



# Uber Supply-Demand Gap

---

Arockia Liborious

*PG Diploma in Data Science March 2019*

# Purpose



Problem Statement



Steps to Clean up Data



Insights from data



Final comments

# Problem Statement

- *Uber is facing - driver cancellation and non-availability of cars leading to loss of potential revenue.*
  - *Clean up the Uber data*
  - *Once done - Identify the Root cause of the problem- Specifically cancellation & non availability of cars*
  - *Suggest ways to tackle the problem*

# Steps to Clean and Optimize up the Raw Data file

- Standardize the Request made Timestamp and Drop Timestamp
- Retrieve the Weekday and Hour information from the datafile

```
#Making the data easier to read by deriving metrics and adding new columns
df1['Request Date'] = df1["Request timestamp"].dt.date
df1['Request Time'] = df1["Request timestamp"].dt.time
df1['Drop Date'] = df1["Drop timestamp"].dt.date
df1['Drop Time'] = df1["Drop timestamp"].dt.time
#adding Weekday and hour information on the data set
df1['Request Weekday'] = df1['Request timestamp'].apply(lambda x: dt.datetime.strftime(x, '%A'))
df1['Request Hour'] = df1['Request timestamp'].apply(lambda x: x.hour)
```

- Calculate the Demand of cars & average requests per driver

```
#Finding out the demand of cars based on the trip completion to requests made data
cars_on_demand = pd.DataFrame({'Total No. of Drivers': [df1['Driver id'].nunique()],
                              'Total Demand Per Day': [df1['Request id'].nunique()/5],
                              'Trips Completed Per Day': df1[(df1['Status']=='Trip Completed')].shape[0]/5})

cars_on_demand
```

# Insights from data

- Uber is losing quite a lot of business due to shortage of available Cabs.

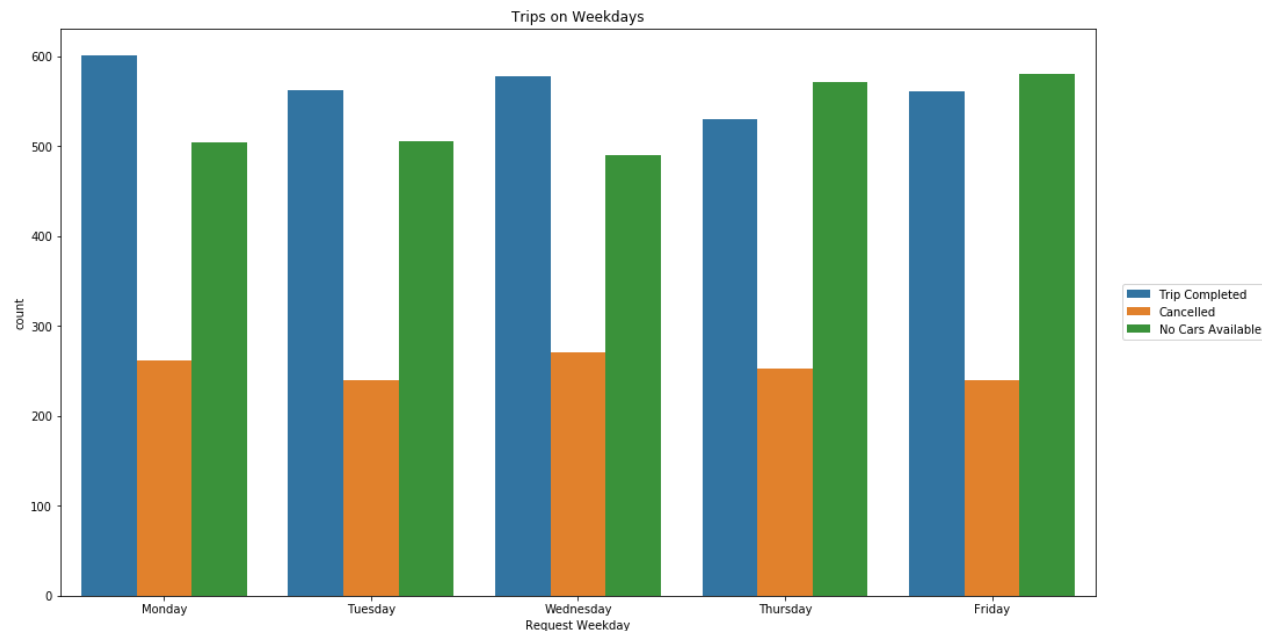
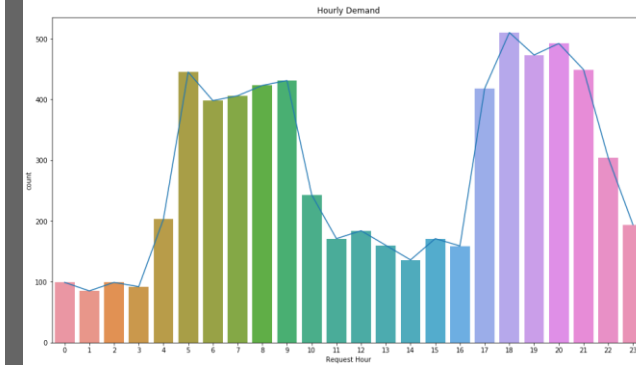
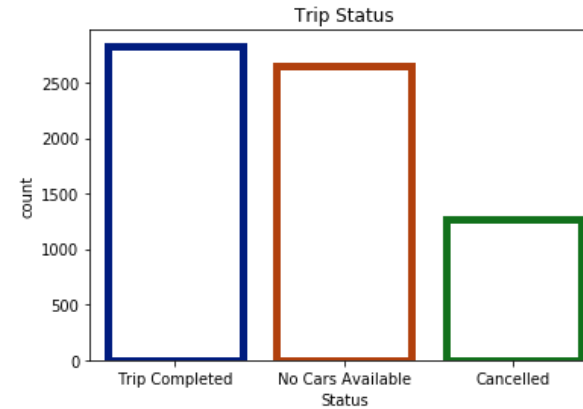
*(Trip Status)*

