Forecasting: Generate future demand predictions based on the best-performing model(s) to support decision-making and planning.

Expected Outcomes:

Enhanced understanding of demand dynamics and underlying patterns.

Identification of the most suitable forecasting model(s) for accurate predictions.

Improved forecast accuracy leading to better inventory management and resource allocation.

Empowered decision-making and proactive planning to meet customer demand effectively.

Literature Review

Demand forecasting methodologies play a pivotal role in supply chain management, as evidenced by a wealth of literature emphasizing their importance. Classical methods, such as Exponential Smoothing, have stood the test of time and have been extensively explored due to their effectiveness in capturing historical trends and seasonal variations. These methods provide a solid foundation for understanding demand patterns and making informed decisions regarding inventory management and production planning.

A recurring theme in the literature is the importance of integrating data visualization with forecasting methodologies. Visual representations of data not only aid in understanding historical patterns but also facilitate communication and decision-making within organizations.

By visualizing trends, anomalies, and forecasted outcomes, stakeholders can gain insights and collaborate effectively to address supply chain challenges.

While classical techniques provide a solid foundation, modern machine learning algorithms offer opportunities for improving forecast accuracy and capturing complex relationships within data. However, the practical implementation of these methods requires careful consideration of organizational capabilities, resource constraints, and business objectives. By integrating classical and modern approaches within a resilient framework, organizations can create robust demand prediction systems capable of navigating dynamic business environments effectively.

Data Overview

The dataset offers a thorough understanding of product demand, encompassing different product categories that are dispersed throughout numerous warehouses. The company's four core warehouses are positioned strategically to cater to global markets, and its production facilities are spread around the globe. This structured format facilitates analysis and allows stakeholders to gain insights into product demand patterns, warehouse-specific trends, and category-wise demand variations over time.

	Product_Code	Warehouse	Product_Category	Date	Order_Demand
1	Product_0965	St john's	Category_006	08-01-2014	2
2	Product_1724	St john's	Category_003	31-05-2014	108
3	Product_1521	Surrey	Category_019	24-06-2014	85000
4	Product_1507	Surrey	Category_019	24-06-2014	7000

Table 1. Dataset Overview

- **Product_Code**: Unique identifier associated with each product.
- **Warehouse**: Location or warehouse where the product demand is recorded.
- **Product_Category**: Category to which the product belongs.
- **Date**: Date on which the demand for the product was recorded.
- **Order_Demand**: Quantity of the product demanded on the specified date.

Data Cleaning

Duplicate values refer to identical records that appear more than once in the dataset. These duplicates can skew the analysis results and lead to inaccurate insights. In our dataset, we identified and removed duplicate records based on all columns to ensure data integrity. The 'Date' column undergoes necessary formatting to meet datetime requirements. We identified columns with null values and decided on an appropriate strategy for handling them, which involved imputation or removal depending on the extent of missing data. The percentage of missing values is less than 1% of the data. Order demand quantities cannot be negative, we converted any negative values to their absolute positive counterparts to rectify this issue. The analysis focuses on positive demand values for further insights. Outliers are data points that significantly deviate from the rest of the dataset. These anomalies can distort statistical analyses and model predictions. To address outliers in the demand data, we employed techniques such as statistical methods (e.g., z-score analysis) to detect and remove or mitigate the impact of outliers.

