# Travel Service Analysis

ARBAJ MOMIN

Pre-processing

I have imported all the necessary data manipulation and visualization libraries. Then, I loaded the data which was downloaded from Kaggle. After that, I loaded it using the read_csv function by providing the path. Next, I checked its data type, null values, and set appropriate data types. Additionally, I created a list of numerical columns in case we need them, so we can directly call them. Moreover, we added a "day name" column and a "month" column to generalize the data. We also added an "hours" column because the trip duration is in seconds, making it difficult to interpret. **After checking the statistical parameters, I noticed that trip duration is in seconds and there are instances of very short trips, such as 1 or 2 seconds, which are not possible. Before proceeding, I removed rows with trip durations less than or equal to 59 seconds. This accounted for just 0.5% of the data, so I removed it to clean the dataset**. Furthermore, we checked the statistical parameters and description of our Data Frame to get an overview of the data.

Trend Analysis

We delved into both monthly and daily pickup trends to gain a comprehensive understanding of the data. On a monthly basis, we examined the frequency of pickups for each month, identifying March as the busiest month, closely followed by April, and January as the least busy. This seasonal pattern suggests heightened demand during the summer months, coinciding with the end and start of the financial year. To visualize this trend, we plotted the pickup counts for each month using a bar chart. Additionally, we analyzed the distribution of pickups across different days of the week. We found that Thursday experiences the highest number of pickups, closely followed by Friday, while Sunday and Monday have the fewest pickups. This daily trend reflects potential variations in transportation needs throughout the week, with Friday serving as a peak demand day. To visualize these trends, we plotted pickup counts for each day of the week using appropriate charts
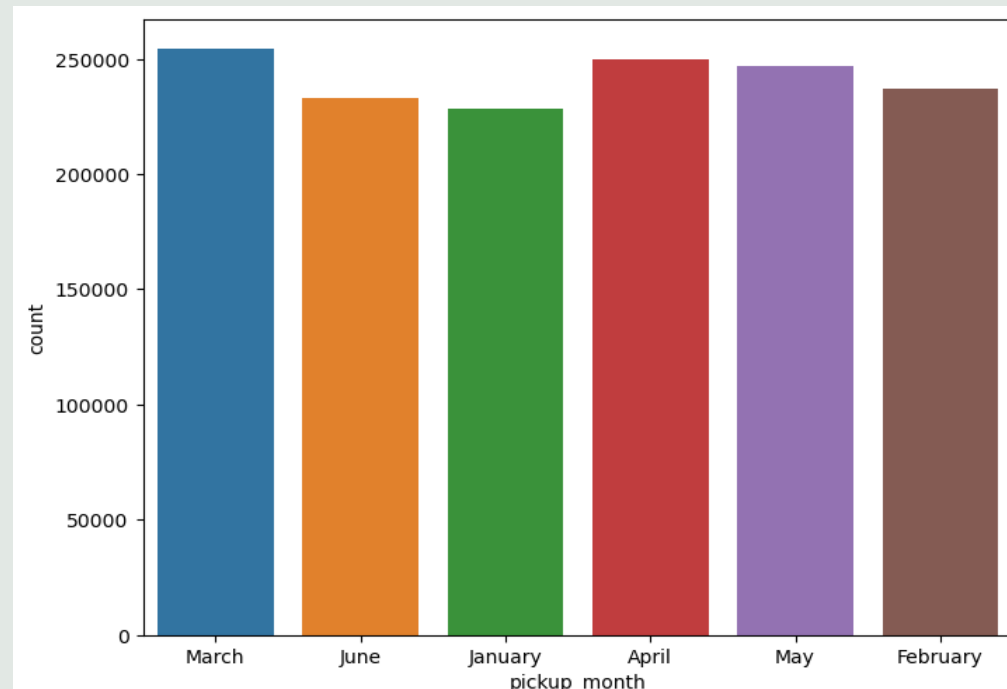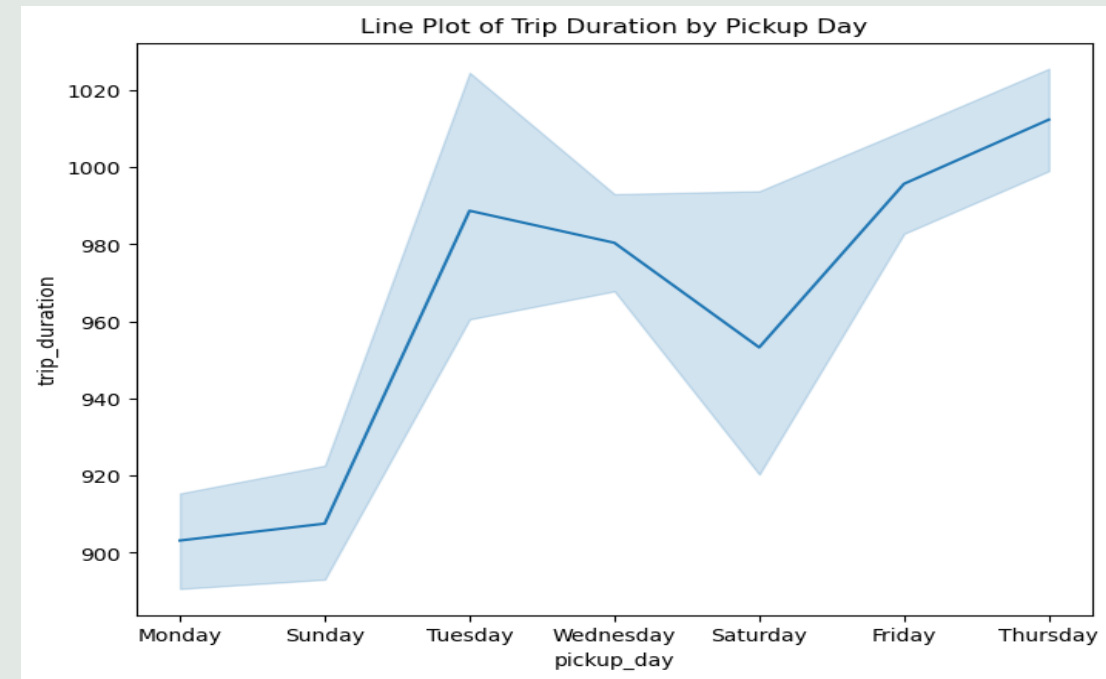
Traveler Group Sizes analysis

