Dorcas Munoko

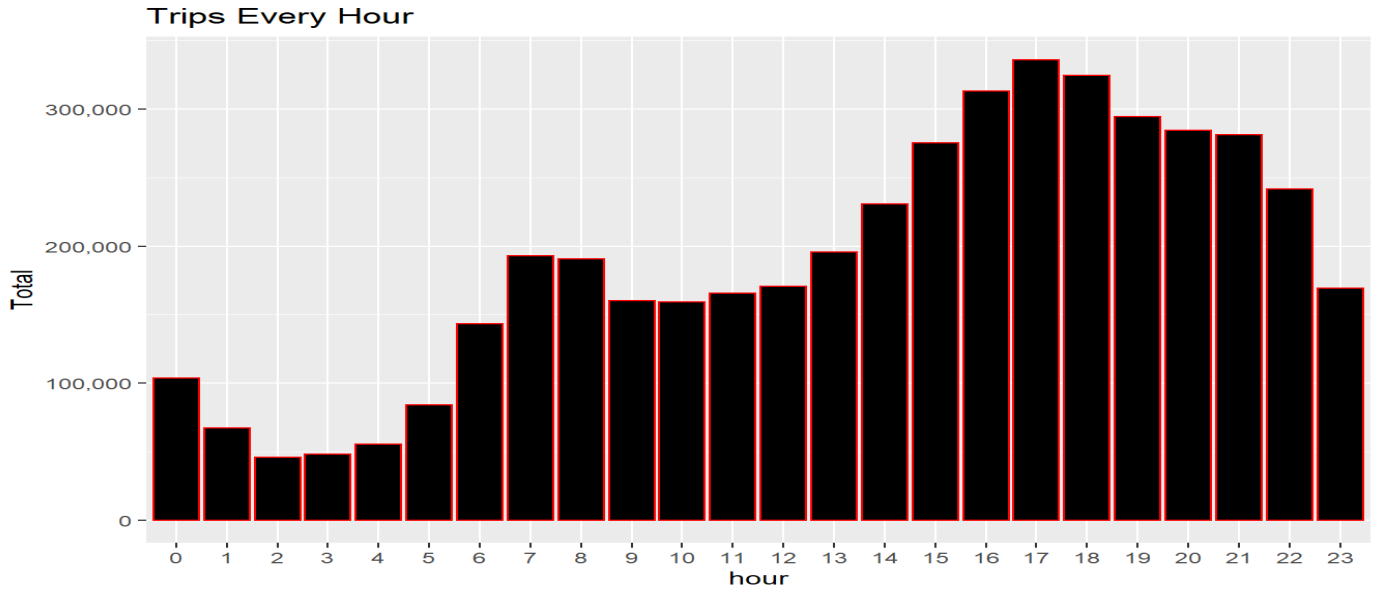
Uber Analysis

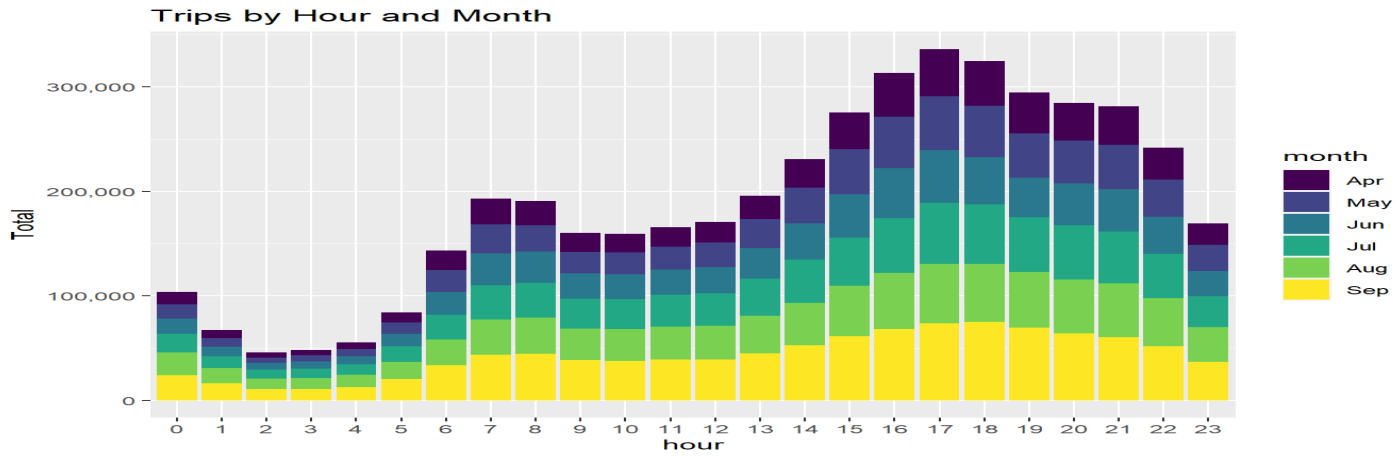
Introduction

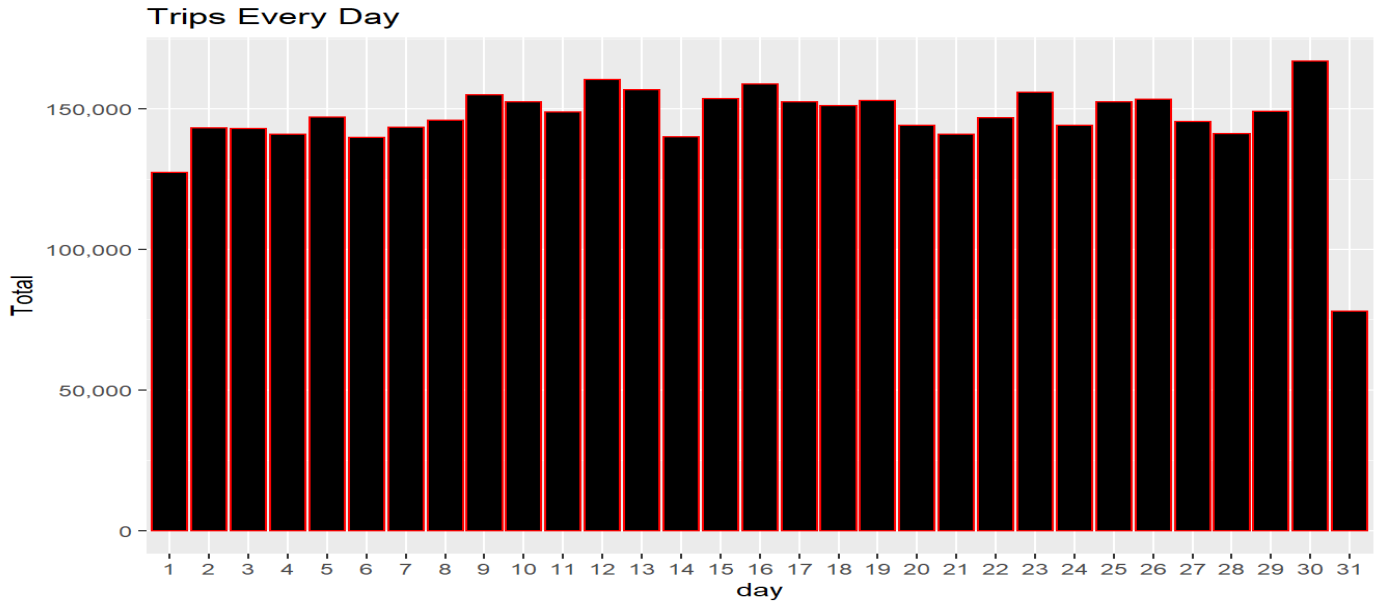
The aim of the project was to analyze Uber pick ups in New York city dataset and create appealing visualization plots. The dataset contained 4,534,327 observations with main features including date and time, latitude, longitude, and base. The first step taken in data preprocessing was to merge 6 dataframes ranging from April to September each containing daily observations. Appropriate formatting of Date.Time column was also conducted to create factors of time objects like day, month, year, hour, minute, and second.

