Uber Case Study

Background

Uber Technologies, Inc. is an American multinational transportation network company based in San Francisco

Ridesharing is a very volatile market and demand fluctuates wildly with time, place, weather, local events, etc.

It has operations in over 785 metropolitan areas worldwide with over 110 million users worldwide.

The key to being successful in this business is to be able to detect patterns in these fluctuations and cater to the demand at any given time.

Objective

To extract actionable insights from the data that we have collected over the past 6 months to optimise resources and identify area of growth and improvement.

We will be majorly focusing on these problems -

- Variables that influence the pickups
- Factors that affect pickups the most and the respective reasons
- Ways to capitalize the fluctuating demand

Data Information

The data contains weather information, location and no. of pickups

Variable	Description
pickup_dt	Date and time of the pick up
borough	NYC's borough
pickups	Number of pickups for the period
spd	Wind speed in miles/hour
vsb	Visibility in miles to nearest tenth
temp	Temperature in Fahrenheit
dewp	Dew point in Fahrenheit
slp	Sea level pressure
pcp01	1-hour liquid precipitation
pcp06	6-hour liquid precipitation
pcp24	24-hour liquid precipitation
sd	Snow depth in inches
hday	Being a holiday (Y) or not (N)

Observations	Variables	Duration
29101	13	6 months

Note:

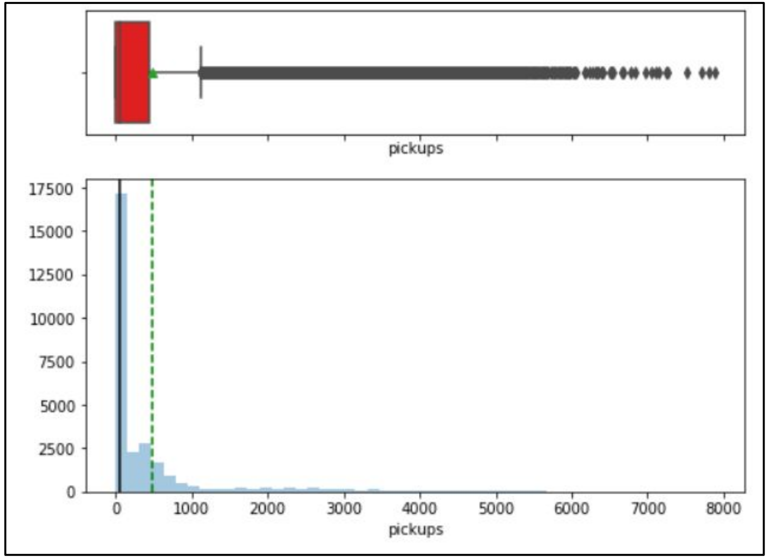
- There are some missing values in the borough column that has been referred to ‘unknown’ in the analysis
- The date column has been splitted into year, month and day separately for the analysis

Exploratory Data Analysis - Number of Pickups

In this data, there are variables like temperature, wind speed, dew point, number of pickups etc. that affects the ride sharing business.

Let us first explore some of the variables and how they are distributed

Number of pickups

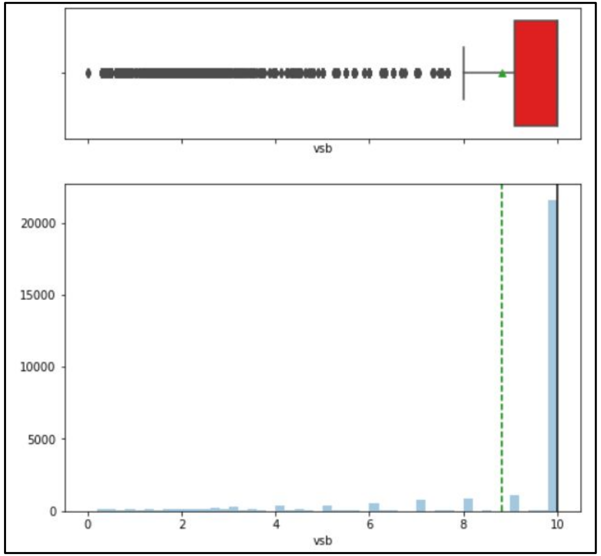


Observations:

- The distribution of hourly pickups is highly right skewed
- Majority of the hourly pickups are close to 0
- Median pickups = 54 but the mean is ~500
- There are a lot of outliers in this variable.
- While most hourly pickups are at lower end, we have observations where hourly pickups went as high as 8000

