



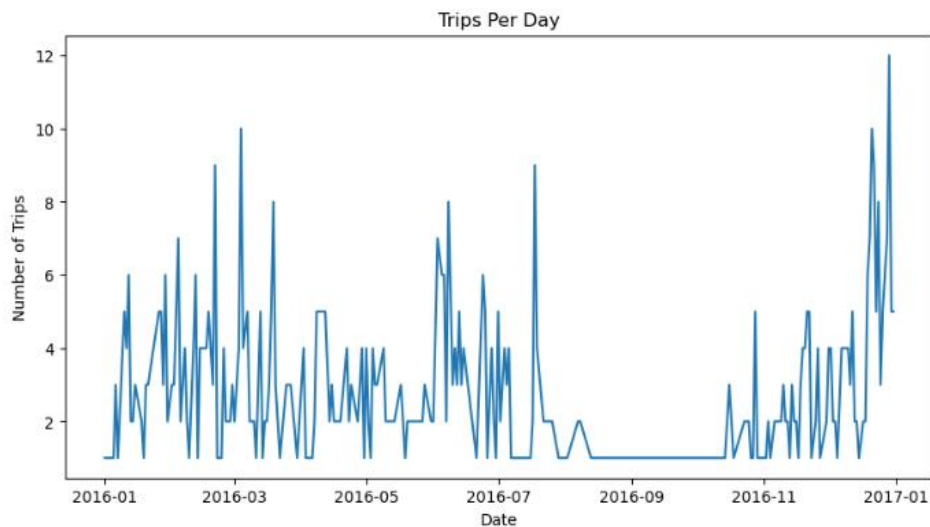
ICS 474: Big Data Analytics

[Project Visualization Document](#)

In this document, I will provide the graphs and diagrams of the work I have done in the project which is (Histogram, Boxplot, Count plot, Heat Map, and Bar Chart), each graph or chart is labeled clearly with titles axis labels and legends.

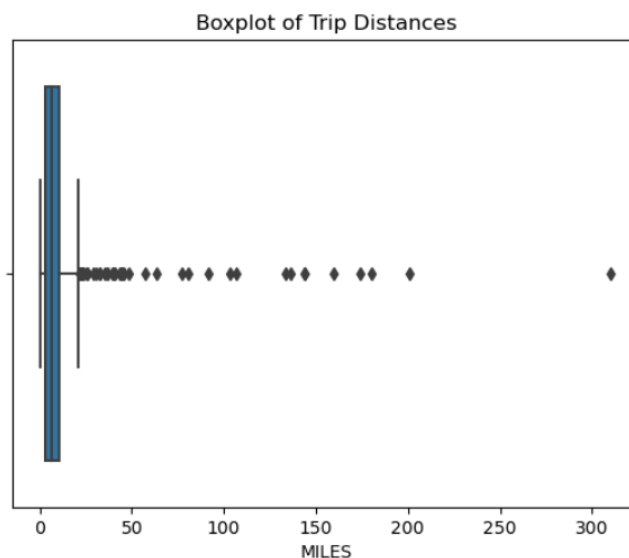
1- Distribution of Trip Distances (MILES)

This histogram shows the distribution of trip distances. The majority of trips are short, with most distances falling under 20 miles. The data is right-skewed, indicating a long tail of longer trips, with very few trips exceeding 50 miles



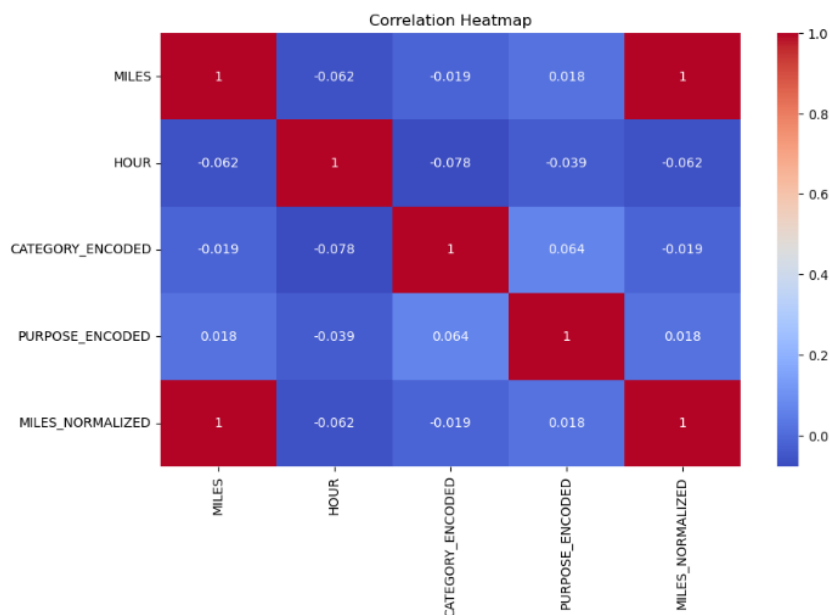
2- Boxplot for (MILES)

This boxplot highlights outliers in the `MILES` column. While most trips fall within a reasonable range, some trips exceed 100 miles. These were identified and treated as outliers during preprocessing.



