UBER

PYTHON PROJECT

UBER RIDE ANALYSIS

JASIKA GUPTA

jasikagupta04 Jasika Gupta jasikagupta04@gmail.com

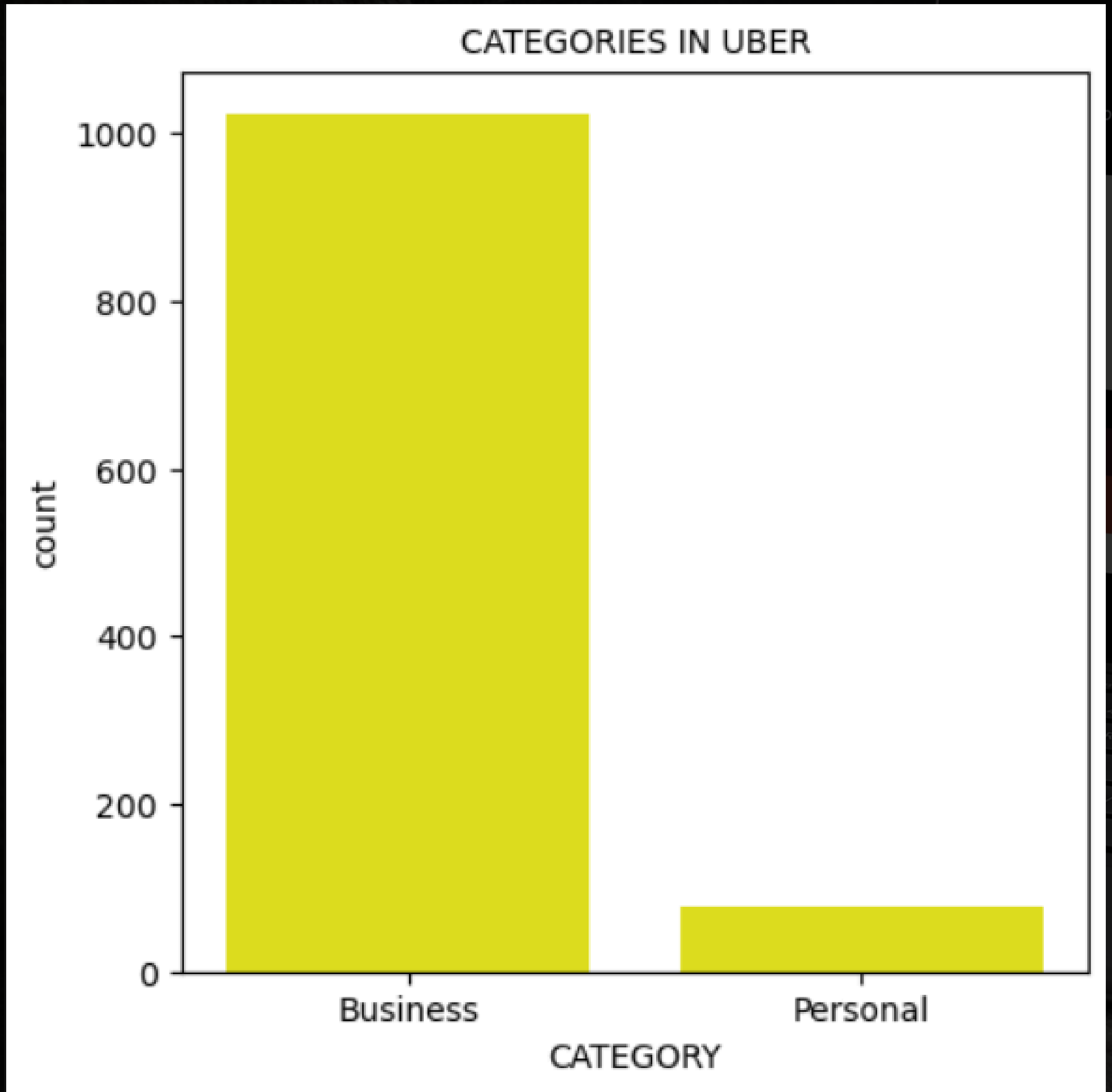
INTRODUCTION

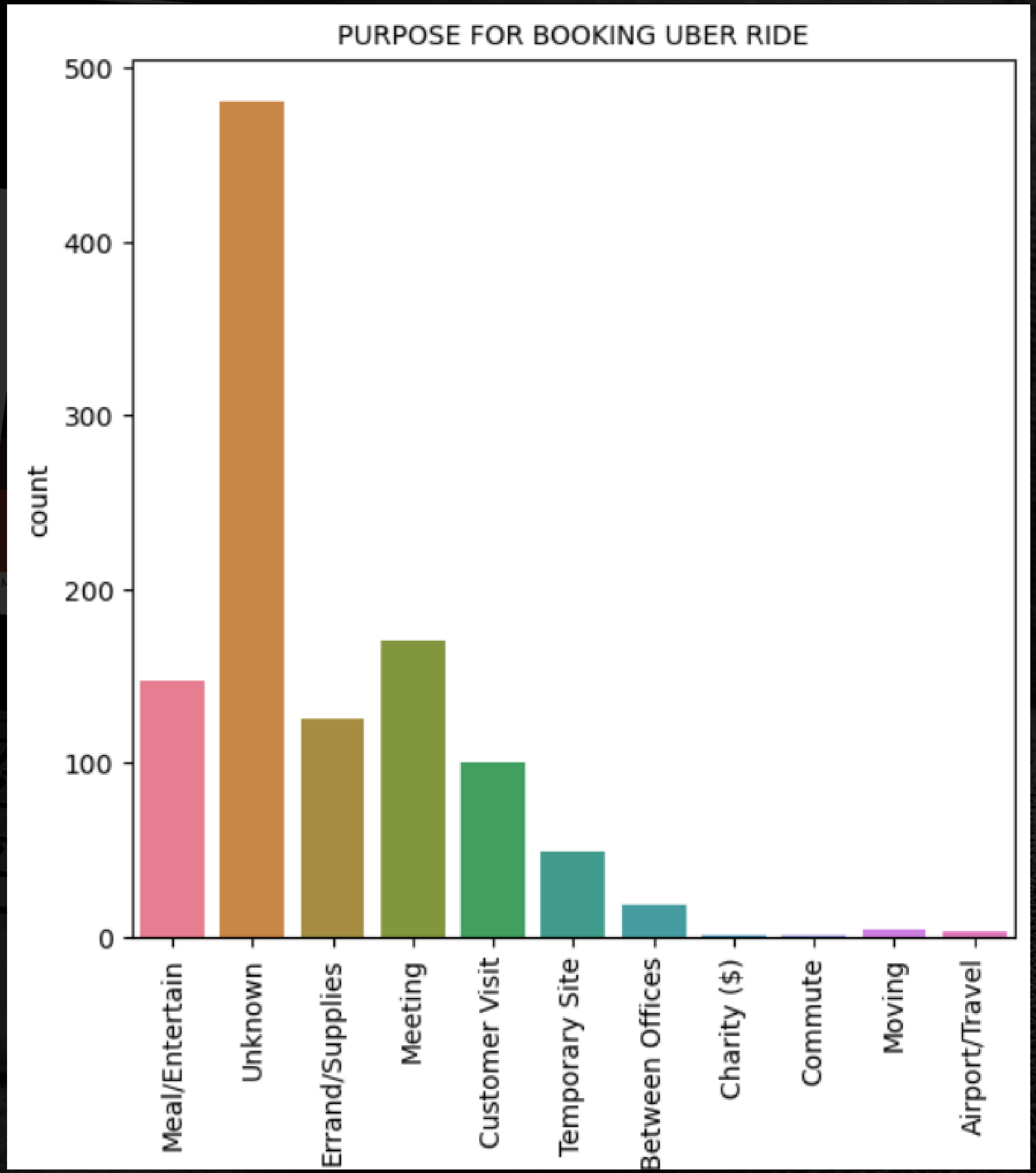
This project focuses on analyzing a dataset of **Uber rides** containing over **1,100 entries**. The dataset includes key information such as ride categories, start and stop locations, mileage, and ride purposes. The primary objectives of this analysis were:

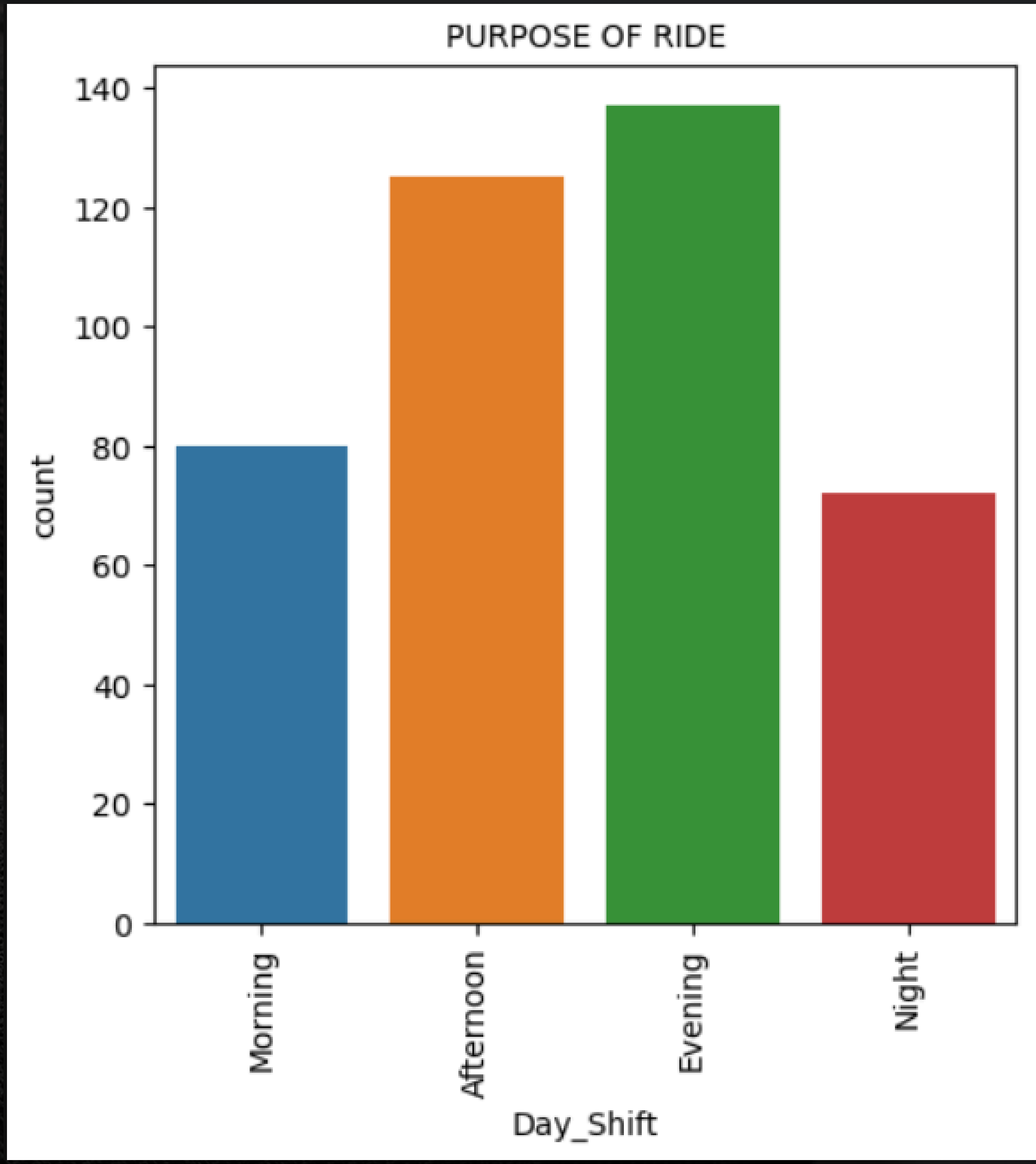
1. **Data Cleaning:** Handling missing values, ensuring data consistency, and preparing the dataset for analysis.
2. **Data Processing:** Extracting meaningful information, such as ride durations, shift patterns, and trends across categories.
3. **Data Visualization:** Using various charts and graphs to uncover insights using Pandas, Numpy, Seaborn and matplotlibrary.

Through this project, we explored the journey of transforming raw data into actionable insights, highlighting key trends in ride-sharing patterns.

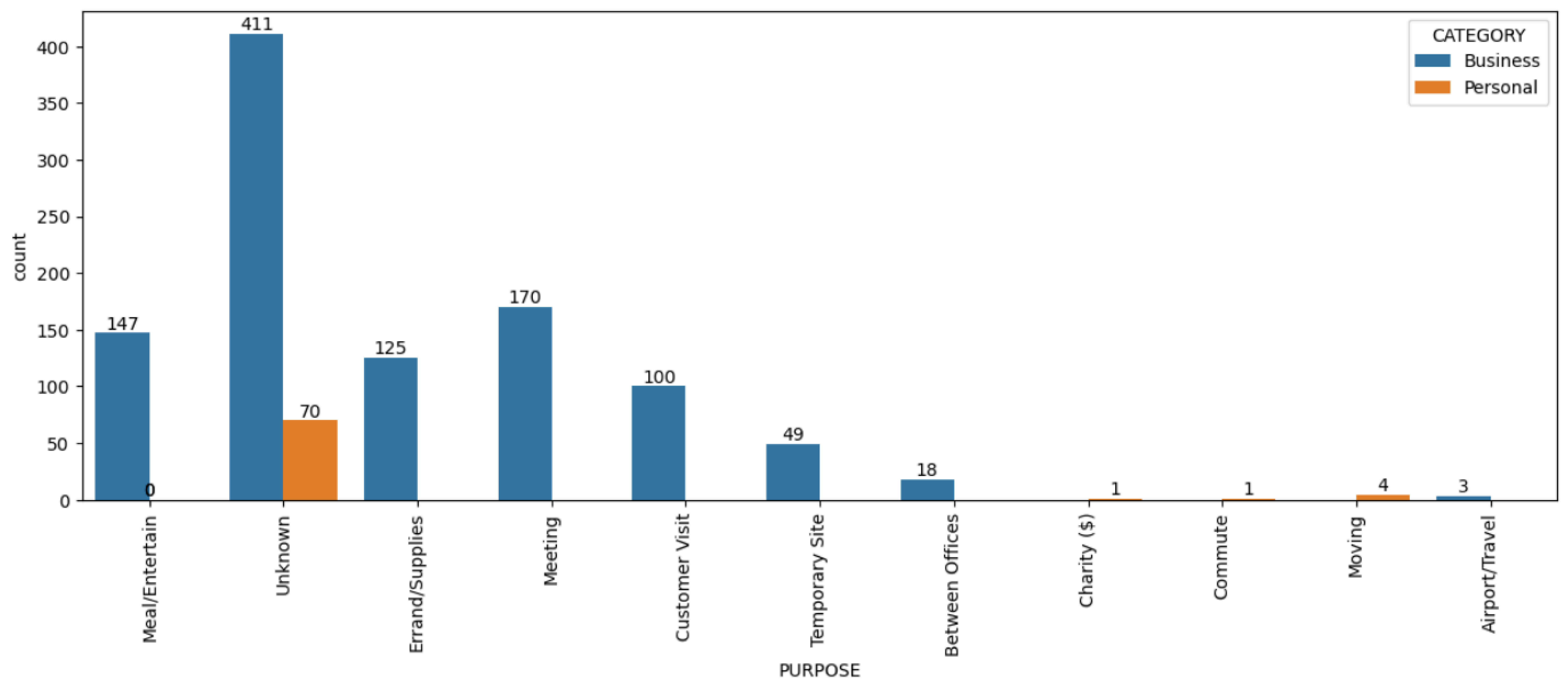
DATA VISUALIZATION (1/3)



