Uber Supply Demand Gap Assignment

UpGrad

Data Exploration

- The Uber case study related data is present in excel file of .csv file format named Uber Request Data.
- The data contains total 6 columns and 6745 rows i.e entries.
- There are six attributes/columns associated with each request made by a customer. They are:

Request id: A unique identifier of the request

Time of request: The date and time at which the customer made the trip request

Drop-off time: The drop-off date and time, in case the trip was completed

Pick-up point: The point from which the request was made

Driver id: The unique identification number of the driver

Status of the request: The final status of the trip, that can be either completed, cancelled by the driver or no cars available

Data Cleaning and Manipulation

- **Data Cleaning is removing null values and duplicated rows and columns.**

1. In the given data we don't have any duplicate rows.
2. 58.03% and 39.29% null values are present in drop timestamp and driver id column respectively.
3. Drop timestamp column should be removed as it is containing high percent of null values but it will also remove the values of cancelled and no cars available from Status column. Therefore these null values are not removed as it will cause the loss of data which is needed for analysis.

- **Data Manipulation enables faster data computation.**

1. Request date, day ,hour and month values are derived into separate columns from request timestamp which are useful for further data analysis.
2. The request hour derived from request timestamp is further categorized into 5 different time slots to find the supply demand gap within each time slot.

Analysing Trends For Each Day

- The data exists only of the year 2016.There are 5 unique request dates in the year 2016, they are 7th day of two months December and November , 13th of July , 14th of July and 15th of July .
- No cars available problem is highest on 2016-07-14 in evening slot when compared to other dates and cancellation of cars is almost equal in proportion in all dates except 2016-12-07.

UpGrad

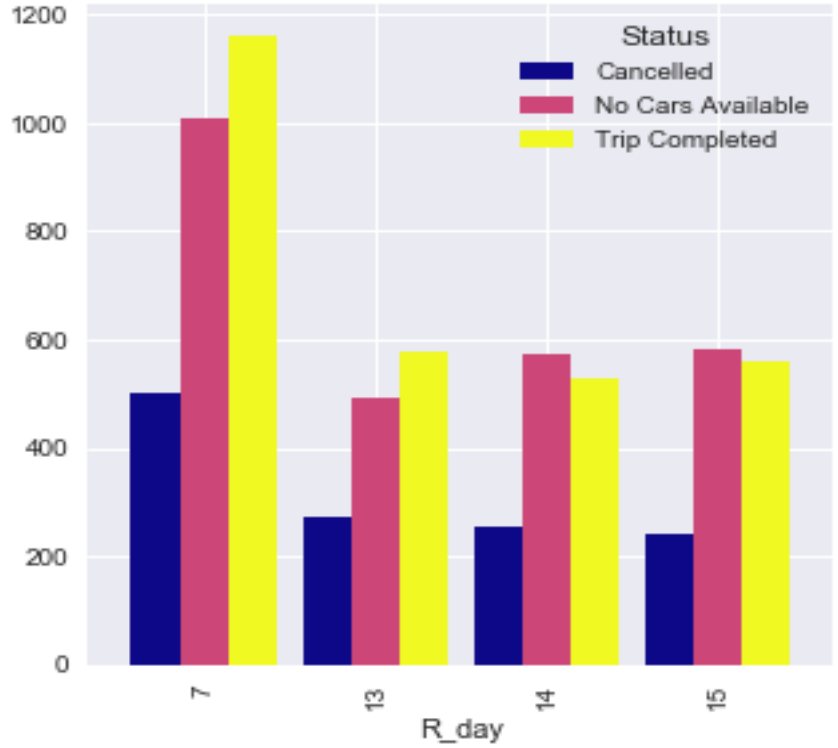
Analysing Trends For Each Day

REFER THE PREVIOUS SLIDE PLOTTING

- On 13th July evening is at a frequency of 248 with no cars availability problem and next comes the morning slot with 175 cancellation problem.
- On 14th July date it is having even more higher evening slot with 315 as no cars availability problem and morning slot with 177 cancellation problem .
- On 7th November , evening slot is having 276 requests id's which occurred with the problem of no cars available

REFER THE PLOTTING ON RIGHT SIDE

- On analyzing all the days we get to conclusion that 7th day on any month is having the highest request id's facing the both the problems and also trip is completed majorly on the 7th day when compared to other dates.