<u>**Data Exploration and Visualization**</u>

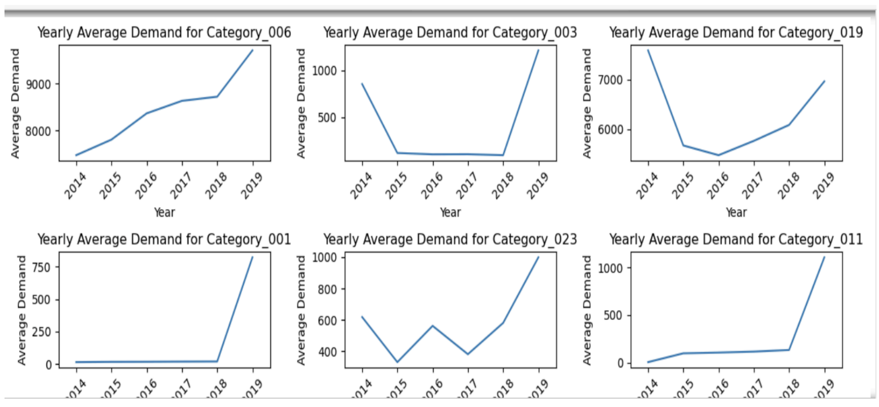

Figure 1. Average Demand for Product Category over the years

Once we've cleaned the data, we group it by year and category. This helps us to find out how much of each product category was demanded each year, as shown in Figure 1. Additionally it aids us to understand the historical performance of each product category overtime and helps with identifying high and low demand categories.

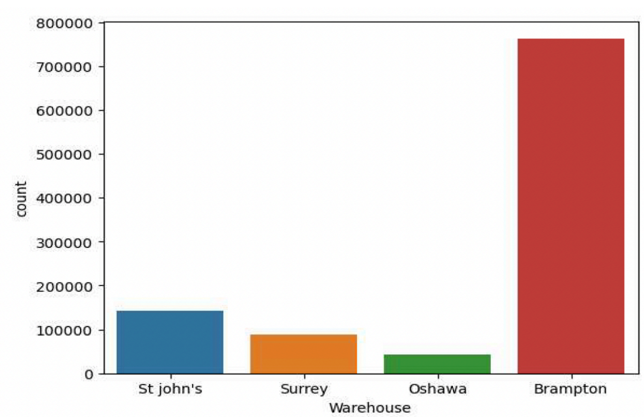

Figure 2. Aggregate Demand among the Warehouses

After establishing the demand patterns of each product category throughout the years, we then plot a bar graph that assists in grasping how products are spread across various warehouses, offering valuable insights into managing inventory and optimizing logistical operations. The bar plot in Figure 2. showcases that Brampton warehouse is the most preferred and highest in demand location for storing product categories whereas Oshawa warehouse is the least preferred.

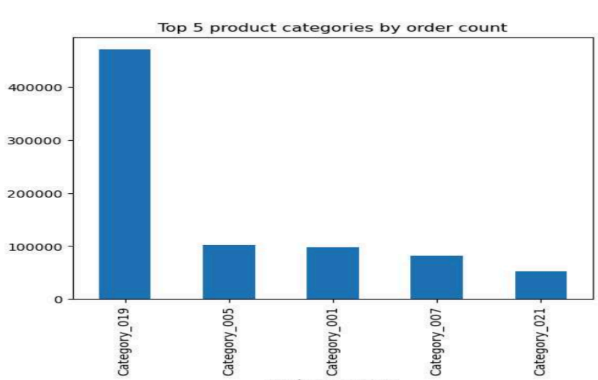

Figure 3. Top 5 Product Categories by Order Count

The graph illustrated in Figure 3. shows the top five most popular product categories based on their order count. Each category is represented by a bar, with the height of the bar indicating the frequency or count of orders associated with that particular category. By visualizing this data, the manufacturing company can quickly discern which product categories are the most frequently ordered, providing valuable insights into consumer preferences and potentially informing business decisions such as inventory management, marketing strategies, and product development efforts. It shows that the product Category_019 has the highest order count. This effectively highlights the relative popularity of this product category within the dataset, allowing stakeholders to focus their attention and resources accordingly.

After visualizing Fig. 3, we also calculated total order demand by product category and its % contribution where we find that 'category_019' dominates other categories with approximately 77% of total demand.