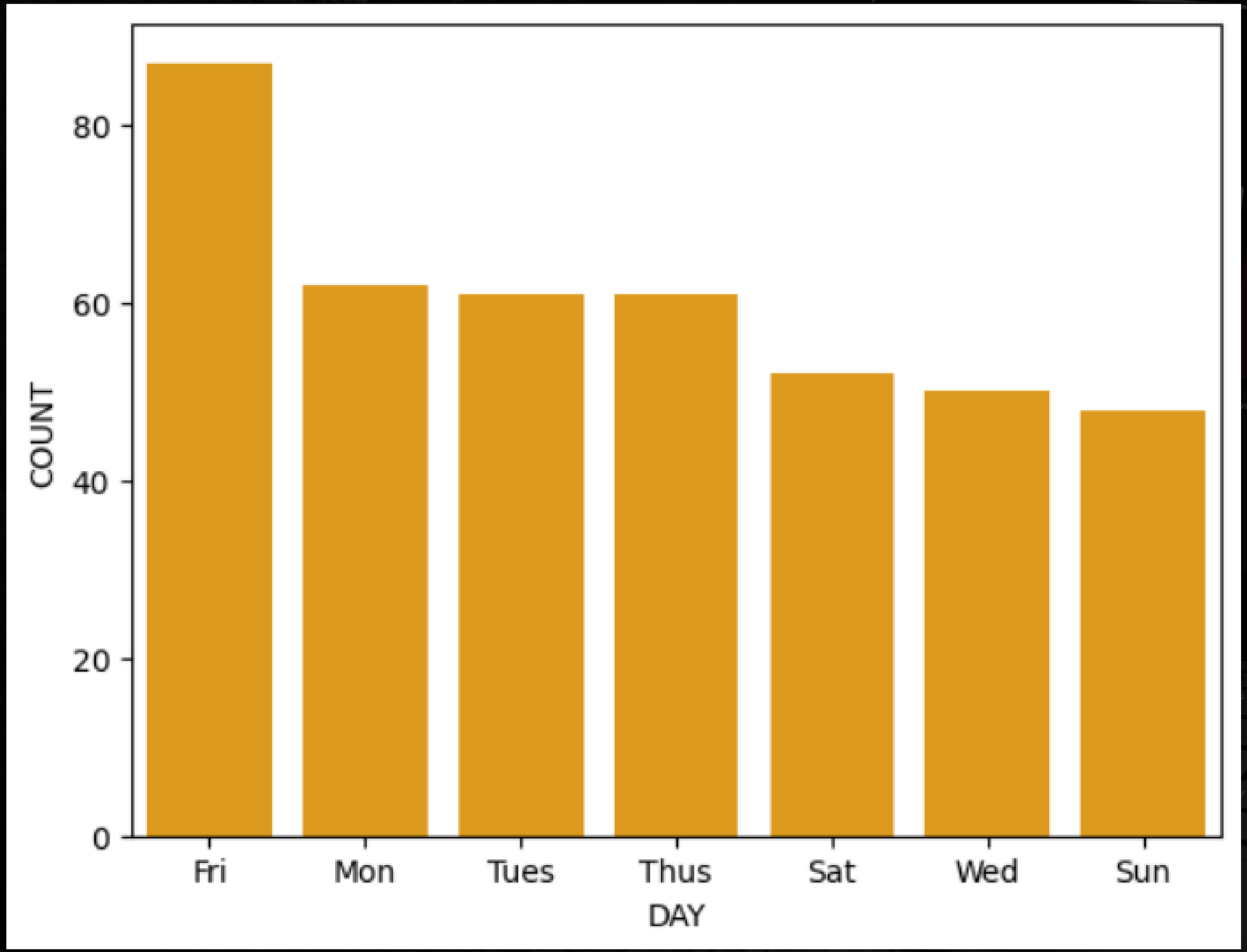


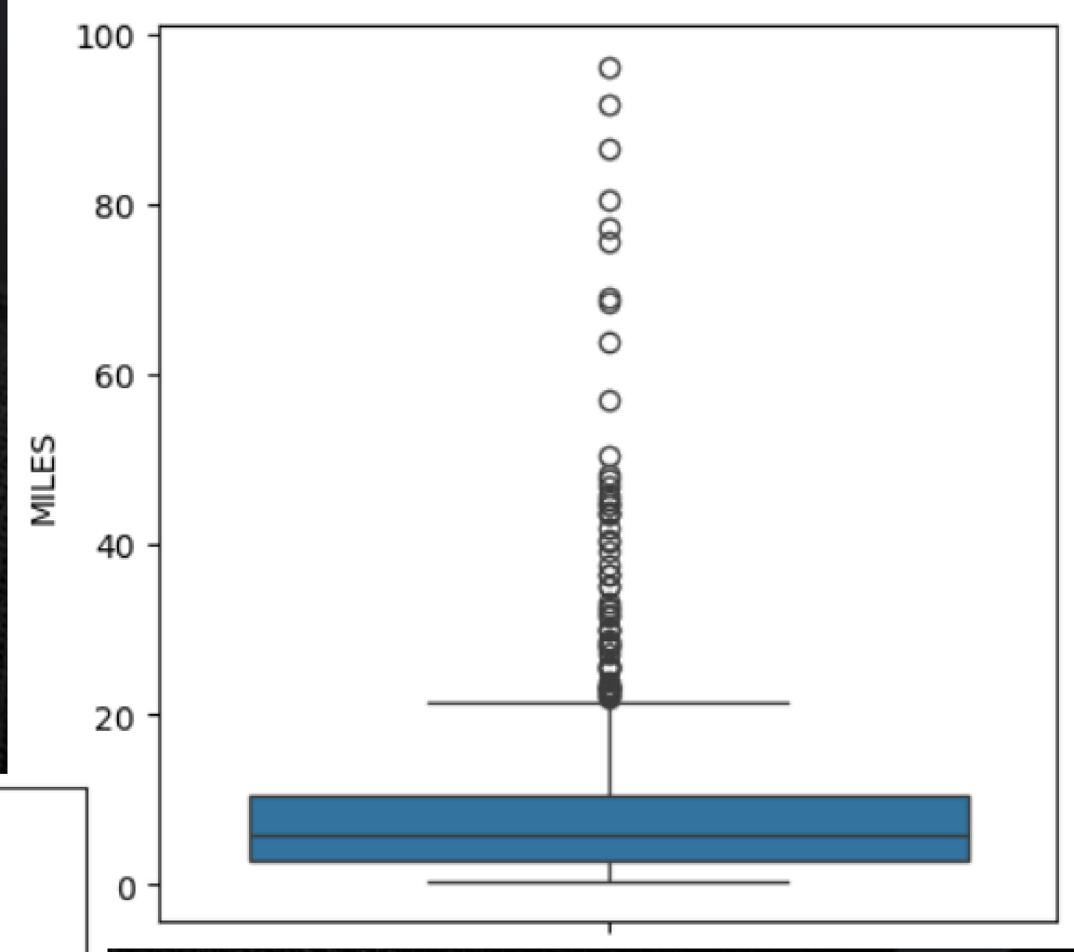