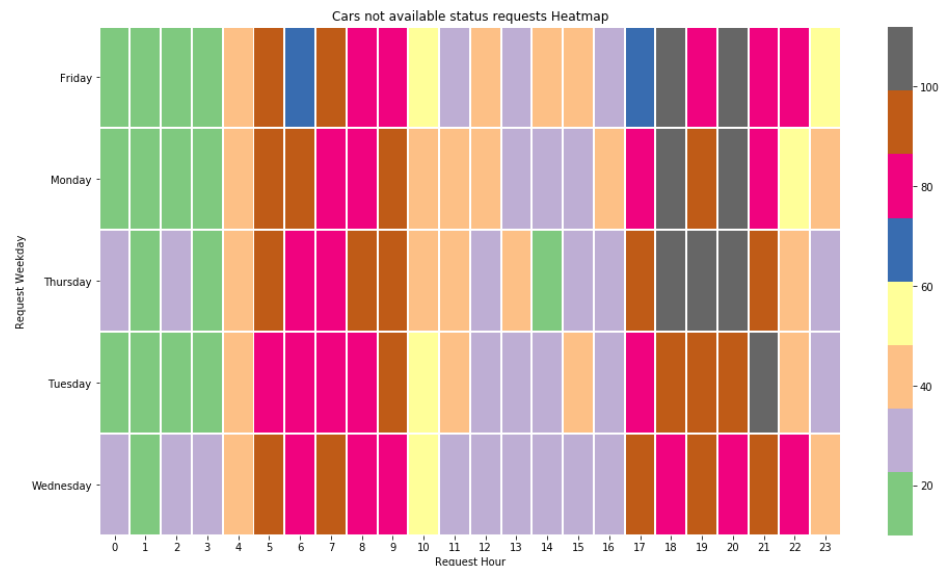
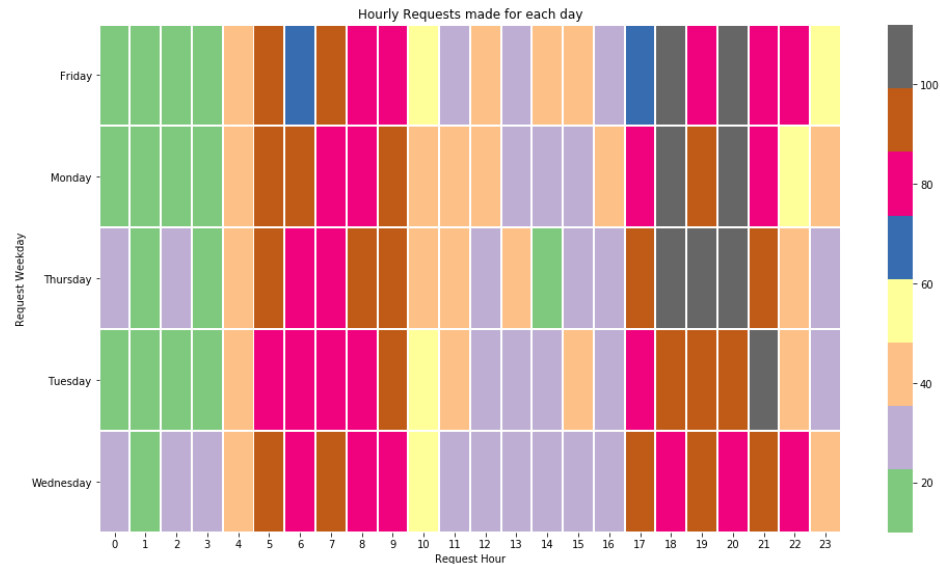
- On certain days of the week, the number of no cars is more than the completed trips. This indicates that the higher number of requests are left unattended.

