Walmart Supply Chain Analysis

Project Name: Analysis of Walmart Weekly Sales

Report: Our project focuses on analyzing the factors behind Walmart's 45 individual store sales, specifically the incentives for revenue growth, and how we can improve them. By analyzing the data "Walmart.csv" found from Kaggle, we were able to find a correlation between store size and product type within the sales data. Furthermore, we were able to design a set of strategies targeting the different sizes of stores to boost sales.

To start with our analysis, we began the process of data preparation, cleaning, and feature engineering. There are 16 variables including (but not limited to): "Store", "Date", "isHoliday", "Dept", "Weekly Sales", "Type", "Markdown", etc. We added a data feature which is 'Size_Type' where we times the Size and the Type of the store so we can get further inside to compare to the other stores. After performing the data cleaning process, we ran a general analysis on the dataset by using several linear correlation graphs as well as identifying the weeks of the holidays through SQL query, where we successfully identified the weekly sales trends throughout the year. After identifying the trends with the graph visualization, we continued to find which variable that is affecting the 'Weekly_Sales' by calculating the p-value correlation and Pearson matrix. With the result from the previous step, we found that the size of the store and the type of product is most correlated to weekly sales. The results are 'Size', 'Type and 'Size_type' are the variables that have the most correlation towards 'Weekly_Sales'. On the other hand, other factors such as "Markdown", "Unemployment", or "CPI" have a less significant impact on weekly sales compared to the size of the store and type of product sold.

To relate size of the store and department of product on weekly sales, we created several histograms regarding average weekly sales, average sales/store, and average sales/department.

First, we found the weekly sales for each store, giving us a baseline of Walmart's average sales numbers across 52 weeks. In this analysis, we could see which weeks had an increase in sales, usually around holidays as expected. According to the graph, there are four inflection points, illustrating that there are four different clusters for weekly sales at Walmart. Then we ran a K-means clustering algorithm on four different clusters in the store to identify which department of item works best for each store. Using the results from our cluster analysis, we then created a histogram. Doing so allowed us to figure out the specific departments of Walmart's products that were boosting the weekly sales for each specific store, along with their respective store dimensions. This means that with the different sizes of Walmart stores, each store should focus on different types of products to increase their revenue, specifically that branch's respective highest-earning item type. A limitation that we ran into was that when we attempted to analyze the factors behind the sales, the results did not apply to every situation. To further improve our analysis and strengthen our hypothesis, we decided to assess different segments and figure out the specific strategy for each cluster. After identifying the best dept of items that works best for each store we split our data to train, test, and validation test then run a machine learning model for a random forest tree to predict the 'weekly_sales' from the related variable. We checked for the best depts, max features and best split in our RandomForestRegressor parameter. Because by nature we understand that RandomForestRegressor acts similarly to clustering, therefore it fits our model very well. We then check for the MSE, RMSE value as well as the WMAE values which are the accepted values. The RandomForestRegressor model generally fits the data well, as suggested by the high R-squared value of 0.979, which means about 97.9% of the variance in your target variable is explained by the model.

Overall, given the varying sizes of Walmart stores, each location should concentrate on distinct product categories to boost its income, particularly the highest-grossing item type unique to that branch. More specifically, for stores that fall within the Rank 1 & 2 clusters, they need to focus on Departments 93 and 96. For a store that sizes within the Rank 3 cluster, they should focus on Departments 37, 93, and 96. For a store that falls into the Rank 4 cluster, they should focus on Department 38. If the branches' were to follow our suggestions, they will find that their weekly sales and overall revenue will increase.