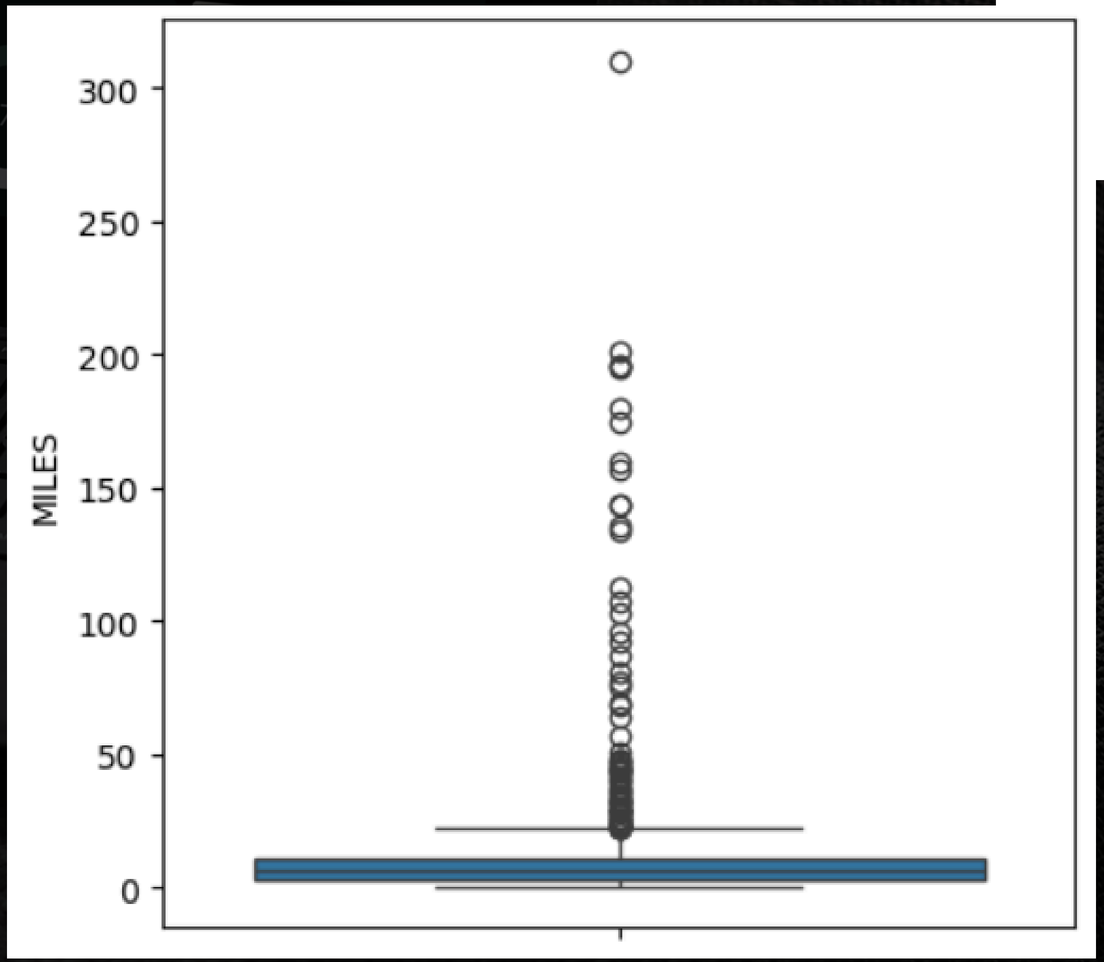
DATA VISUALIZATION (2/3)