*(Trips on weekdays)*

- Higher number of requests made between 5AM to 9AM and between 5PM to 9PM.

*(Hourly Demand)*



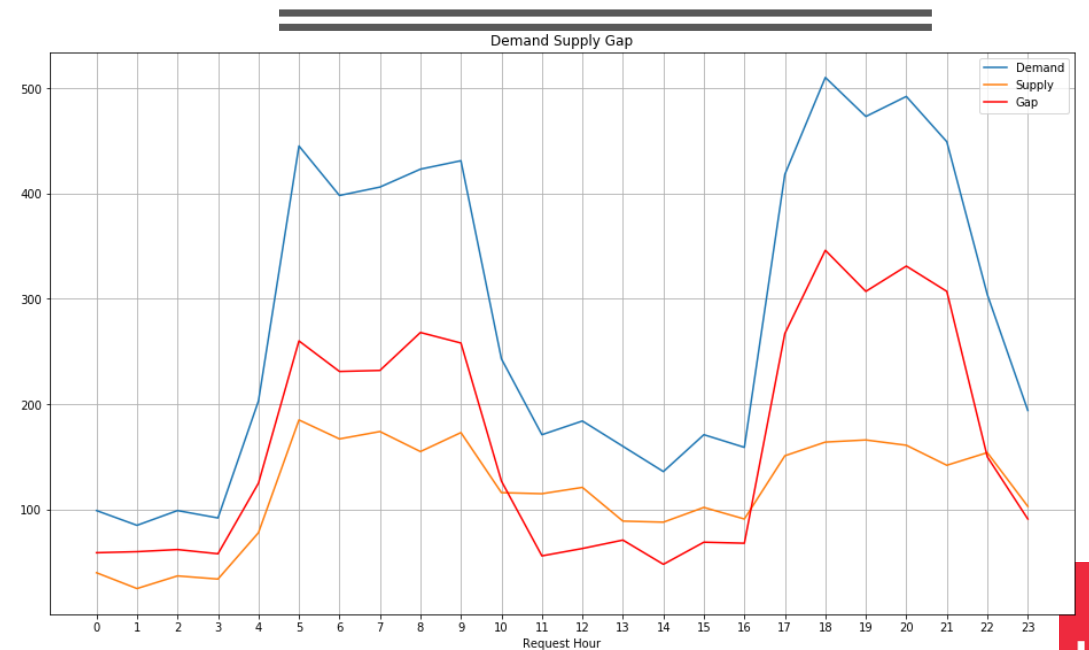
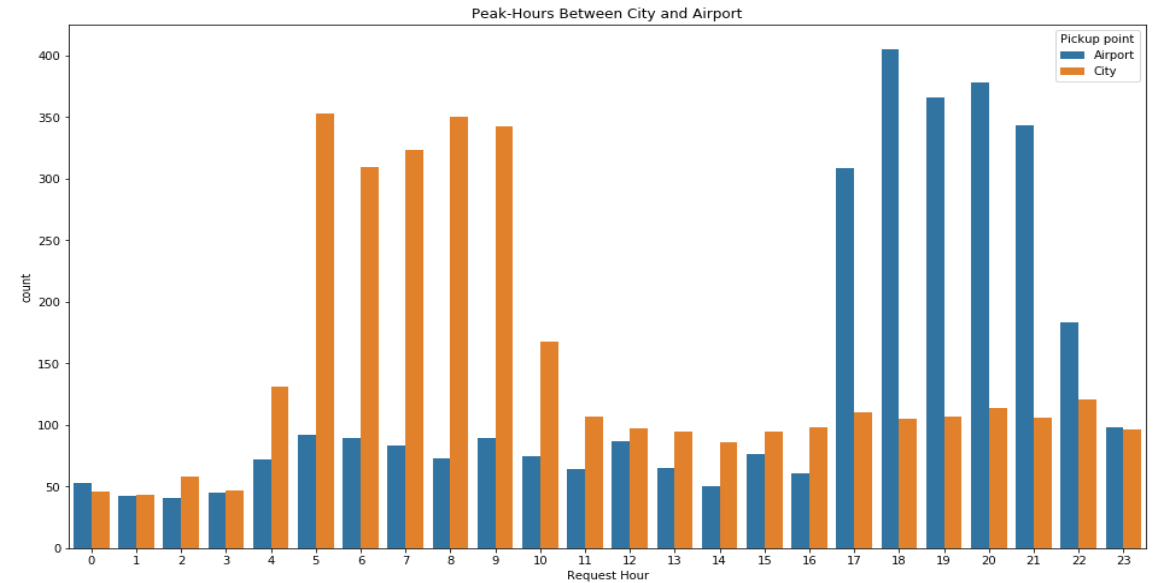


# Insights from data contd...

- Uber is unable to meet the demand at the peak request hours (5 am to 9 am and 5pm to 9 pm) . The heat map plotted between Request made hourly and No Cars available hourly data indicates the same.

