IE509 Report Write Up

Nisarg Jain

23N0454

1. Data Source: https://www.kaggle.com/datasets/suddharshan/retail-price-optimization/data
2. Various Steps:
   a. Importing various libraries for use.
   b. Used Pandas to import the dataset and viewed few rows to get an idea of the dataset.
   c. Understood Feature Description, to get an idea for what column represent what.
   d. Performed some sanity checks and summary for dataset, where in I learned how many rows are there, are there any null values, unique values etc.
   e. Did exploratory data analysis of various features, getting various insights on the distribution. Analyzed various boxplots to understand outliers.
   f. Created a correlation matrix for targeted features to get an idea for the relation.
   g. Did monthly revenue analysis and created a regression plot for monthly demand vs sales.
   h. Did feature selection for the model based on insights taken from data analysis above.
   i. Did dataset split of 80/20 and created a stratification list
   j. Trained a Random Forest Regressor Model, and calculated metrics of Performance.
   k. Plotted our prediction compared to Actual.
   l. Used Permutation Algorithm to calculate feature significance in our model.
3. Tools Used:
   a. Libraries Used: Pandas, NumPy, SKLearn, Matplotlib, Seaborn, eli5
   b. Functions and Plots Used: Histplot, Boxplot, Correlation Matrix, pd.groupby, regplot, train_test_split, RandomForestRegressor, Scatterplot, PermutationImportance.
   c. Code Written by Own: Most of the code upto Plotting and Analysis is written by myself. For model building some inspiration is taken from this notebook: https://www.kaggle.com/code/harshsingh2209/retail-price-optimization
4. Conclusion: Most of the insights from plots are generated are on my own, and analysis part is wholly done by myself. For selecting a model, inspiration was taken from the notebook mentioned above. Strategies and feature selection was done based on my decisions.