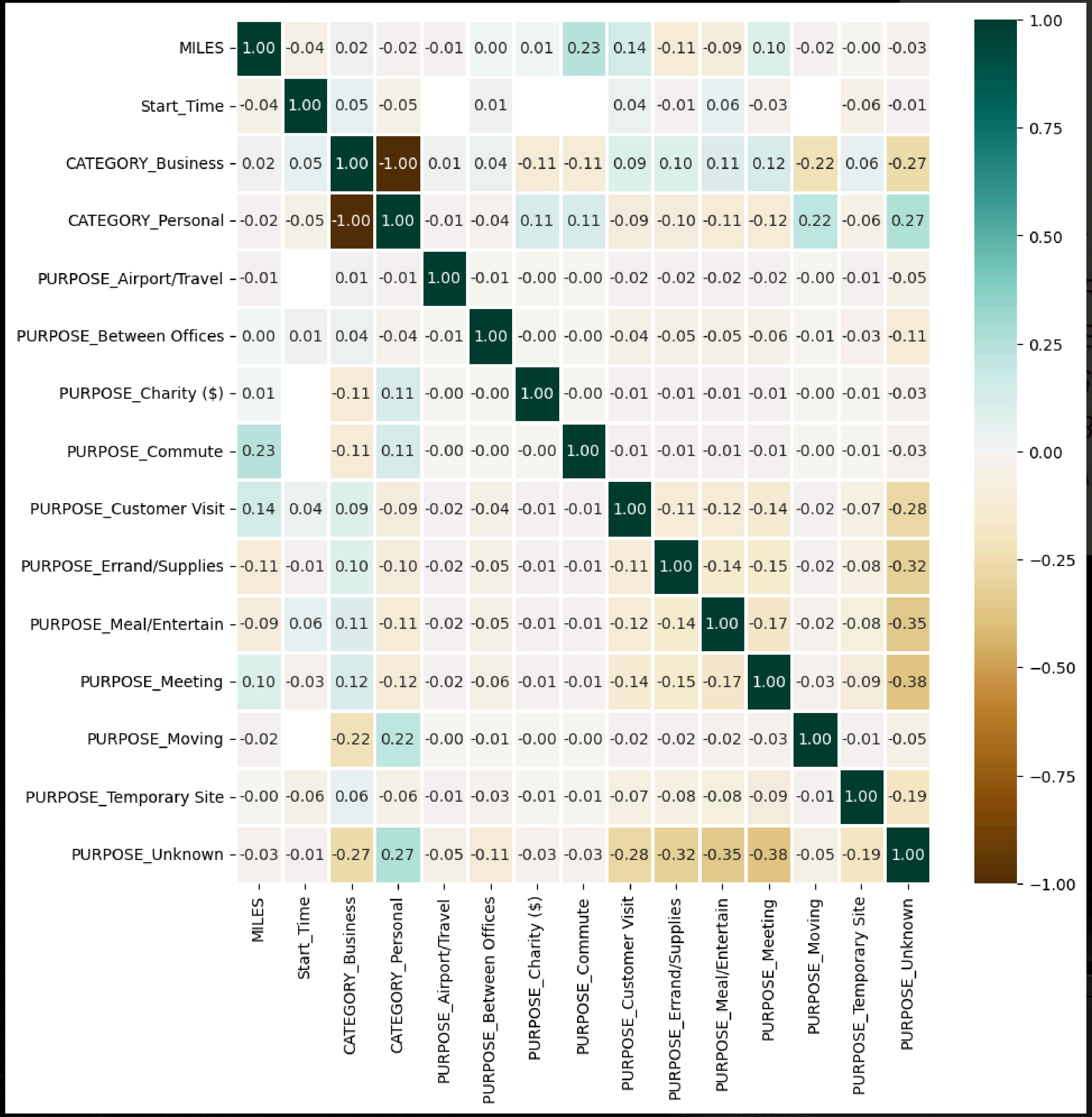
DATA VISUALIZATION (3/3)