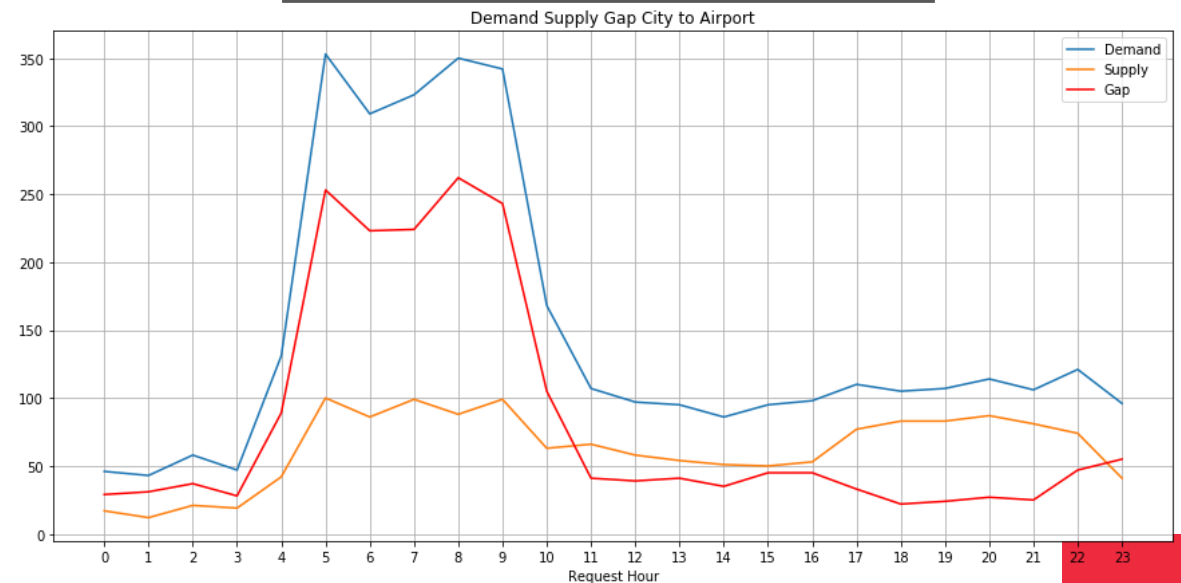
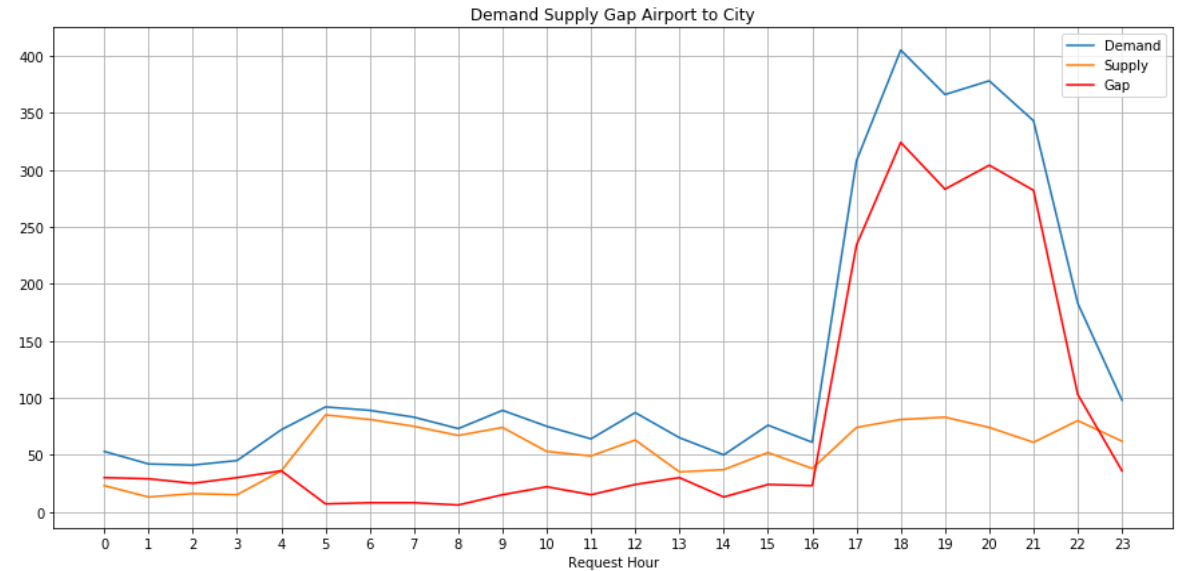
# Insights from data contd...

Pin point the supply and demand gap in City and Airport between the peak requests hours.