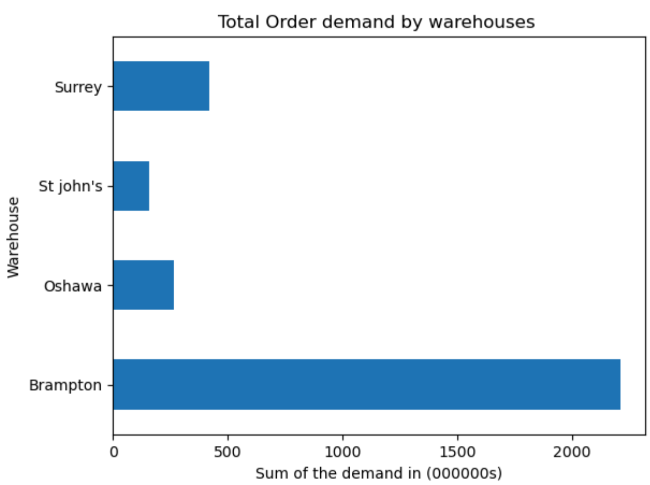

Figure 4. Total Order Demand by Warehouses

The plot generated by the provided code showcases the total order demand aggregated by the warehouse. By visualizing this data, the stakeholders can quickly discern which warehouses have the highest demand for products, providing valuable insights into distribution patterns and the effectiveness of warehouse management. Here, Brampton exhibits the highest total order demand by warehouse and St.John's has the lowest. Overall, the plot would involve decision-making processes related to supply chain management and logistics.

Figure 5. Average Demand for Warehouse

After we figure out the total order demand by warehouse, we then provide a demonstration of the average order demand across different warehouses as shown in Figure 5. Plotting this data as horizontal bars makes it simple to compare the average demand levels between the warehouses, which helps to understand how demand is distributed across various locations and informs inventory management decisions.

Oshawa has the highest average demand of around 7000 whereas St. John's has the lowest average demand of around 1200.

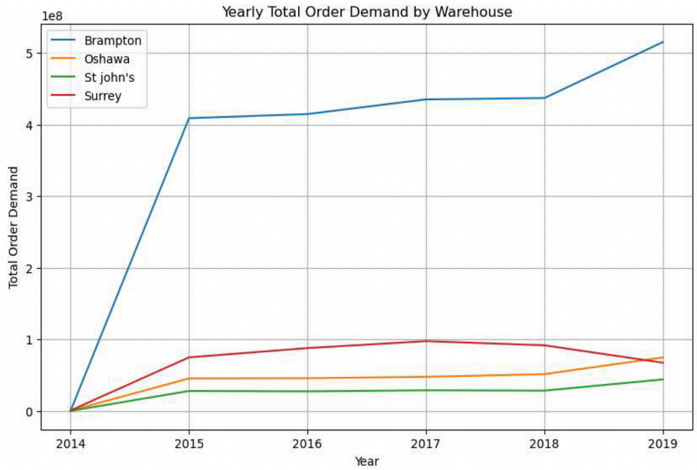

Figure 6. Yearly Total Order Demand by Warehouse

The purpose of this code is to visualize the yearly total order demand across different warehouses over time. As depicted in Figure 6., Brampton consistently exhibits the highest total order demand among the warehouses, while St. John's consistently shows the lowest yearly total order demand. Interestingly, total order demand peaks across Brampton, Oshawa and St. John's warehouses in 2019, but Surrey experiences a decrease between the years 2018-2019.