Description of Features

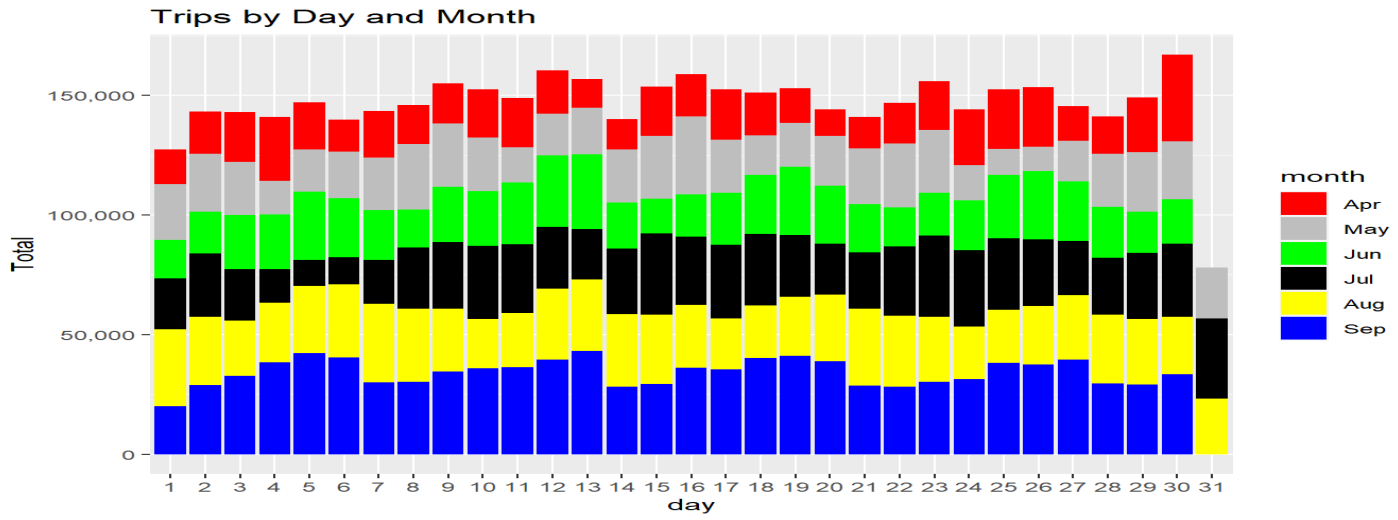
The number of trips per day were plotted. It was observed that the number of trips is higher in the evening around 5:00 and 6:00 PM. Data was also plotted based on every day of the month. It was observed that 30th of the month had the highest trips in the year which was mostly contributed by the month of April.