# Insights from data contd...

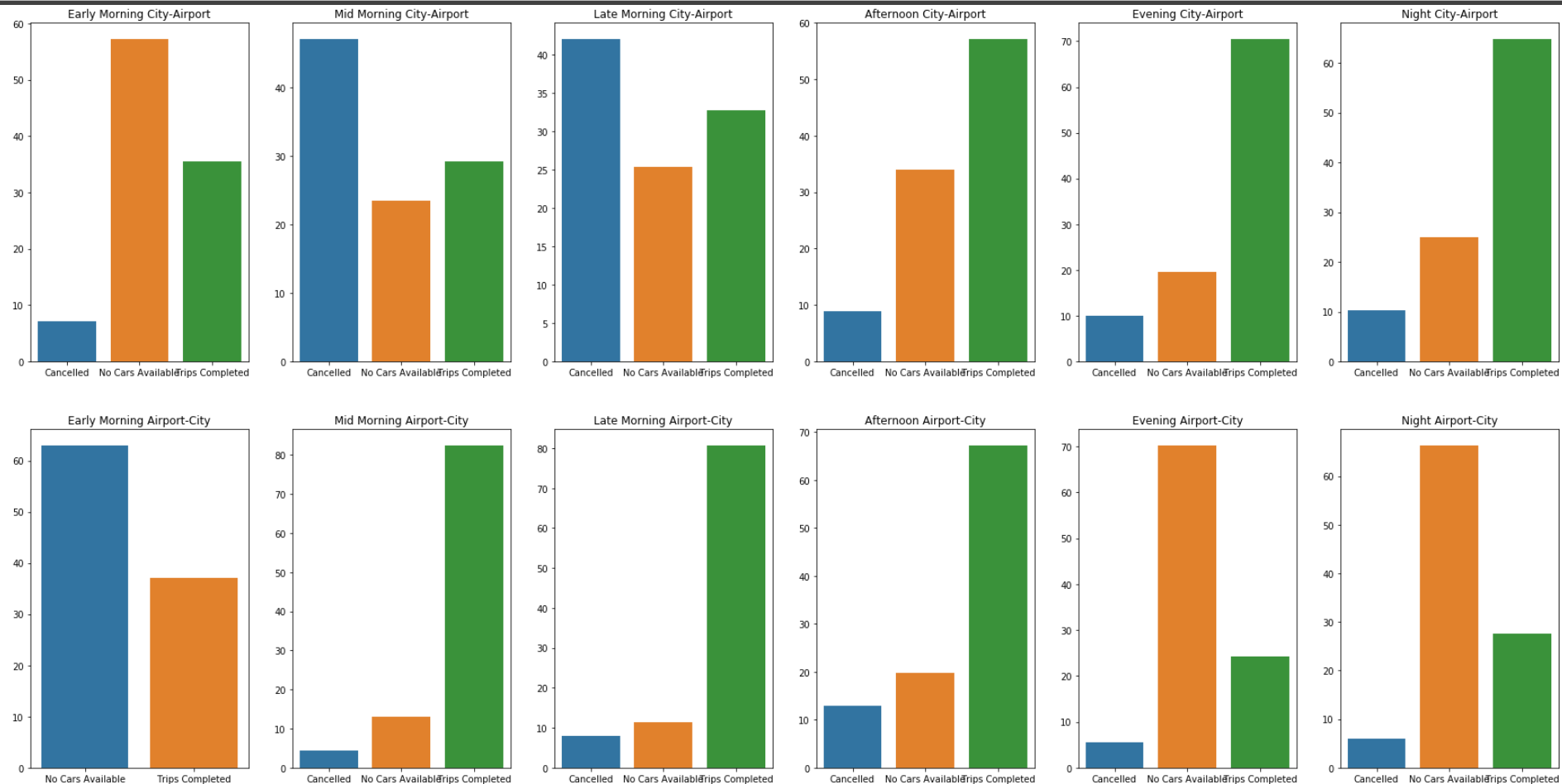
- Demand in the City is at its peak between 5AM and 9AM
- Demand in Airport is at its peak 5PM and 9PM at the Airport.





# Insights from data contd...

- Distribution of Trips Completed, No Cars and Cancelled requests between different time frames in the day to check how the data looks to and from the airport



# Suggestions to tackle the problem

- Create incentive for the cab drivers to work at high demand hours by
  - Having higher rates for rides from City to airport during Early morning hours and from Airport to City during evening and night hours.
  - Also provide food (dinner, breakfast, snacks allowances)
- To handle the demand during the evening and night at the airport, start directing cabs from city towards the airport starting late afternoon. Direct cabs towards the city from the airport to supply high demand at early morning hours.
- Survey to identify the reason behind higher percentage of cancellations during mid morning and late morning hours in the city. Attract users through lower fares than normal time of the day.