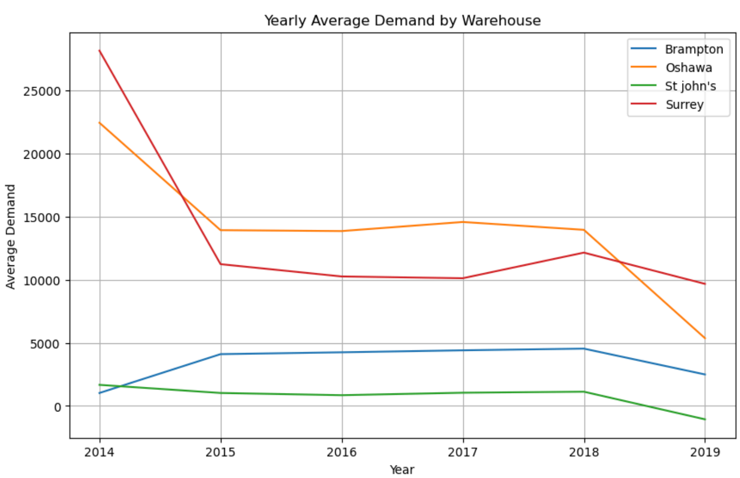

Figure 7. Yearly Average Demand by Warehouse.

The plot generated by the provided code illustrates the yearly average demand for each warehouse over the specified period. By visualizing this data, the company can observe how the average demand fluctuates over time for each warehouse. The plot enables stakeholders to identify patterns such as seasonal variations or long-term trends in demand, facilitating strategic decision-making related to inventory management, resource allocation, and supply chain optimization. It can be seen that St. John's and Brampton exhibit the lowest average demand across the years 2014-2019. Surrey had the highest average demand in 2014, which dipped by 2015 and remained consistent. Oshawa also saw this dip in 2015.

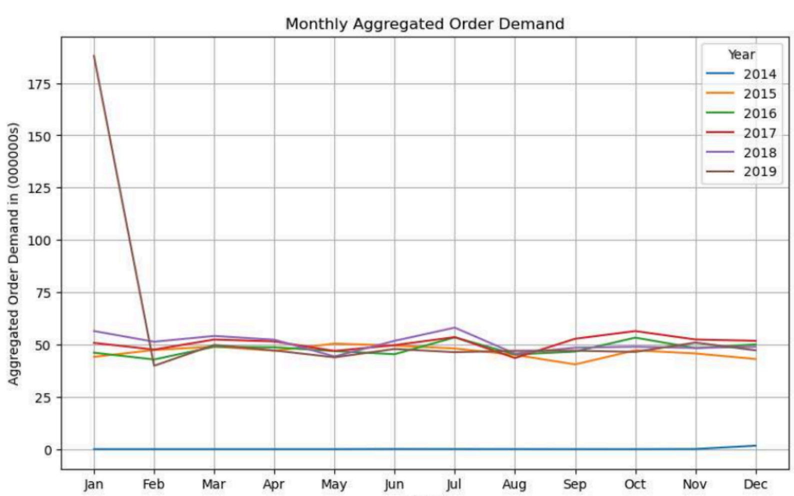

Figure 8. Monthly Aggregated Order Demand

The plot generated above illustrates the aggregated order demand every month over multiple years. This plot enables stakeholders to identify seasonal trends and patterns in order of demand, such as peak months or periods of low demand. Analyzing the aggregated demand data every month allows businesses to anticipate and prepare for fluctuations in demand, aiding in inventory management, production planning, and resource allocation. It can be seen that demand remained consistent across all months for all years around 50 (000000's). The demand was 0 throughout 2014. It can be inferred that the company started operating in December of 2014, which is why a slight rise can be seen in December of that year. The demand was the highest in January of 2019, and a steep decline was seen from January till February after which it remained consistent.