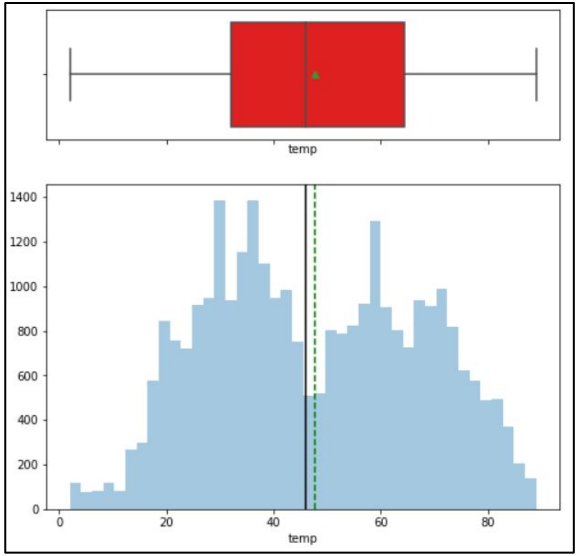
Exploratory Data Analysis - Visibility, Temperature & Dew point

Visibility



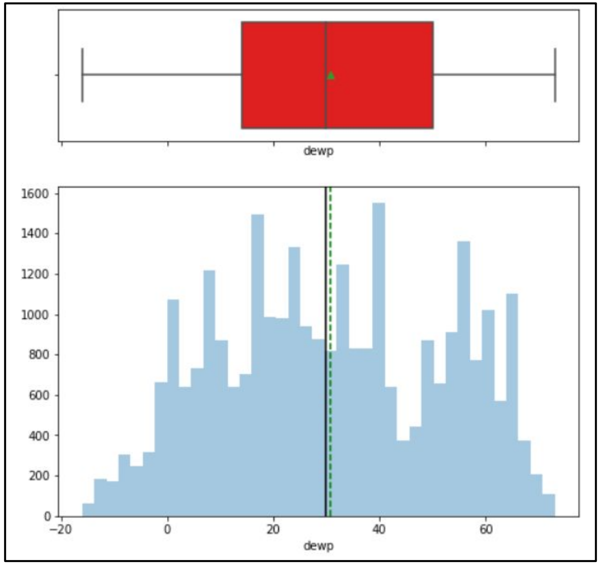
- Both the mean and median are high indicating that the visibility is good on most days
- There are however outliers towards the left, indicating that visibility is extremely low on some days.

Temperature



- Two peaks (Bi-modal) can be seen for temp, one at around 35F other at around 60F. The hump is greater at 35F (~1.5 C) indicating cold weather conditions.

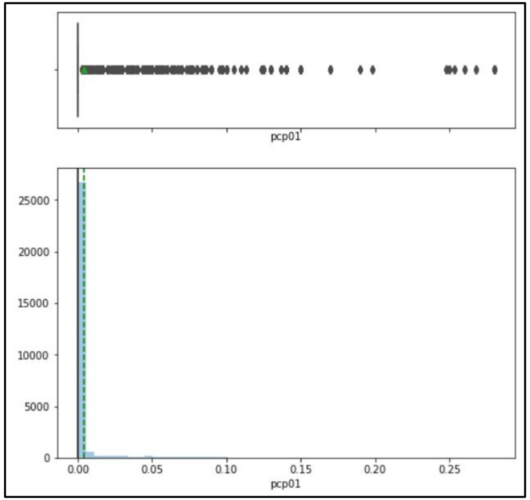
Dew point



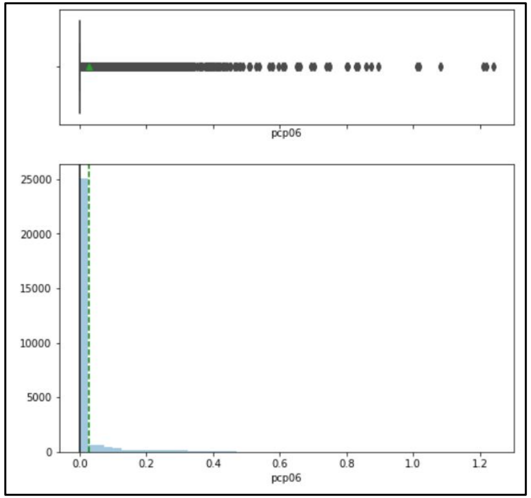
- The distribution is similar to that of temperature. It suggests possible correlation between the two variables.
- Dew point is an indication of humidity, which is correlated with temperature.