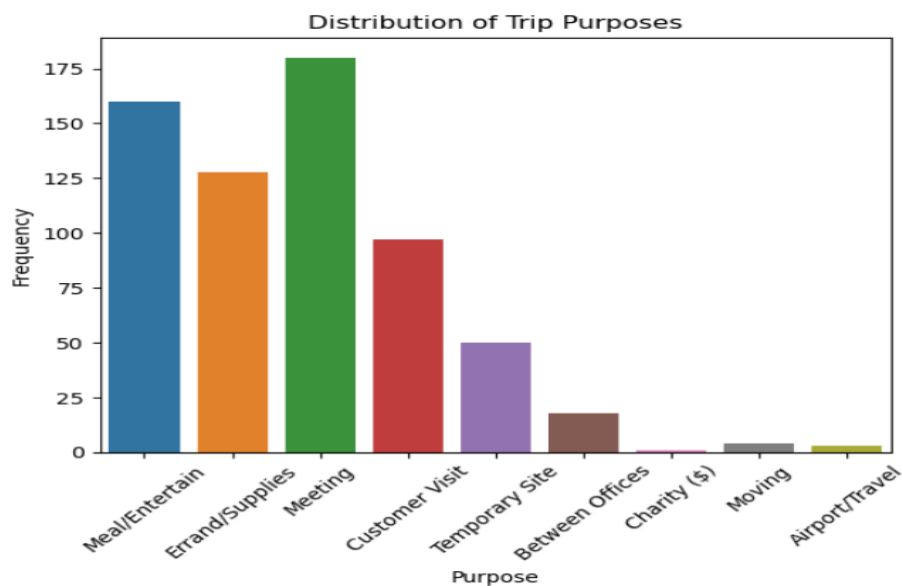
3- Correlation Heatmap

This heatmap visualizes correlations between numerical features. The `MILES` column shows a strong correlation with the target variable `DISTANCE_CATEGORY`, while other features, such as `HOUR` and `CATEGORY_ENCODED`, have weaker relationships



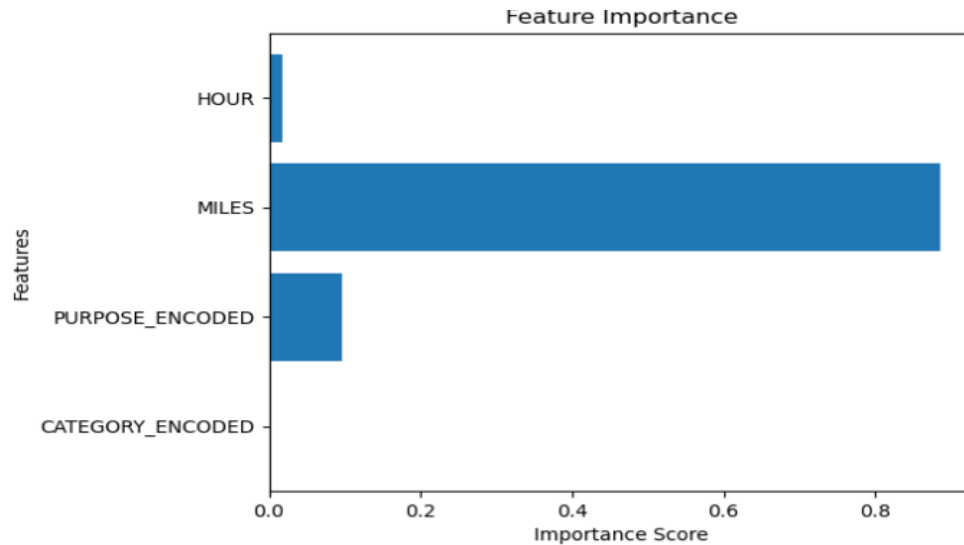
4- Count Plot for (PURPOSE)

this count plot shows the frequency of different trip purposes. The most common purposes are 'Meeting' and 'Customer Visit,' indicating that the dataset is primarily business-oriented. Less frequent purposes include 'Charity (\$)' and 'Moving'.



5- Feature Importance Bar Chart

This bar chart displays feature importance scores from the Random Forest model. `MILES` is the most important feature for classifying trips into distance categories, followed by `PURPOSE_ENCODED`. `HOUR` and `CATEGORY_ENCODED` have minimal impact on the model's performance.



6-Boxplot for Predictions

This boxplot compares the model's predicted `DISTANCE_CATEGORY` with actual trip distances (`MILES`). The model accurately categorizes trips into `Short`, `Medium`, and `Long` based on their distance. `Short` trips cluster below 5 miles, `Medium` trips range between 5–20 miles, and `Long` trips extend beyond 20 miles.