+ We examined the distribution of traveler group sizes by plotting the frequency of passenger counts using a bar chart.

+ The analysis reveals that most trips are undertaken by individuals traveling alone. Trips with multiple passengers are less common, indicating a preference for solo travel among passengers in the dataset. This insight suggests potential opportunities for transportation services to cater to the needs of solo travelers, such as offering smaller vehicle options or implementing solo traveler discounts.
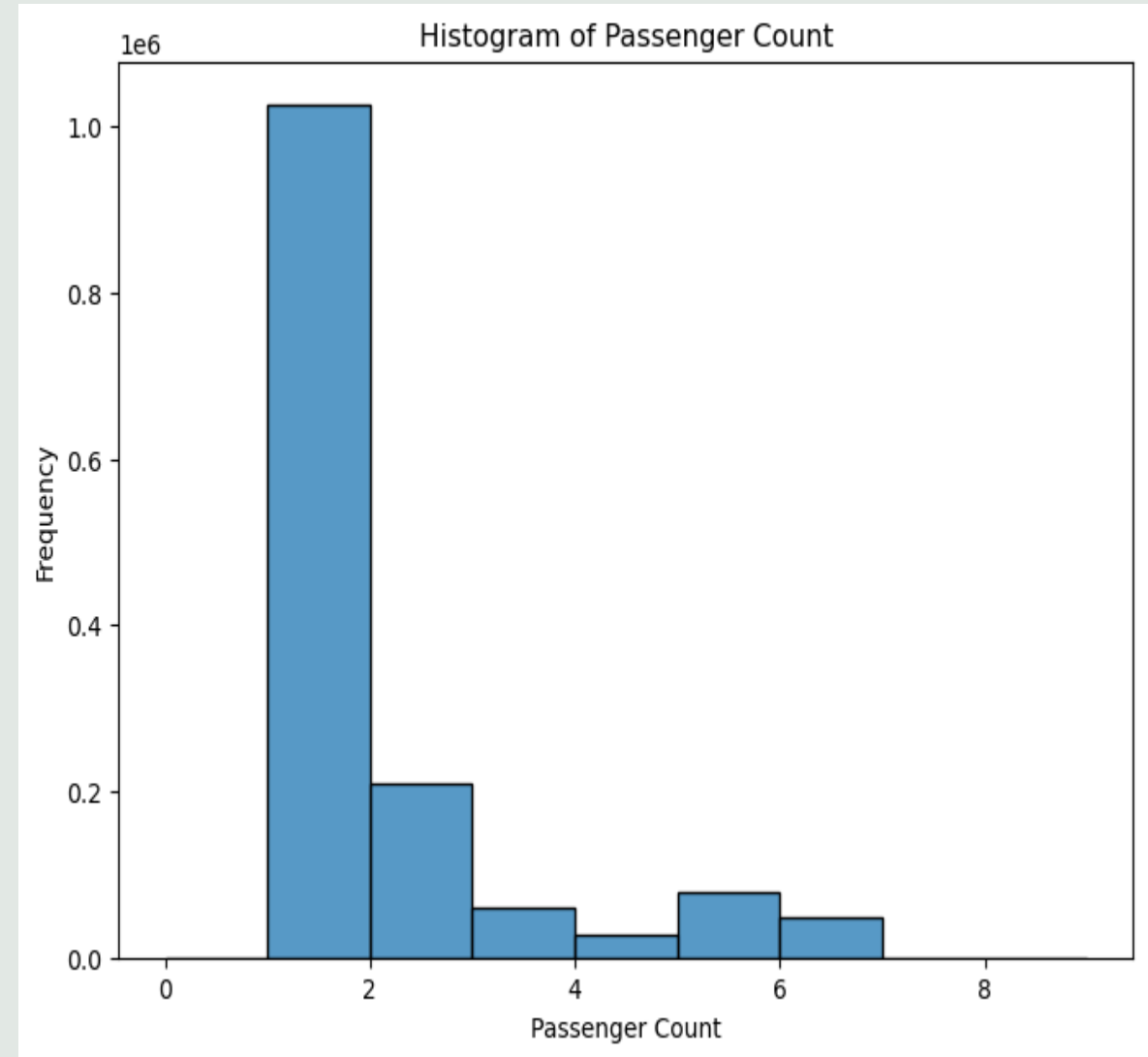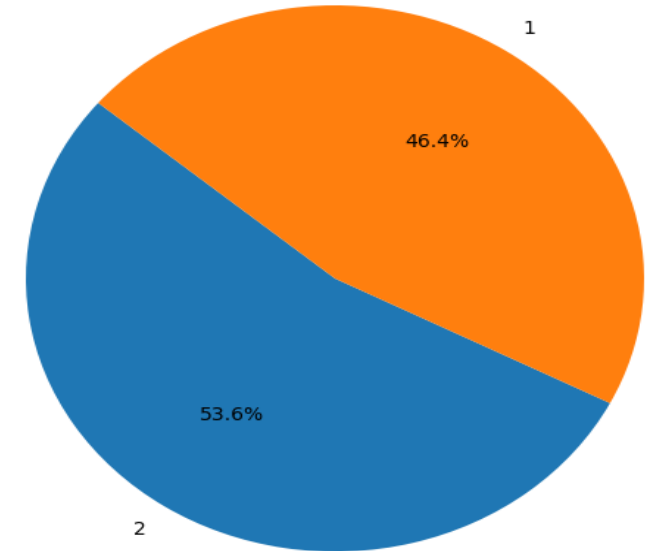
Peak Travel Time Analysis

We categorized pickup times into preferred time slots by dividing the hours of the day into specific intervals. Each pickup was assigned to one of these intervals. We then calculated the proportion of trips occurring in each time slot to understand travel patterns. To visualize the distribution, we created a bar chart showing the frequency of pickups for each time slot.