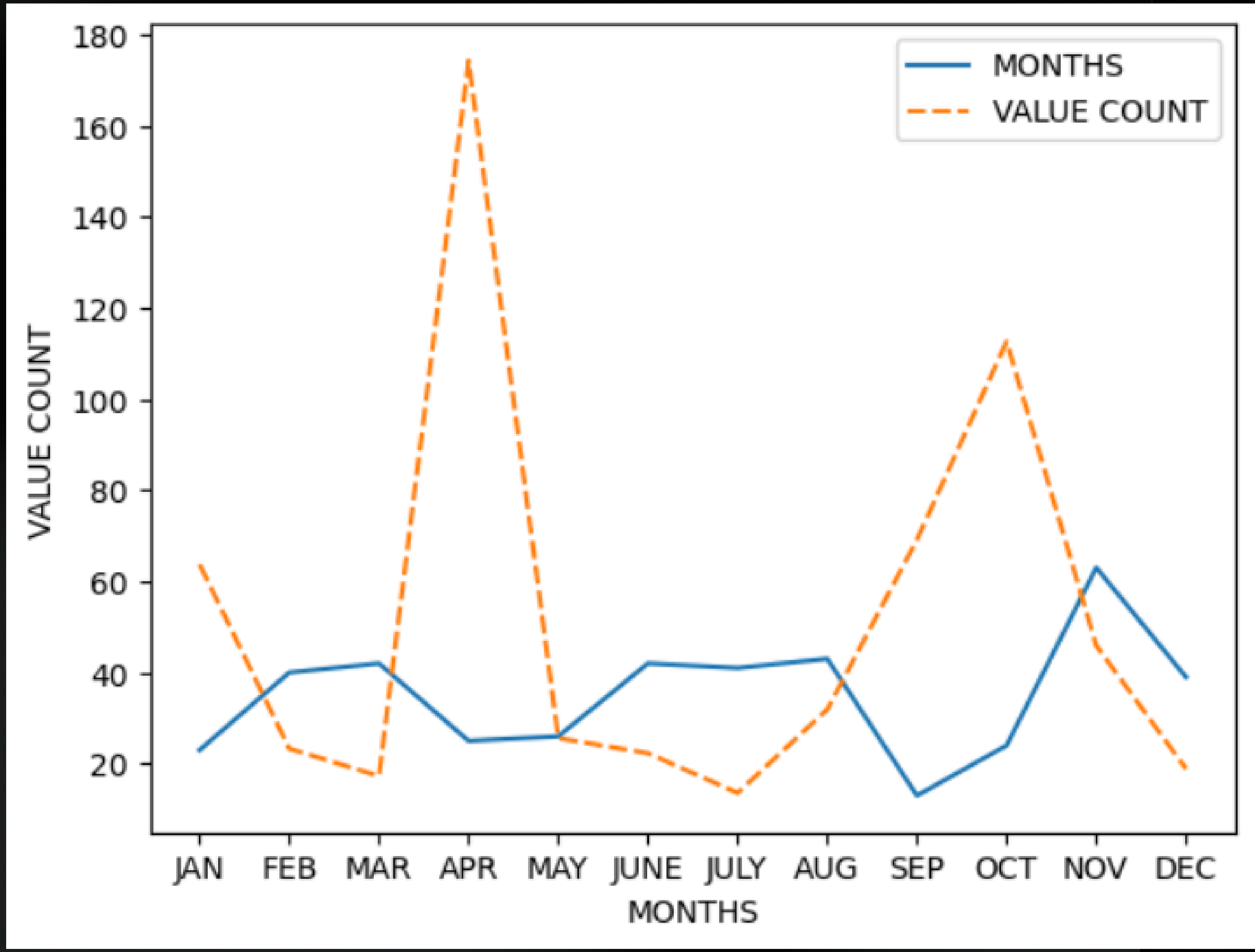
CORRELATION HEATMAP



OBSERVATIONS:

- **MILES:** A strong postive correlation (1.00) with itself, as expected.
- **TIME:** No strong correlation with most variables.
- **PURPOSE Variables:**
  - PURPOSE_Between Offices and PURPOSE_Temporary Site show very weak correlations with other variables, meaning they might not influence mileage or time significantly.
  - PURPOSE_Meal/Entertain shows weak correlations across the board, meaning it doesn't strongly affect trip time or mileage.
- **CATEGORY and PURPOSE Interactions:**
  - CATEGORY_Business is positively correlated with PURPOSE_Customer Visit, PURPOSE_Meeting, and PURPOSE_Errand/Supplies, aligning with common business activities.
  - CATEGORY_Personal has slight correlations with non-business purposes.
- **Negligible Correlations:**
  - Most variables have weak correlations (close to 0) with mileage and time, indicating that no single variable strongly predicts these metrics in the dataset.

