**Uber Data Visualization**

New York City (January-June 2014)

**Introduction:**

In this project,I am trying to visualize the Uber rides in New York City during January-June 2014, and the objective of the project is to understand the uber rides at different locations, times in different months which will help uber to know the people demand by location, time and eventually help them in taking any decision to improve the business. The statistics are very much important for a business to grow.

Individual datasets have Multiple of 100,000 observations in each Month but when joined with Base, it all came down to 1416 rows and 10 columns, where information about uber rides of each month in separate excel sheets, here I have performed union and combined all of them with Base of vehicle.

Some of the questions we are going to answer from the project are to the number of rides based on Base models, number of rides by weekday, map by geolocation, number of rides by hours, pickups by weekday and time slot and some interactive Dashboards.

The data has the following attributes:

1. Date/Time
2. Lat
3. Lon
4. Base
5. Active Vehicles
6. Trips
7. Dom
8. Weekday
9. Hour
10. Month
11. Weekday
12. Time slot
13. Pick up

**Data Description:**

1. **Date/Time**: The date and time of the Uber pickup
2. **Lat**: The latitude of the Uber pickup
3. **Lon**: The longitude of the Uber pickup
4. **Base**: The TLC base company code affiliated with the Uber pickup
5. **Active Vehicles**: Number of Active Vehicles
6. **Trips**: Number of Trips
7. **Dom**: Day of the month
8. **Weekday**: Name of the day in a week
9. **Hour**: Hour of the day
10. **Month**: Name of the month
11. **Weekday**: The day of the week
12. **Time slot**: The part of the day
13. **Pick up**: The ride Pick up

The Base codes are for the following Uber bases:

B02512: Unter  
B02598: Hinter  
B02617: Weiter  
B02682: Schmecken  
B02764: Danach-NY

**Methodology**: The data for this project was taken from Kaggle which is publicly available to download. This is the link for the data source(<https://www.kaggle.com/hugomenz/uber-data-visualization/data>).

During the project I have done some Data Preprocessing to consume less memory and create some new columns like Hour, Day etc. from Date/Time column. And then I created separate sheets for each month in a single excel workbook. After loading data into Tableau, I joined them all by Base model. Python helped me in performing the Analysis and Preprocessing.

**Data Preparation:**

Here I used python to load and do some preprocessing steps on datasets. The dataset is very clean without any missing values, and I changed each column name.

Look at python snippet below.

Graphical user interface

Description automatically generated with low confidence

Here I changed datatypes of few columns and created some new columns (dom, week\_day, hour, month) from existing Date/Time column.

Text

Description automatically generated

Below, I dropped all rows from dataset which have at least one ‘nan’ value. After that I have created a single excel workbook with individual sheets for months.

Application

Description automatically generated with low confidence

**Analysis:**

**Q1: Number of Trips by Base Model**

**Chart, bar chart

Description automatically generated**

In the above plot, I am just calculating number of trips by Base and Base (B02764, Green) has more trips (around 1900,000) compared to other Base Models, we can also see other Base models trips in different colors.

**Q2: Number of Trips by Weekday**

**Chart

Description automatically generated**

In the above plot, I am calculating number of Trips on each day of a week, and we can see representation of each month in different colors which can tell number of trips on each day and number of trips for month also. Clearly, we have a greater number of trips on Saturday and that makes sense because its weekend. And on the right side, we can select weekday manually also.

**Q3: Number of Trips by Week (Animation)**

**Graphical user interface, text, application, email

Description automatically generated**

The above plot is an Animation which shows the number of trips by week. If we click a little play button on the right, the plot keeps moving to the right showing number of trips by week.

**Q4: Active Vehicles by Base**

**Chart, pie chart

Description automatically generated**

The above plot is a pie chart that shows the number of Active Vehicles by Base model, all the Base models have approximately same number of Active Vehicles.

**Q5: Drill Down by Base and month**

**Graphical user interface

Description automatically generated with medium confidence**

The above plot is a Drill Down approach which shows number of Trips by Base but when I click on any Base model it’ll open by all months and shows trips for that month with that Base model.