The analysis reveals that the most preferred time slot for travel is between 18:00 and 21:00, accounting for nearly 19% of trips. This is followed by the time slots between 15:00 and 18:00, and between 12:00 and 15:00, which represent about 17% and 16% of trips, respectively. Conversely, the least activity is observed during the early morning hours, between 01:00 and 06:00, which accounts for approximately 8% of trips. This insight indicates peak travel times in the evening and highlights a significant drop in travel during early morning hours.
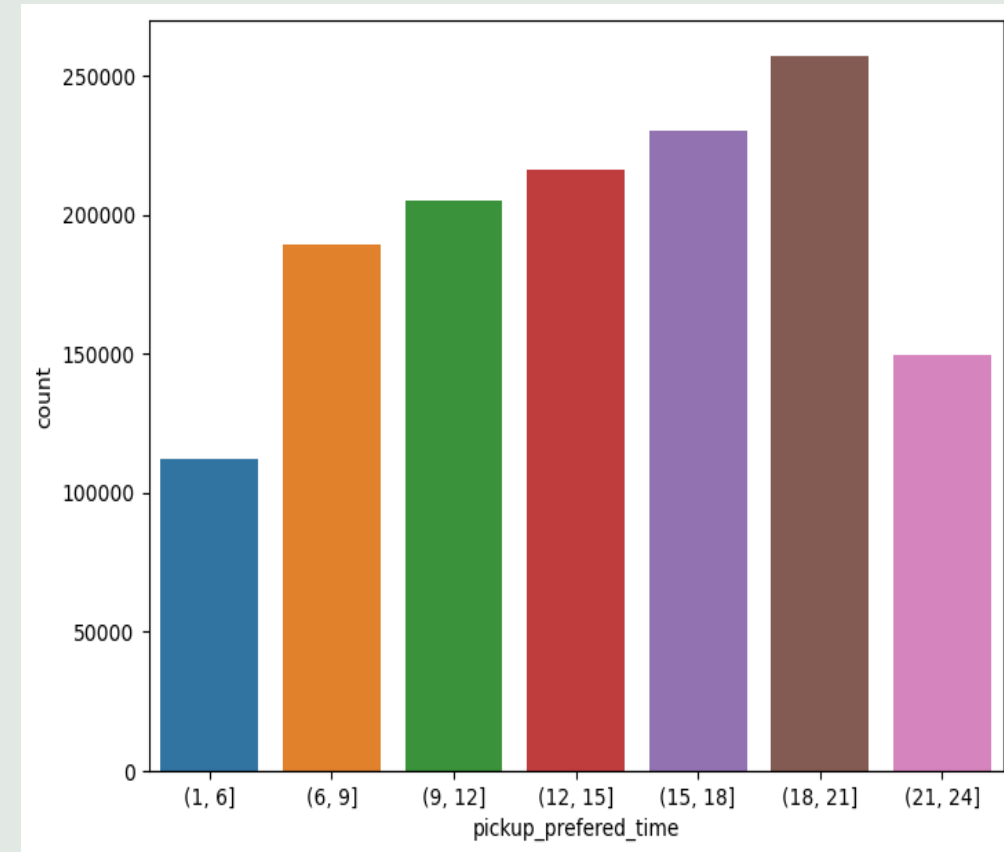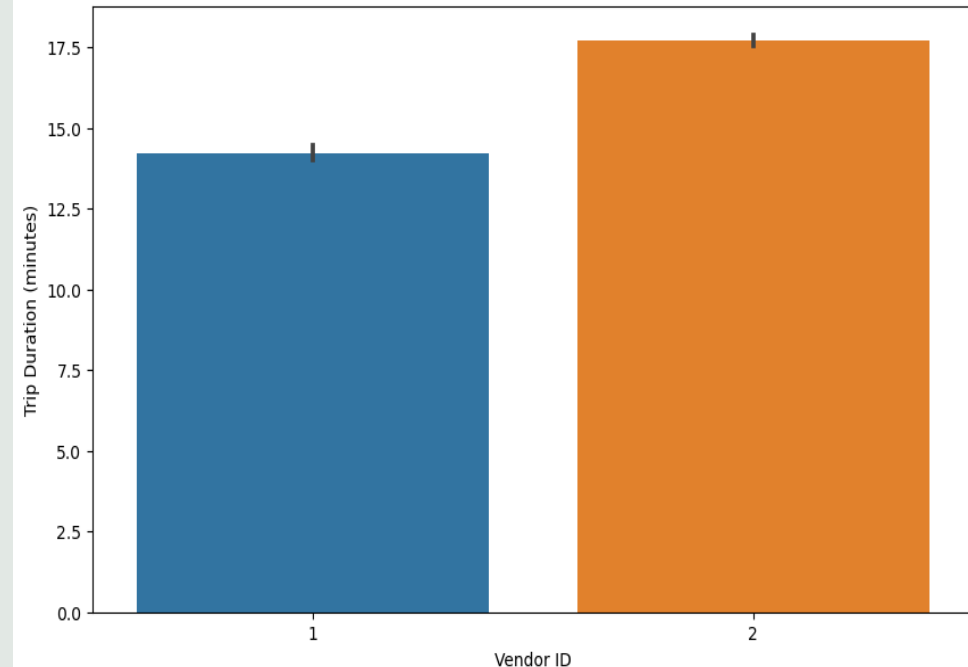
Vendor Analysis

+ We analyzed vendor performance by counting the number of trips associated with each vendor and examining the average trip duration for each vendor. First, we counted the occurrences of each vendor ID to determine the number of trips per vendor. Then, we visualized the distribution of trip counts using a bar chart. Additionally, we compared the average trip durations for each vendor by converting trip duration from seconds to minutes and plotting this information in a bar chart.

+ The analysis reveals that vendor ID 2 has the highest number of trips, with 776,872 trips, compared to vendor ID 1 with 673,332 trips. Additionally, vendor ID 2 also has the highest average trip duration. This indicates that not only does vendor ID 2 handle more trips, but the trips tend to be longer on average compared to vendor ID 1. This insight can help in understanding vendor performance and potentially guide decisions related to vendor partnerships and service optimization.

+**Thank You.!**