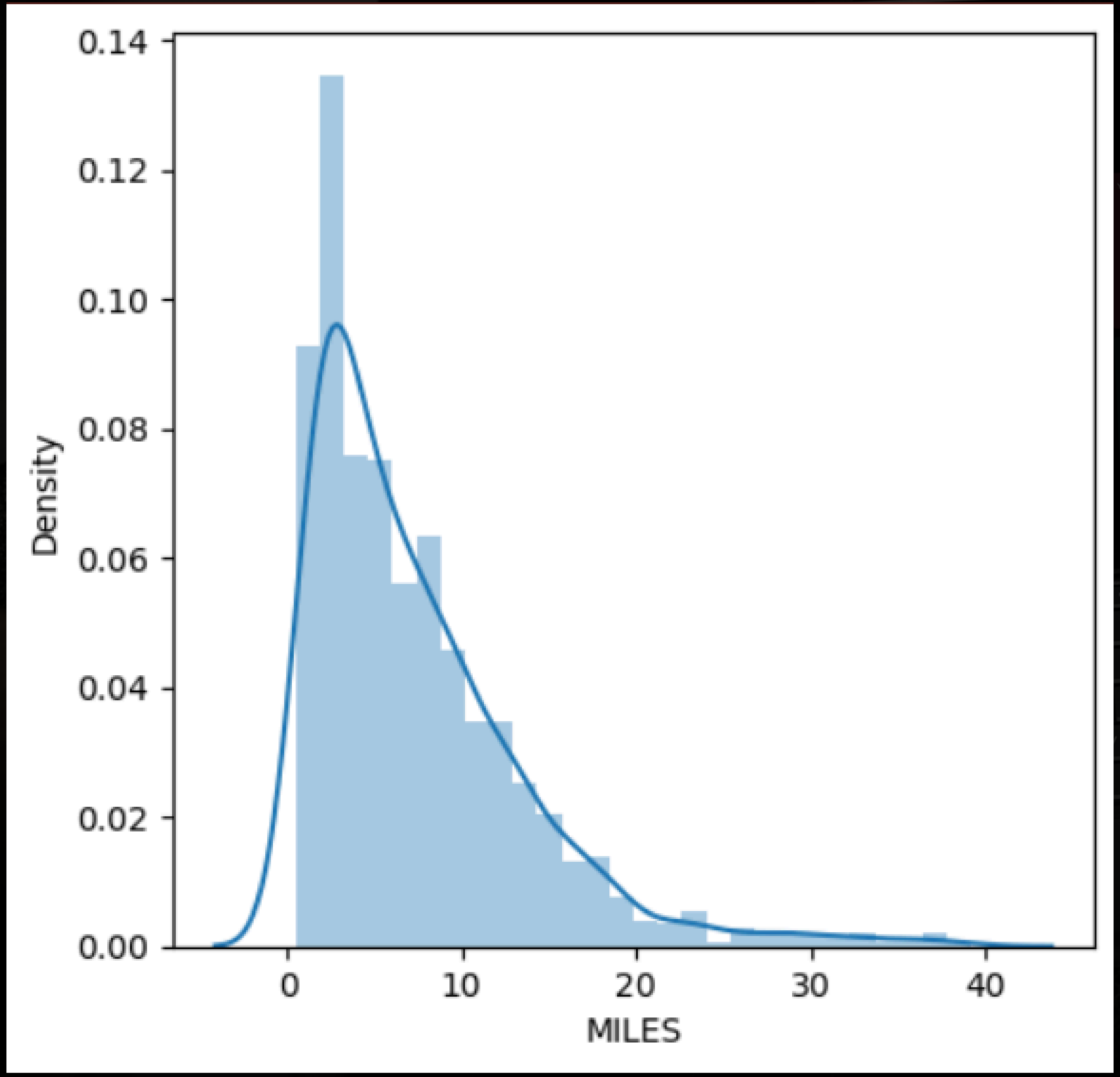
MONTHLY VALUE COUNT



OBSERVATIONS:

- **Peak in April:** There is a significant spike in value counts in April, indicating an unusually high activity or number of events during that month.
- **Steady Trends:** Months like February, May, June, and August show relatively stable value counts without significant fluctuations.
- **Decline After April:** Following the peak in April, the value counts drop sharply, suggesting a seasonal or event-driven factor influencing this pattern.
- **Rise in September and October:** There is another noticeable increase in value counts during September and October, although not as dramatic as April's peak. This suggests some recurring or periodic activity during these months.
- **Low Activity in July and November:** The months July and November appear to have low value counts, indicating these could be off-peak months for whatever the data represents.

SALES LAST MONTH



OBSERVATIONS:

- **Distribution:** The distribution is skewed to the right, indicating that a majority of the data points fall within lower mileage values. There are fewer observations with higher mileage values.
- **Central Tendency:** The mean (center of mass) is likely to be lower than the median (middle value) due to the right-skewed nature of the distribution.
- **Spread:** The distribution appears to have a moderate spread. We can quantify this using measures like standard deviation or interquartile range.
- **Potential Outliers:** While not explicitly visible in the plot, there might be a few potential outliers on the higher end of the mileage range, contributing to the right-skewed shape.

THANK YOU

jasikagupta04

Jasika Gupta

jasikagupta04@gmail.com