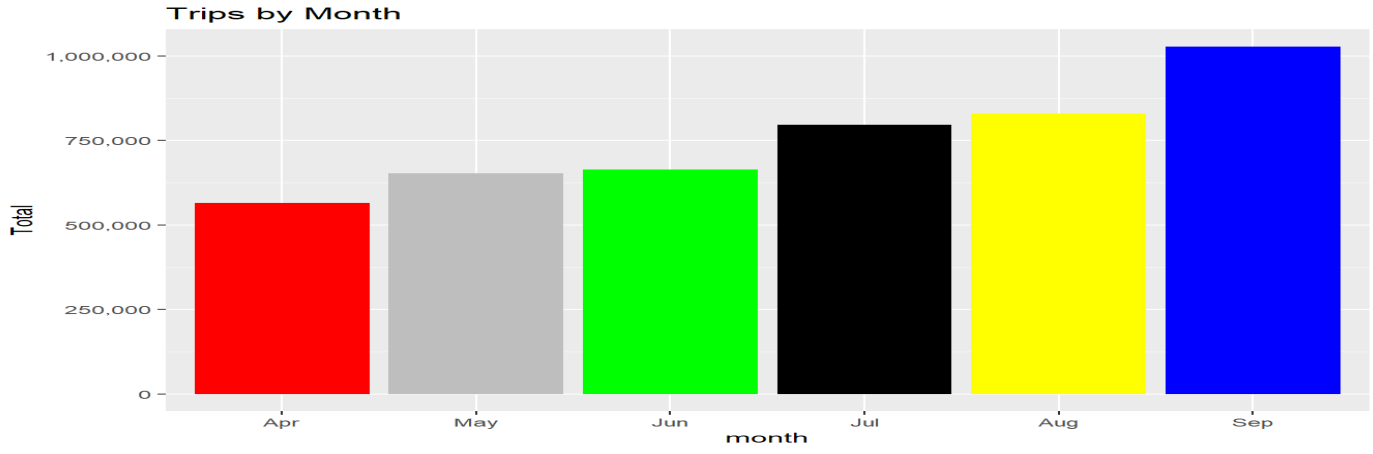


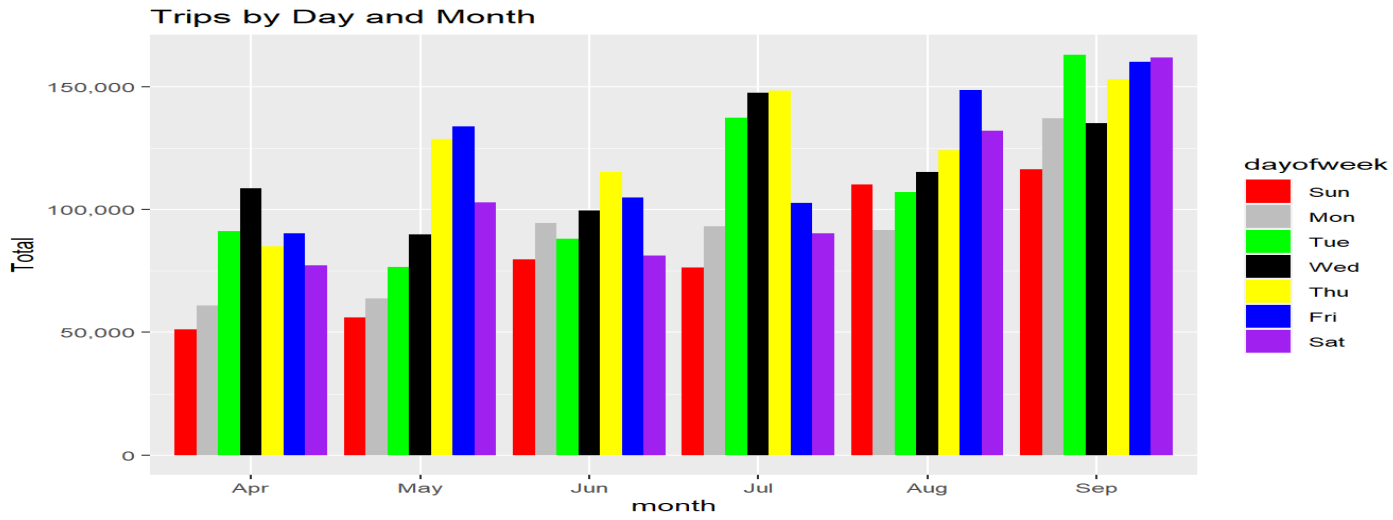




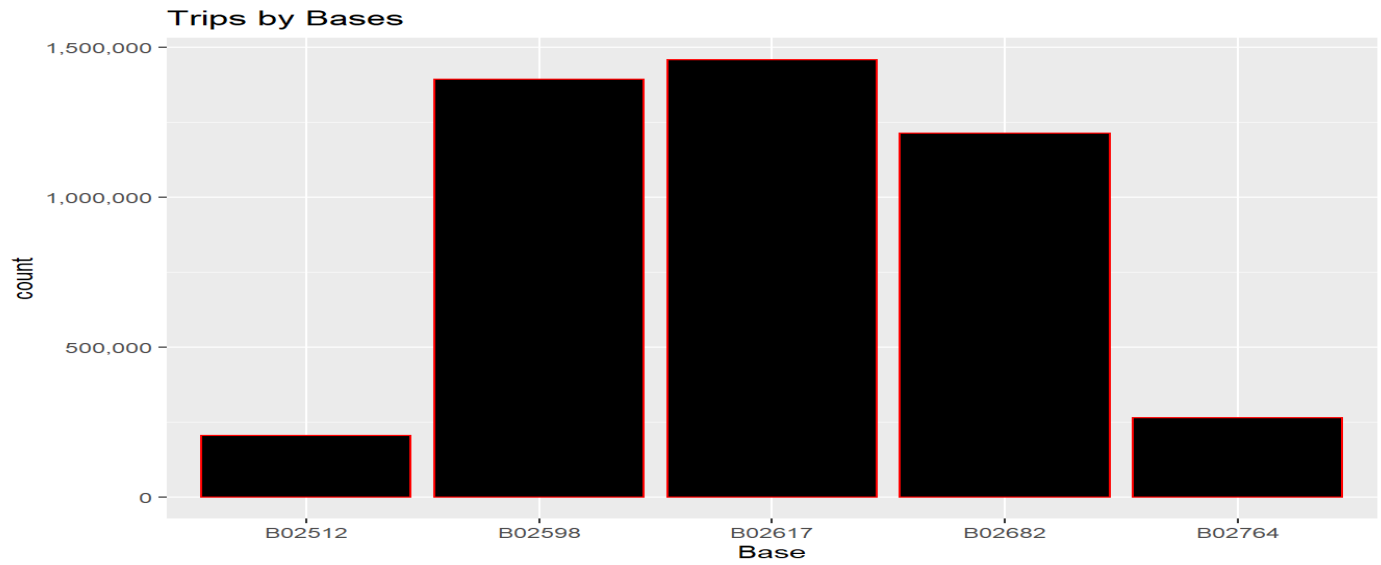
Further analysis was carried out to determine the number of trips that took place each month of the year. It was observed that most trips were made during the month of September. Visual reports of the number of trips that were made on every day of the week were also obtained.