Exploratory Data Analysis - Univariate

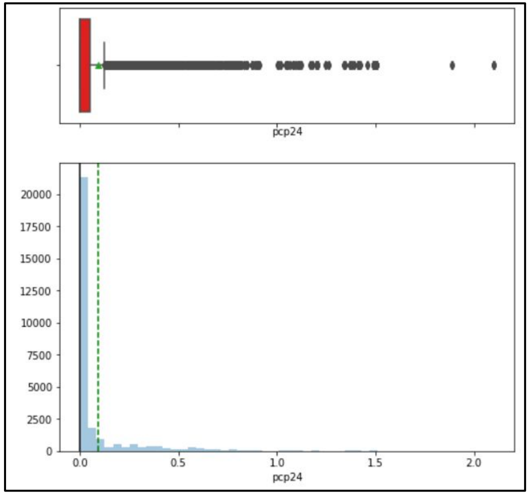
1 hour liquid precipitation



6 hour liquid precipitation



24 hour liquid precipitation



Observations:

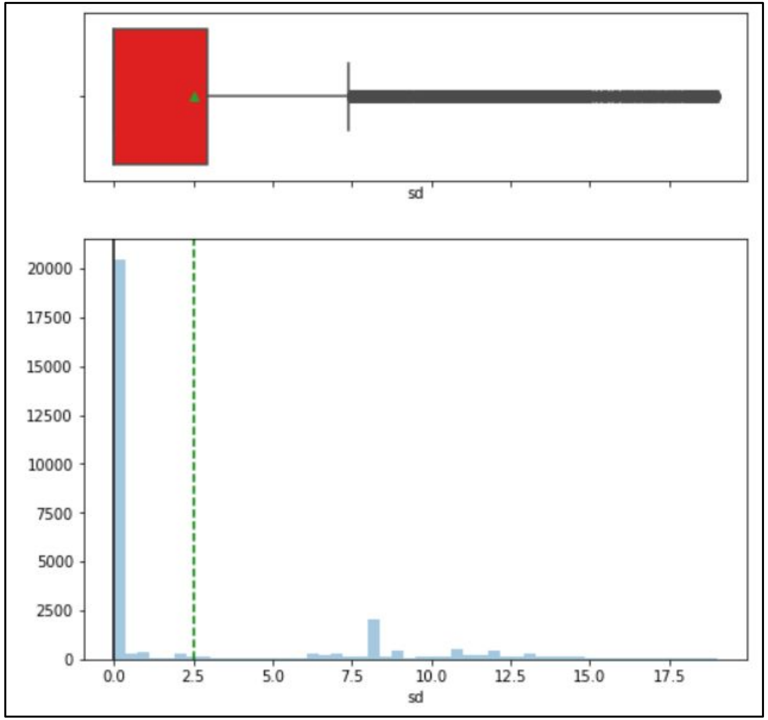
- It rains on relatively fewer days in New York.
- Most of the days are dry.
- When it rains, and sometimes when it rains heavily, we get outliers.

Exploratory Data Analysis - Snow depth

Observations:

- We observe that there is a snowfall in the time period that we are analysing.
- There are outliers in this data.
- We will have to see how snowfall affects pickups. We know that very few people are likely to get out if it is snowing heavily, so our pickups will decrease when it snows.