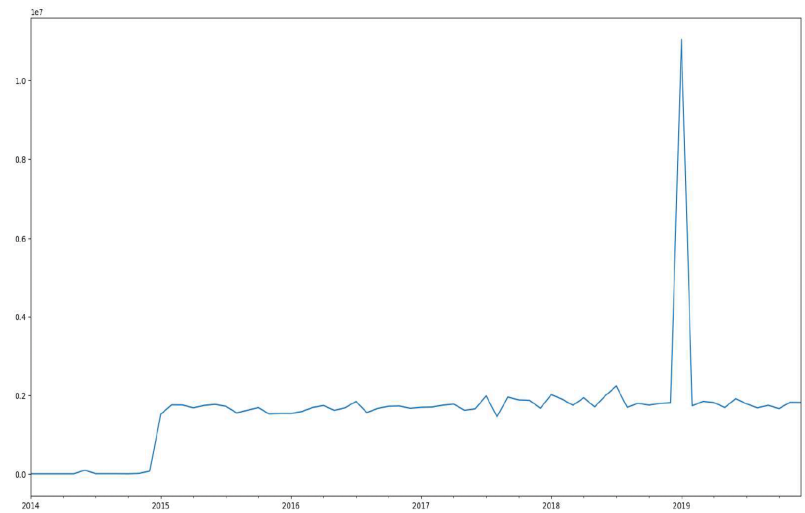

Figure 9. Sales Trend over time

In Figure 9., we have visualized the monthly average sales over a period of several years, from 2014 to 2019. We are able to identify that there is a positive trend in sales, indicating an increase in average monthly sales over time each year with highest sales being recorded in the year 2019.

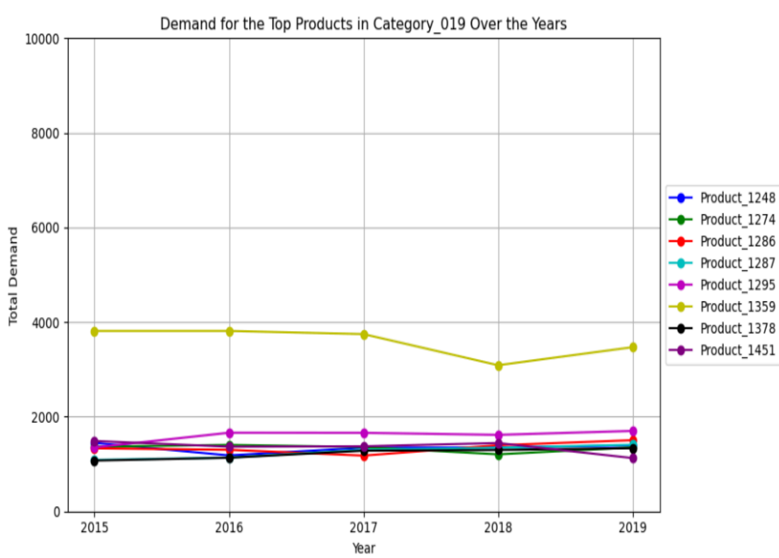

Figure 10. Demand for the Top Product Categories in Category_019 over the years

The graph in Figure 10. is plotted to visualize the demand trends for each of the top products over the years among the highest demanded Category_019. Each product is represented by a separate line on the plot, with different colors assigned to differentiate them. Product_1359 had the highest demand across all years while other products were comparable to each other.

Time Series Analysis And Forecasting

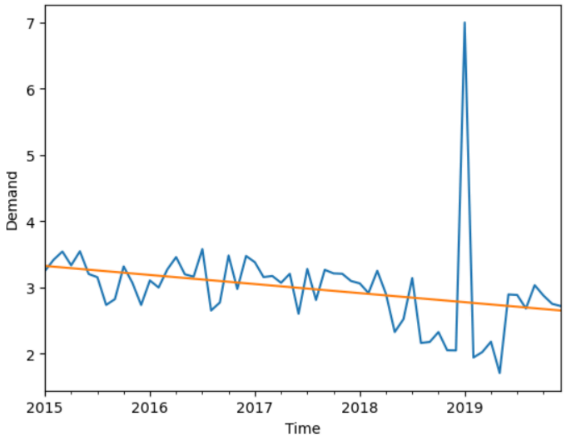

Figure 11. Time Series for Order Demand across the years

After fitting the model, a plot is generated to visualize the original demand time series along with the predicted values from the linear trend model. This trend line is fitted using a linear regression model, with time as the independent variable and demand as the dependent variable. By visualizing the demand time series and the fitted trend line together, stakeholders can assess the general direction and magnitude of the demand trend over the observed period. This plot aids in understanding the underlying patterns and tendencies in demand behavior, providing insights that can inform forecasting, planning, and decision-making processes related to inventory management and resource allocation.

