WalMart Capstone Project – Retail Sales Forecasting

Table of Contents

1. Problem Statement

The retail store chain, Walmart, with multiple outlets across the country, faces challenges in effectively managing inventory to match demand with supply. This misalignment results in either overstocking, leading to increased holding costs and potential losses from unsold goods, or understocking, causing lost sales and customer dissatisfaction. The company is suffering from inefficiencies due to sub-optimal inventory levels, impacting overall financial performance.

2. Project Objective

The objective of this project is to develop a predictive model that accurately forecasts weekly sales for Walmart stores using historical data and relevant external factors. By improving the accuracy of sales forecasting, the project aims to optimize inventory management, reduce costs associated with overstocking and understocking, and enhance overall operational efficiency.

3. Data Description

The dataset available is "walmart.csv," which contains 6435 rows and 8 columns.

Data Description, various insights from the data.

Feature Name	Description
Store	Store number
Date	Week of Sales
Weekly_Sales	Sales for the given store in that week
Holiday_Flag	If it is a holiday week
Temperature	Temperature on the day of the sale
Fuel_Price	Cost of the fuel in the region
CPI	Consumer Price Index
Unemployment	Unemployment Rate

The data spans from 2010 to 2012 and includes information on store sales, holidays, weather conditions, and economic indicators. Further EDA revealed trends in weekly sales, seasonality, and the impacts of various external factors.

4. Data Preprocessing Steps And Inspiration

The preprocessing of the data included the following steps:

1. **Importing Libraries:** Essential libraries such as pandas, NumPy, matplotlib, and seaborn were imported to facilitate data manipulation, analysis, and visualization.
2. **Loading the Dataset:** The Walmart sales dataset was loaded into a pandas DataFrame for further processing.
3. **Data Inspection:** Initial inspection involved examining the dataset's structure, identifying data types, and checking for missing values.
4. **Feature Engineering:** Temporal features (week, month, year) were extracted from the Date column. Sales trends were analyzed over time, by store, and for holiday/non-holiday periods.
5. **Feature Categorization:** Categorical features were created for Temperature, Fuel_Price, CPI, and Unemployment for simplified analysis
6. **Dropping**: CPI_Cat and Fuel_Price_Cat columns were dropped

5. Choosing the Algorithm For the Project

Description for the XYZ algorithm for the project. I have chosen the XYZ algorithm for this project for the following reasons:

I have chosen the following algorithms:

1. Time Series Analysis (ARIMA, SARIMA) - will be used for Time Series Analysis to analyze on seasonal data
2. Linear Regression - will be used for linear trends on data

6. Motivation and Reasons For Choosing the Algorithm

I have chosen the XYZ algorithm for this project for the following reasons:

1. Time series analysis is ideal for handling the temporal dependencies in the sales data.
2. ARIMA models effectively capture seasonality and trends.
3. Regression models are versatile for exploring relationships between sales and external features.

7. Assumptions

The following assumptions were made in order to create the model for XYZ project.

1. Historical sales data is representative of future trends.
2. External factors (holiday, temperature, fuel price, CPI, unemployment) have a linear relationship with sales.
3. The dataset contains reliable and accurate information.

8. Model Evaluation and Technique

The following techniques and steps were involved in the evaluation of the model

1. **Data Splitting:** The dataset was divided into training and testing sets.
2. **Model Training:** The selected models were trained using the training data.
3. **Performance Evaluation:** Models were evaluated on the testing data using RMSE and MAE.
4. **Model Comparison:** The performance of different models was compared to identify the best one for sales forecasting.

The evaluation report suggests the following:

1. Linear Regression shows promising initial results, as well as, time series analysis.

9. Inferences from the Same

The model performance, inferences: Time Series Analysis helps to identify on weekly sales Fuel prices and Temp influence the most

10. Future Possibilities

The future possibilities, limitations:

Future sales are subject to change. Model improvement

11. Conclusion

The project aims to provide Walmart stores with actionable insights for inventory management and improve sales forecasting accuracy. By understanding the impact of various factors and developing a robust predictive model, the project will contribute to better decision-making and increased profitability.

12. References

- https://www.analyticsvidhya.com/blog/2022/06/time-series-forecasting-using-python/
- https://github.com/ajitsingh98/Time-Series-Analysis-and-Forecasting-with-Python
- https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html