**Dashboard 1:**

**Chart, bar chart

Description automatically generated**

This is the Dashboard having all the above plots at one place. And this is an Interactive Dashboard where can visualize by each weekday, each Base when selected that color in the plot.

**Q5: Map by Base**

**A picture containing text, screenshot, electronics, display

Description automatically generated**

This is a map showing all the locations of trips booked in New York city, by base in different colors and we can select individual base on the right side to see the locations of trips booked of that Base model.

For better visual, I used Dark theme for the map and selected Streets, States and Province names in Map Layers.

**Q6: Map (Longitude, Latitude) by Hour**

**Graphical user interface, map

Description automatically generated**

The above map shows the locations of trips booked in New York city and we can select an Hour on the right side to view number of trips booked by specific Hour of the day.

**Q7: Number of Records by Hours**

**A picture containing text

Description automatically generated**

The above plot shows the number of records by Hour and each color indicates different month.

**Dashboard 2:**

**A picture containing map

Description automatically generated**

The above visualization is Dashboard that has Trips by Hours, Trips by Base Models and Month Wise.

**Q8: No of Trips by Base (Pie)**

**Chart, pie chart

Description automatically generated**

This is simple Visualization which tells Number pf Trips by Base. Base model (B02764) has more Trips.

**Q9: Base Model by each hour of the day**

**Chart, line chart

Description automatically generated**

The above plot shows the Base Model by Hour of the day in different colors.

**Q10: No of rides by each pickup**

**Chart, bar chart

Description automatically generated**

We can see number of rides from 2 different (Airport, City) pick up points. Clearly, there are more pick-ups in the city compared to Airport.

**Q11: Rides distribution across time**

**Chart

Description automatically generated**

This is the Area plot which tells the Number of records on different weekdays by hour of the day. And there is high demand in the mornings and evenings which is evident because people use commute during these working hours.

**Q12: Day vs Time slot of the day**

**Chart, bar chart

Description automatically generated**

The above plot shows the number of records by time slot of each weekday. This plot also conveys the same, high demand for rides in the morning and evening.

**Q13: Pickup vs Timeslot**

**Chart, treemap chart

Description automatically generated**

The above plot shows the percentage of rides at different time slots in each pickup. The pickups at Airport (Evening) and City (Morning) are high compared to other.

**Q14: Pickup vs Day**

**Chart, bar chart

Description automatically generated**

The above visualization shows the rides at each pickup spot on all weekdays. Friday, which is weekend has a greater number of rides.

**Q15: Pick up vs Weekday vs Timeslot**

**Chart

Description automatically generated**

This is a Drill Down visualization which shows the number of rides at each pickup spots on each weekday and time slot of the day.

**Dashboard 3:**

**Chart

Description automatically generated**

This is an Interactive Dashboard where we can just select any color for weekday or pickup spot in the 1st plot to see highlighted area of the weekday in 2nd plot.

**Conclusion:**

After Analyzing the dataset on various factors, I was able to come up with the following insights

1. The Number of Rides on weekends are more compared to other weekdays.
2. The pickups in the city are in greater number compared to Airport pickups.
3. During any weekday, the rides have high demand in the time slots **morning** and **evening** because people tend to commute to their work in those slots.
4. In the year 2014, the number of rides kept increasing from month to month. This can be easily visualized through Animation.
5. And if we want to know the number of rides by location, Map is plotted between Latitude and Longitude.
6. During a day (0- 24 hours), the number of rides is higher from 4 to 10 and then from 16 to 22 hours.
7. And there are 3 Interactive Dashboards which can be used to visualize all visualizations all at one place.

**Future Works**:

1. After the analysis, I got an idea that we can also add weather data for that specific month and then we can visualize number of rides on sunny and rainy days. We will also have weather related information like Visibility, Wind and Sea level fields which we can use in our visualizations.
2. We can also use this analysis and visualizations to build some predictive models to predict the cost per ride as demand increases at different places and in different time slots.