UBER DATA ANALYSIS- REPORT

Completed by: Angad Singh Pental

Data storytelling is the methodology for communicating information tailored to a specific audience with a compelling narrative. It is an essential component of Machine Learning through which we can understand the background of various operations. The purpose behind such a project is to gain insights from complex data that will guide us towards making better decisions.

I have used to Uber TLC FOIL Response dataset, which contains over 4.5 Million records of Uber Pickups in New York City.

I have implemented data visualization concepts using R. I have combined the datasets from April-September 2014 and prepared plots to measure the trips made by Uber drivers.

The packages from R I have used include:

1. **ggplot2:** It is the most popular data visualization library and is widely used to create aesthetic plots.
2. **dplyr:** We use it in data manipulation.
3. **tidyr:** It helps tidy the data.
4. **DT:** It helps us prepare datatables.
5. **Scales:** It helps in automatically mapping the data to the correct scales with well-placed axes and legends.

After importing the required packages and data, we perform some formatting for the Date-Time variable. We can create time objects such as day, month, year, hour, minute, and second through this step.

The following are the findings uncovered from data visualization using the Uber Pickups Data:

1. The number of trips is higher in the evening, around 5:00 and 6:00 PM.
2. The 30th day records the highest number of trips, indicating that business tends to pick up by the end of a month.

3. Thursdays tend to record a higher number of trips, while September shows the highest number of pickups.
4. Base B02617 records the highest number of pickups, especially during September, with a higher number of trips on Thursday.
5. From the Day and Month Heatmap, we see that most of the months have slow starts, with the number of trips picking up by the end. There also seem to be fluctuations in the number of trips throughout the month in April and May.
6. The latter months of August and September show a high number of trips on Fridays and Saturdays.
7. We also see that the best performing base, B02617, shows high numbers during July, August, and September, on almost all days of the week, especially from Tuesday to Friday.

Apart from the different graphs, I created a geo-plot using ggplot2 to visualize the Uber Pickups during the specified period. I have also displayed these pickups in terms of the different Bases.