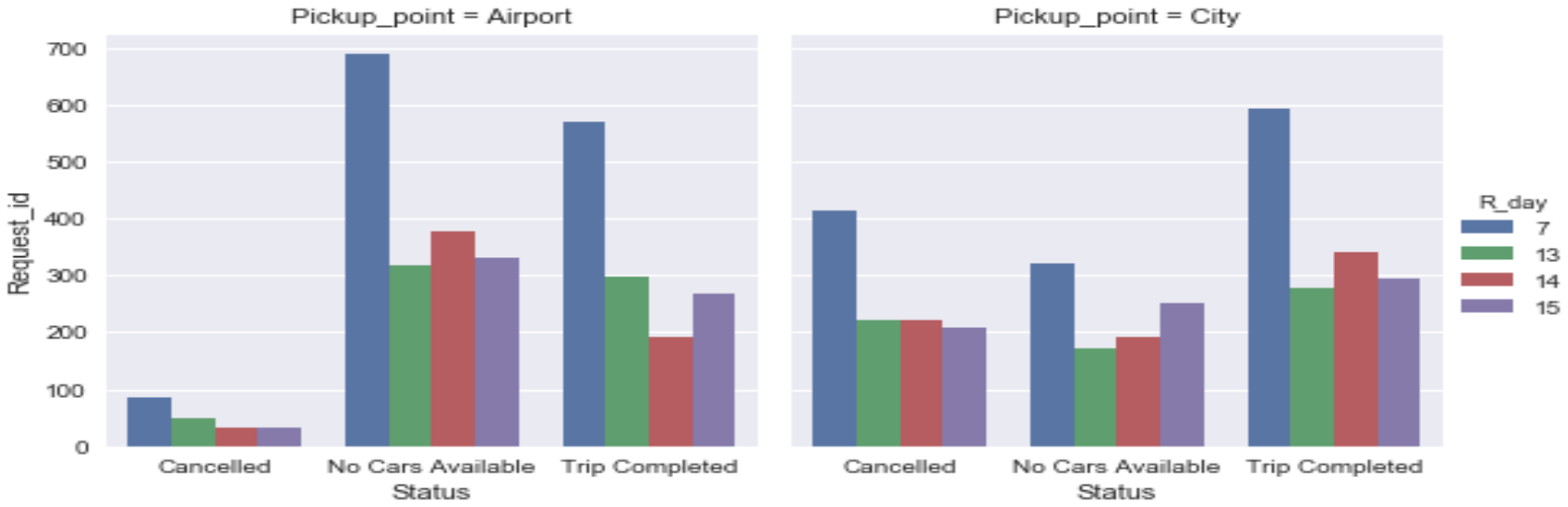


UpGrad

Combining Data For All Days

- The request from airport to city is facing higher number of no cars available problem maximum on 7th day and minimum on 13th day.
- The request from city to airport is facing higher number of cancellations problem max on 7th day and minimum on 15th day

This analysis is irrespective of month.