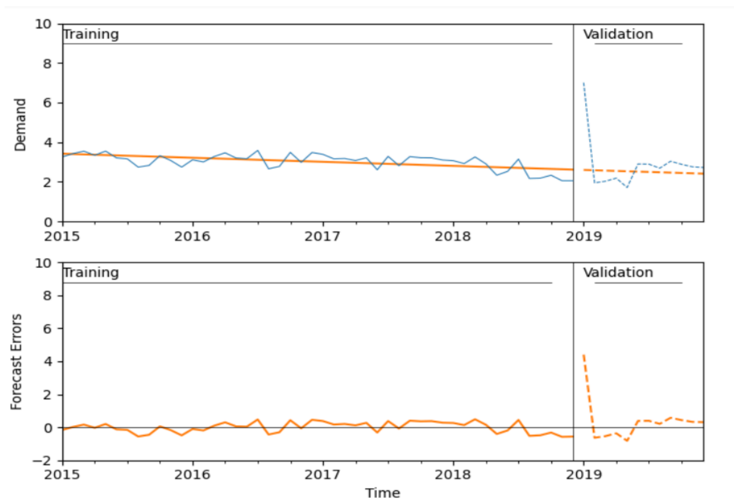

Figure 12. Training and Validation Visuals.

The top subplot titled "Demand" shows two lines. The solid blue line represents the predicted demand from the training set, and the dashed blue line represents the predicted demand from the validation set. The x-axis represents time, but the scale seems to be years between 2014 and 2020. The y-axis shows the value of demand, but the scale goes from 0 to 120,000.

The bottom subplot titled "Forecast Errors" shows two lines, similar to the top subplot. The solid blue line represents the residuals (the difference between actual and predicted demand) for the training set, and the dashed blue line represents the residuals for the validation set. The x-axis again represents time, and the y-axis shows the value of the forecast error.

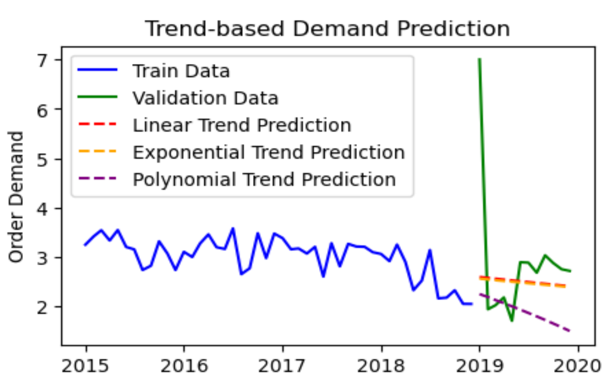

Figure 13. Trend Based Demand Prediction

The demand prediction plot showcases the performance of various trend models against actual and validation data. The blue line represents the actual order demand observed in the training data, while the green line denotes the demand in the validation set. Predictions from three trend models are overlaid: a linear trend prediction (red), an exponential trend prediction (orange), and a polynomial trend prediction (purple). The plot illustrates how each model captures and forecasts the order demand trend differently over time. Such visualizations offer valuable insights into the efficacy of different modeling approaches for demand forecasting, aiding decision-making processes in supply chain management and inventory optimization.