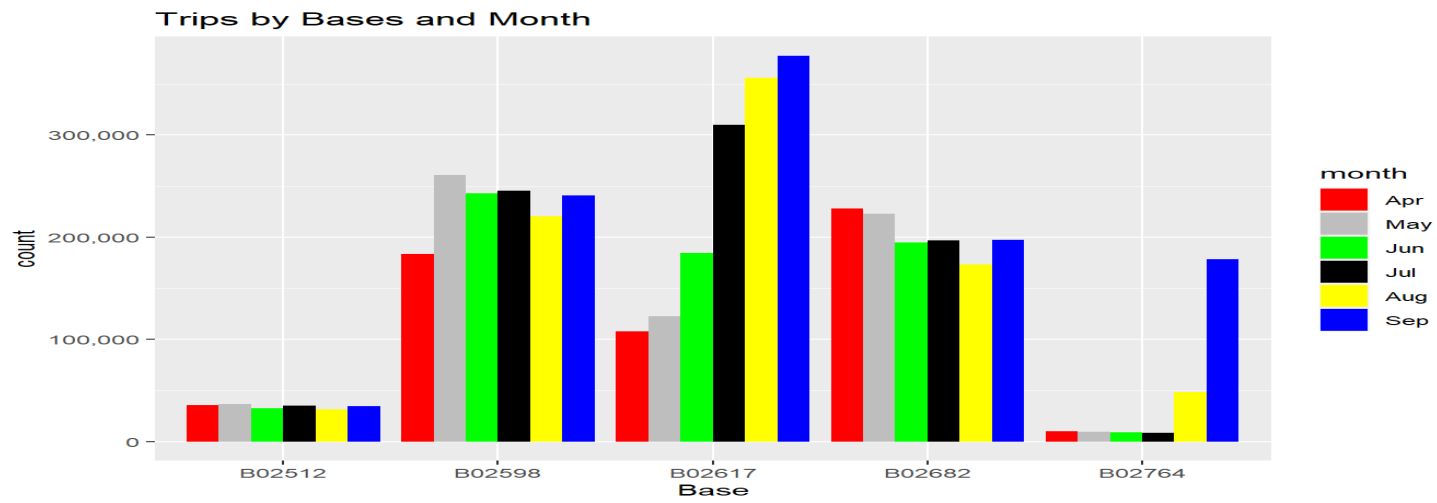


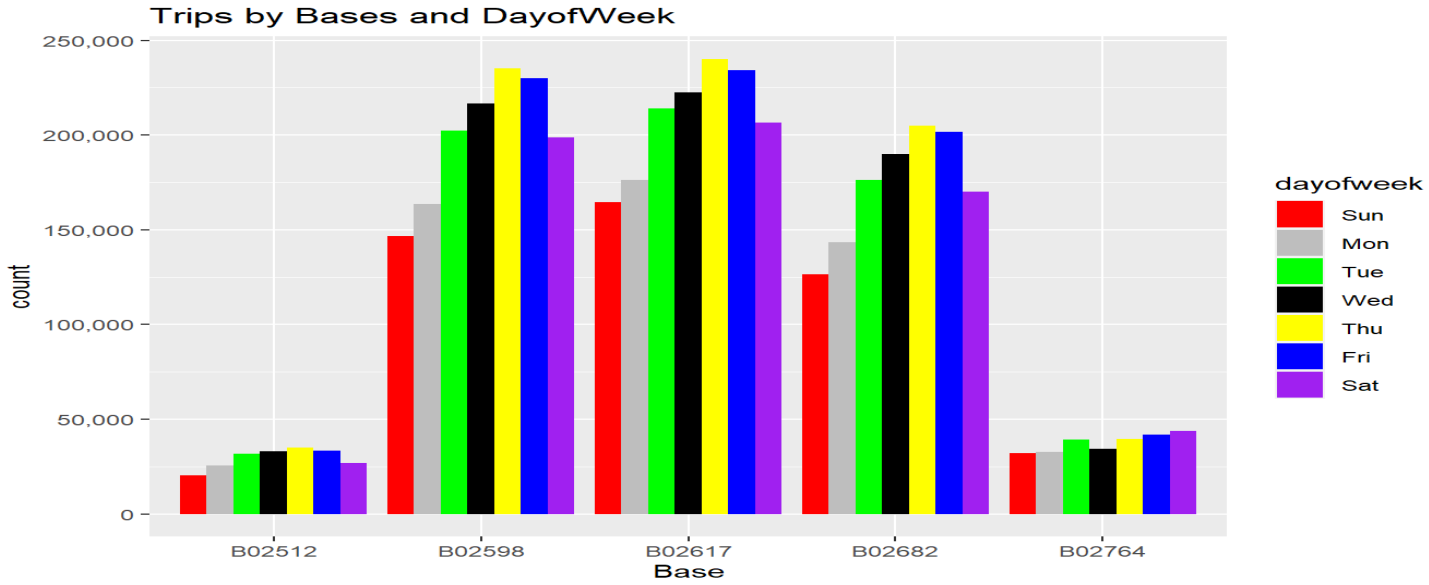




The number of trips that were taken by the passengers from each of the bases were also plotted. It was observed that B02617 had the highest number of trips. Furthermore, this base had the highest number of trips in the month. Thursday observed highest trips in the three bases – B02598, B02617, B02682.







Finally, the rides in New York city were also visualized by creating a geo-plot that visualized the rides during the 2014 (April – September) and by the bases in the same period.