Snow depth

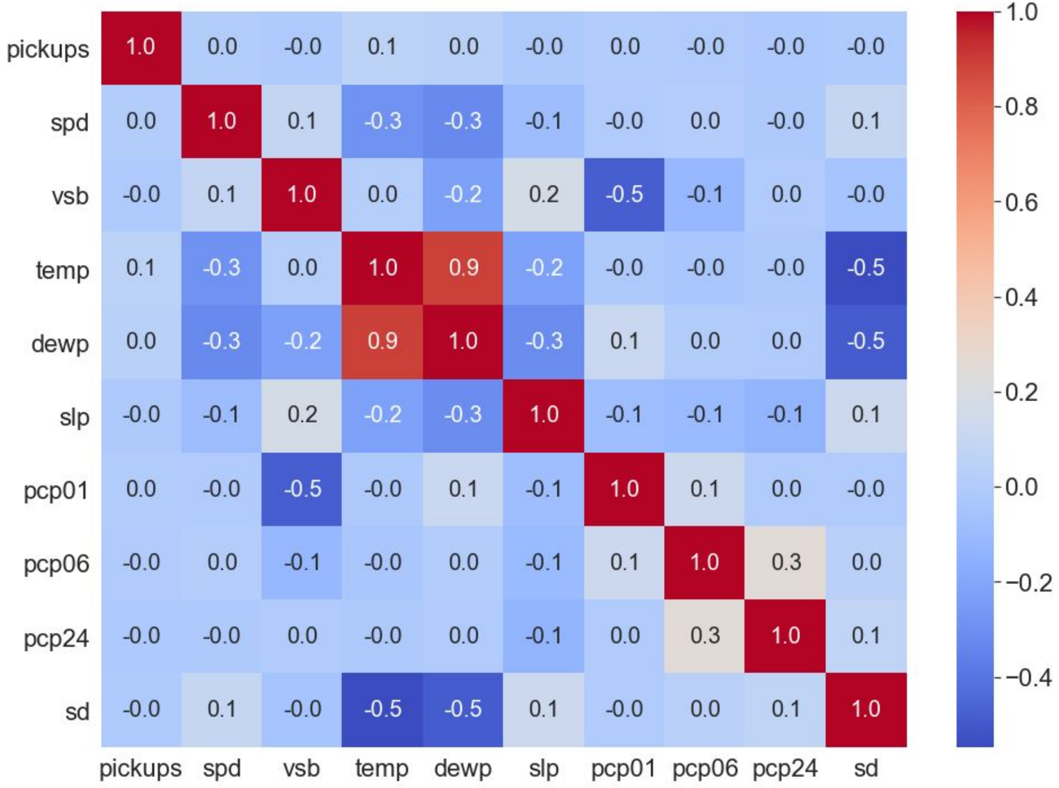


Exploratory Data Analysis - Correlation matrix

Observations:

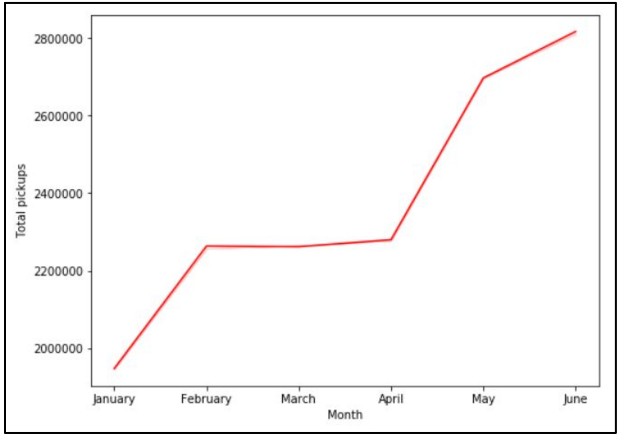
- Temperature shows a high correlation with dew point
- Visibility is negatively correlated with precipitation. If the rains are high during the hour, visibility is low.
- Snow depth is also negatively correlated with temperature.
- Wind speed and sea level pressure is negatively correlated with temperature. As the temperature increases, wind speeds decrease and so does sea level pressure.
- There does not seem to be a strong relationship between number of pickups and weather stats.

Correlation matrix

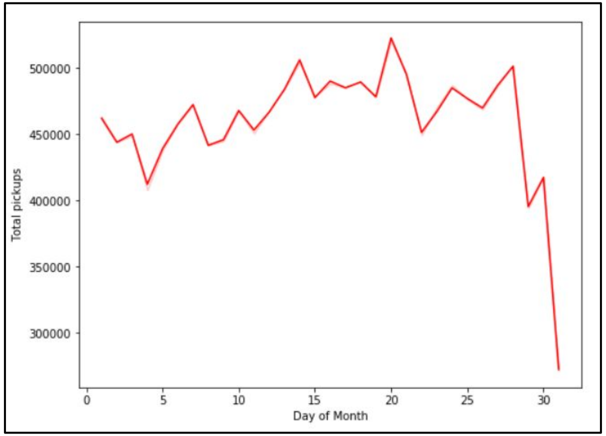


Exploratory Data Analysis - Total Pickups

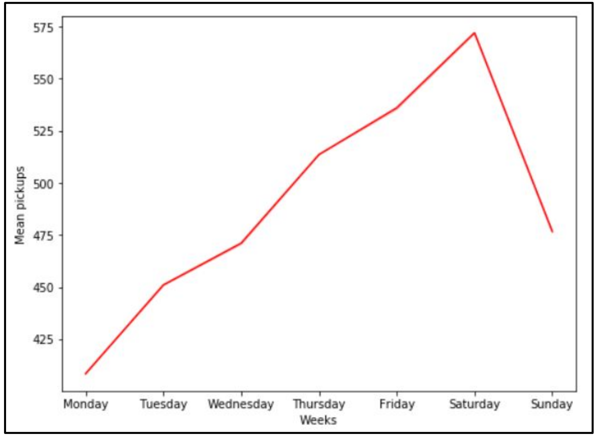
Total Pickups per month



Total pickups per day



Total pickups per weekday



Observations:

- There is clear increasing trend in monthly bookings
- Bookings in June are almost 1.5 times that of Jan

Observations:

- There is a steepfall in the bookings around the last days of the month
- There is a peak in the bookings around 20th day of month

Observations:

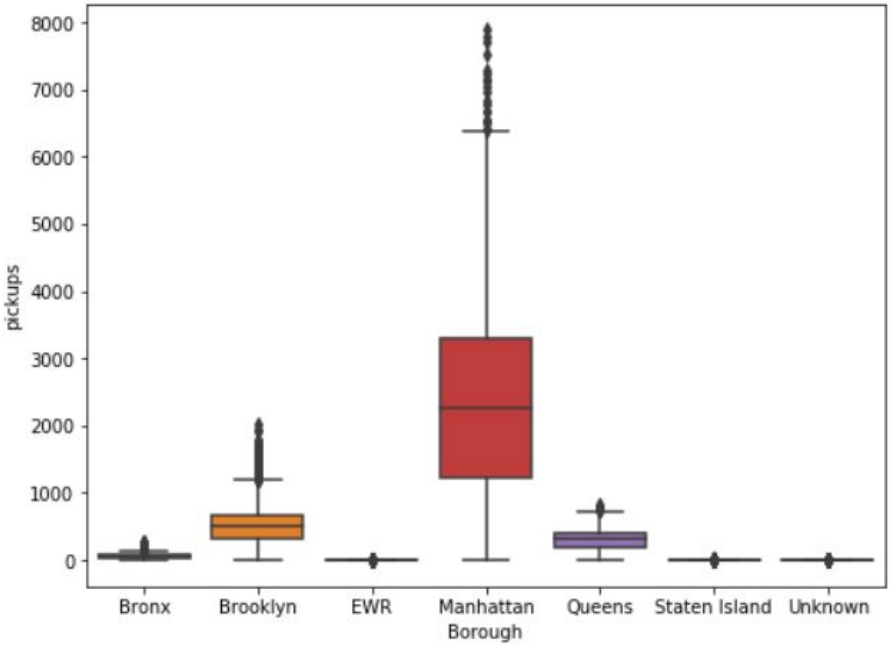
- Pickups gradually increase as the week progresses and starts dropping down after saturday.
- Demand is usually low at the beginning of the week

Exploratory Data Analysis - Pickup across borough

Observations:

- There is a clear difference in ridership across the different boroughs.
- Manhattan has the highest no. of bookings
- Brooklyn and Queens are distant followers
- EWR, Unknown and Staten Island have very low bookings. The demand is so small that probably it can be covered by the drop-offs of the inbound trips from other areas.

Pickups across borough

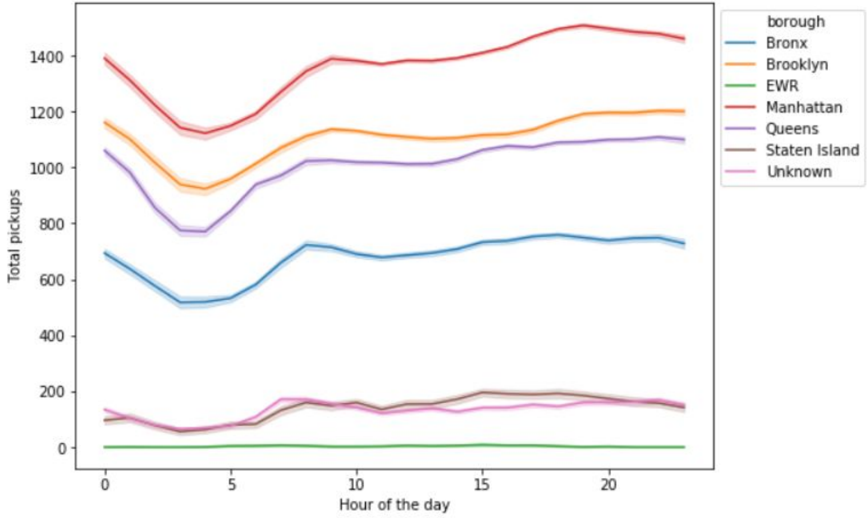


Exploratory Data Analysis - Pickups across hours

Observations:

- Hourly pattern can be seen in all the boroughs.
- Especially on the second plot where a logarithmic scale has been applied to Y axis, it is obvious that the 4 major boroughs follow the exact same pattern.
- The value of Staten Island and Unknown are more scattered
- EWR seems to have a random demand with the majority of the values being zero with a few 1s and 2s.
- Borough and hour of the day combined could be good predictors of pickups.
- Manhattan sees the most uber pickups.

Total pickups per hour across boroughs

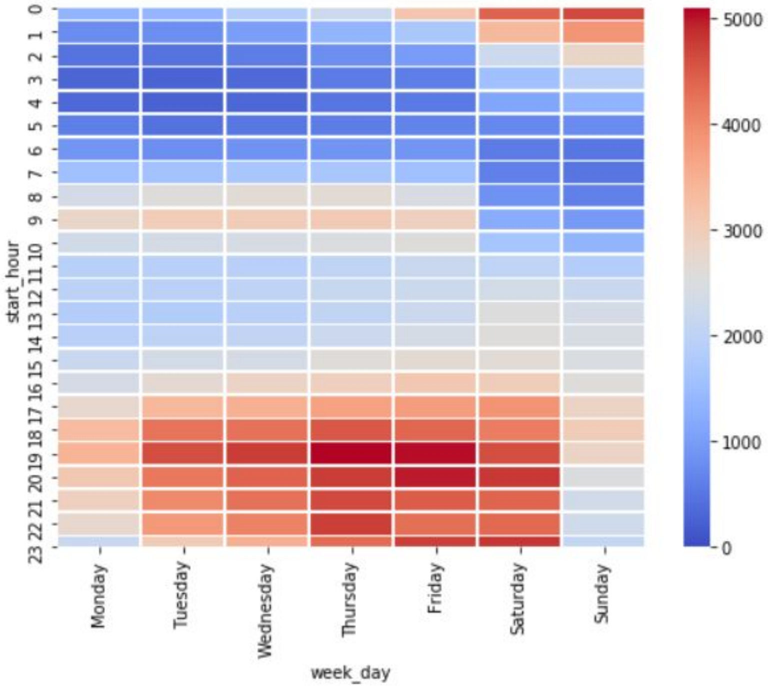


Exploratory Data Analysis - Manhattan Pickups Heatmap

Observations:

- The demand for Uber peaks during the late hours of the day when people are returning home from office.
- Demand continues to be high during the late hours of the day (midnight) on fridays and saturdays.
- It is odd that the demand of Uber is no as high on monday evenings as compared to other working days.

Manhattan pickups - weekday+hour



Conclusion

After all the analysis, we have been able to can conclude that

1. Uber cabs are most popular in the Manhattan area of New York.
2. Weather conditions do not have much impact on the number of Uber pickups.
3. The demand for Ubers has been increasing steadily over the months (Jan to June).
4. The rate of pickups is higher on the weekends as compared to weekdays.
5. It is encouraging to see that New Yorkers trust Uber taxi services when they step out to enjoy their evenings.
6. People use Uber for regular office commutes. The demand steadily increases from 6AM to 10AM in the morning, then declines a little and starts picking up at 12PM. The demand peaks at 7-8 PM at night.
7. We need to further investigate the low demand for Ubers on Mondays.

Recommendations

Based on the analysis, there are following recommendations that can help the business grow:

1. Manhattan is the most mature market for Uber. Brooklyn, Queens and Bronx show a lot of potential.
2. There has been a gradual increase in Uber rides over the last few months and we need to keep up the momentum.
3. Riderships are high at peak office commute hours on weekdays and during late evenings on Saturday. Cab availability must be ensured during these times.
4. The demand for cabs is highest during saturday nights. Cab availability must be ensured during this time of the week.
5. We need to procure data for fleet size availability to get a better understanding of demand-supply status and build a machine learning model to accurately predict pickups per hour, to optimise the cab fleet in respective areas.
6. We need to procure more data on price and build a model that can predict optimal pricing.

greatlearning

Power Ahead