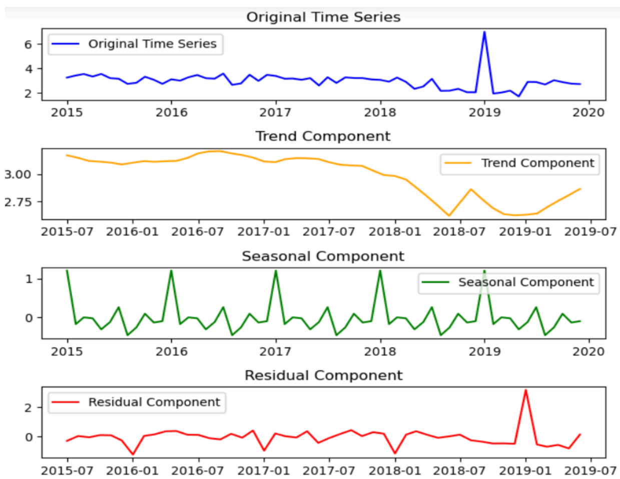

Figure 14. Components of Time Series

By decomposing the time series into these additive components, we can gain insights into the underlying patterns and structures of the data. Here we can see that there are a lot of noise components in the data. Further we will try to use simple exponential smoothing to help reduce the effect of noise in our forecasting.

After depicting and understanding the components of time series as shown in Figure 14., we then calculate the OLS Regression Results. OLS stands for Ordinary Least Squares, which is a type of linear regression. Here, "Order_Demand" is the variable being predicted by the model and "C(Month)" indicates that categorical variables have been created for each month.
Through the OLS Regression Model, we are aiming to find the R-squared (R^2) value, Adjusted R-squared value and F-statistic and Prob (F-statistic) values.

- **R-squared (R^2)**: This statistic measures the proportion of the variance in the dependent variable (Demand) that is predictable from the independent variables (Month). In this case, the R-squared value is 0.300, indicating that approximately 30% of the variance in the demand can be explained by the month variable.
- **Adjusted R-squared:** This is a modified version of R-squared that adjusts for the number of predictors in the model. The adjusted R-squared value here is 0.086.
- **F-statistic and Prob (F-statistic):** The F-statistic tests the overall significance of the regression model. The Prob (F-statistic) value is the p-value associated with the F-statistic. In this case, the p-value is 0.214, indicating that the regression model as a whole is not statistically significant at the conventional significance level of 0.05.

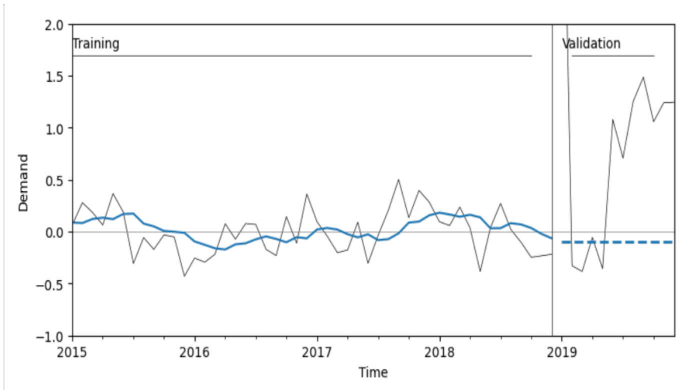

Figure 15. Exponential Smoothing

We apply exponential smoothing to the residuals (residuals_ts) using the ExponentialSmoothing class. The frequency of the time series is monthly. We fit the exponential smoothing model with a smoothing level of 0.2 (smoothing_level=0.2). Then, we plot the fitted values and forecasts of the exponential smoothing model on the same plot. We can see that it doesn't forecast well on the validation set. Further The mean absolute error (MAE) value we've obtained is approximately 2.8081. This indicates, on average, how far off our predicted demand values are from the actual demand values in our validation dataset.

Conclusion

In conclusion, this report provides insights into demand prediction methodologies and their application in supply chain optimization. Through thorough data exploration, visualization, and modeling, key insights into historical demand patterns, warehouse distribution, and product preferences were revealed. The analysis demonstrated the effectiveness of various forecasting techniques, enabling informed decisions on inventory management and resource allocation. By integrating classical and modern approaches, organizations can enhance operational efficiency and customer satisfaction.