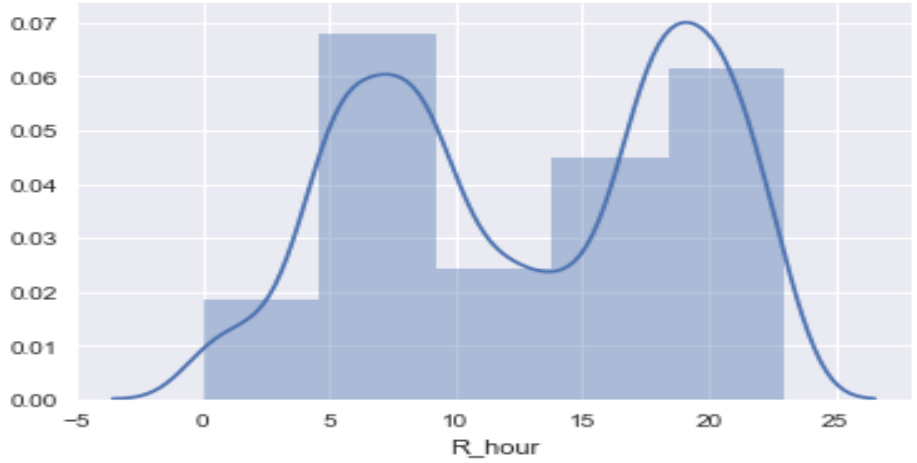


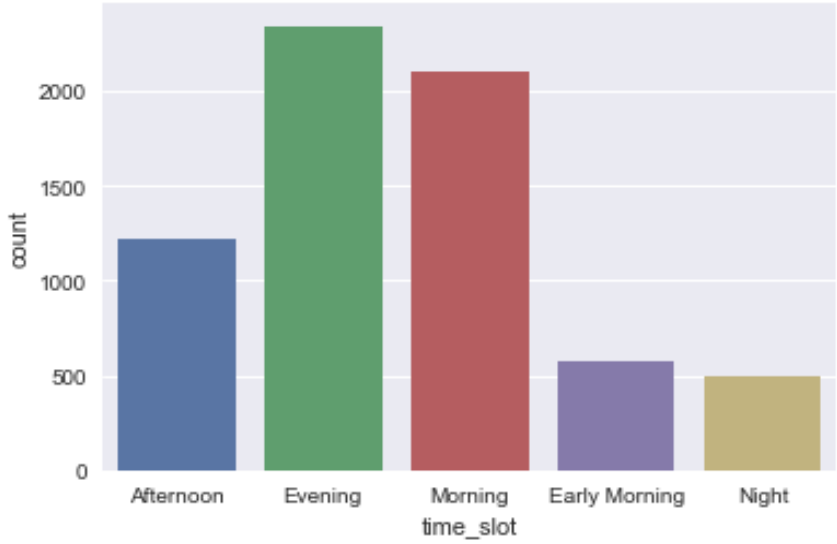
Binning Time Into 5 Categories

The kernel density estimate may be less familiar, but it can be a useful tool for plotting the shape of a distribution. Like the histogram, the KDE plots encodes the density of observations on one axis with height along the other axis.

- Kde is the thin curve drawn on the distplot showing that the morning and evening time slots have higher frequency .
- The same is shown in the bar plot after diving the hours into time slots.
- The hours 0 to 23 is divided into 5 timeslots accordingly.

1. 0-4 hours ->Early morning
2. 5-9 hours ->Morning
3. 10-16 hours->Afternoon
4. 17-21 hours->Evening
5. 22-23 hours->Night





Problem Identification – Morning and Evening

- The no. of supplies given to the no. of demands made by the customer is known as supply demand gap.
- During morning slot the cancellations are having higher supply demand gap when compared to cancellations in another slots.
- During evening slot the no cabs availability is having higher supply demand gap of all other slots.the no. of requests made in each slot are:

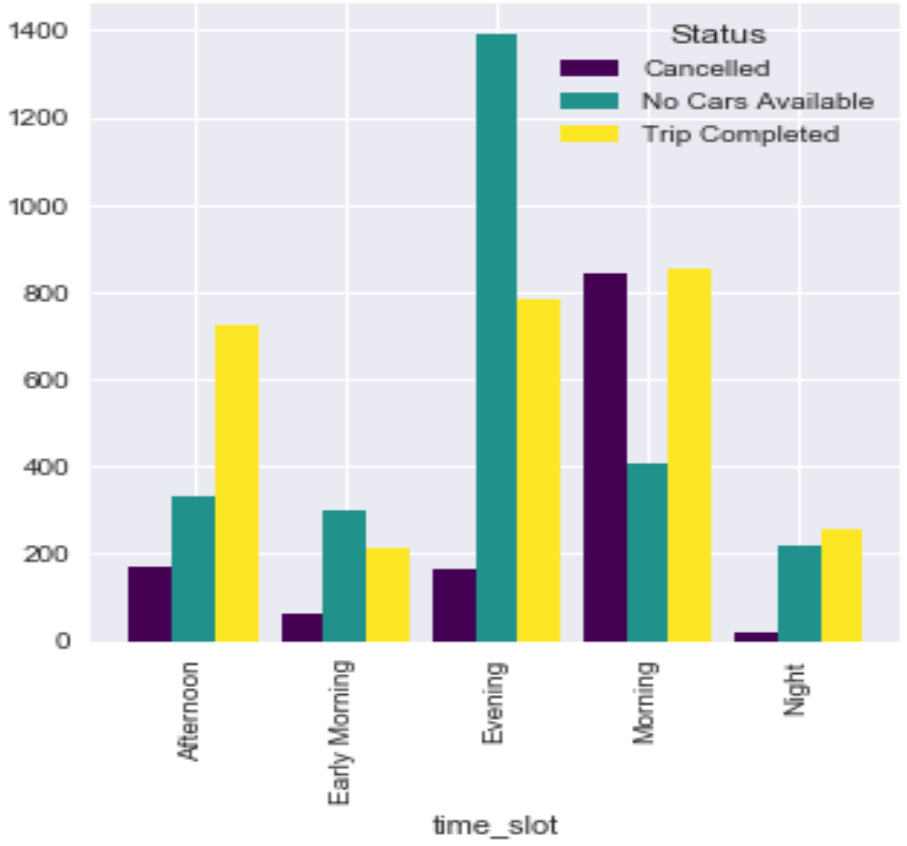
Evening =2342

Morning =2103

Afternoon=1224

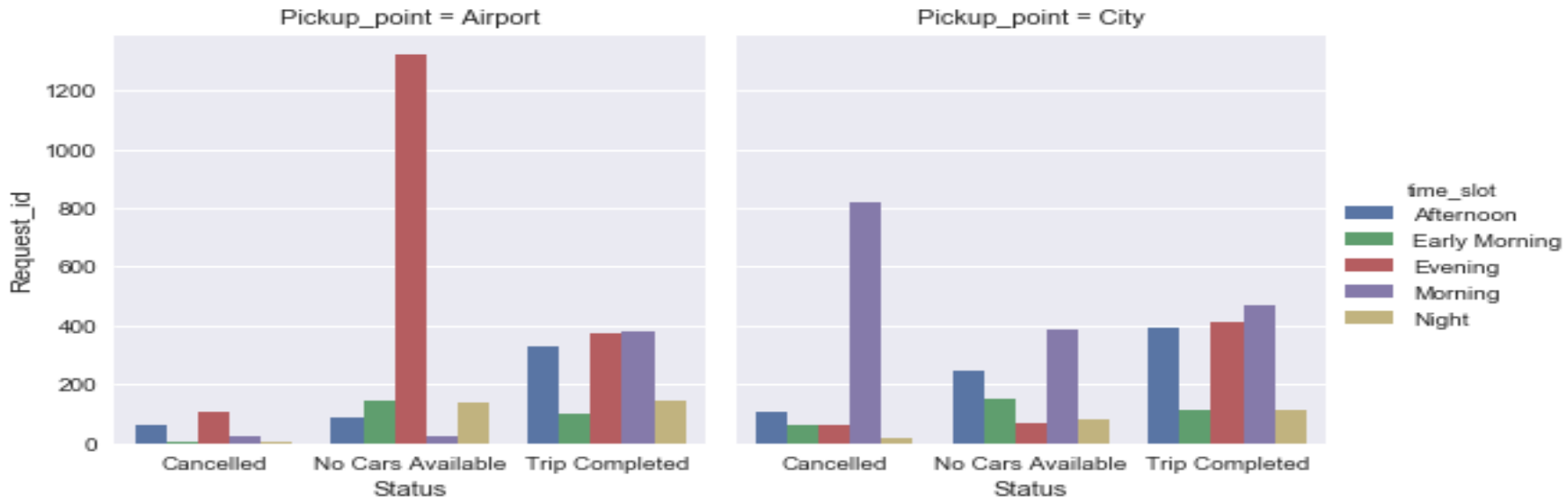
Early Morning= 578

Night= 498



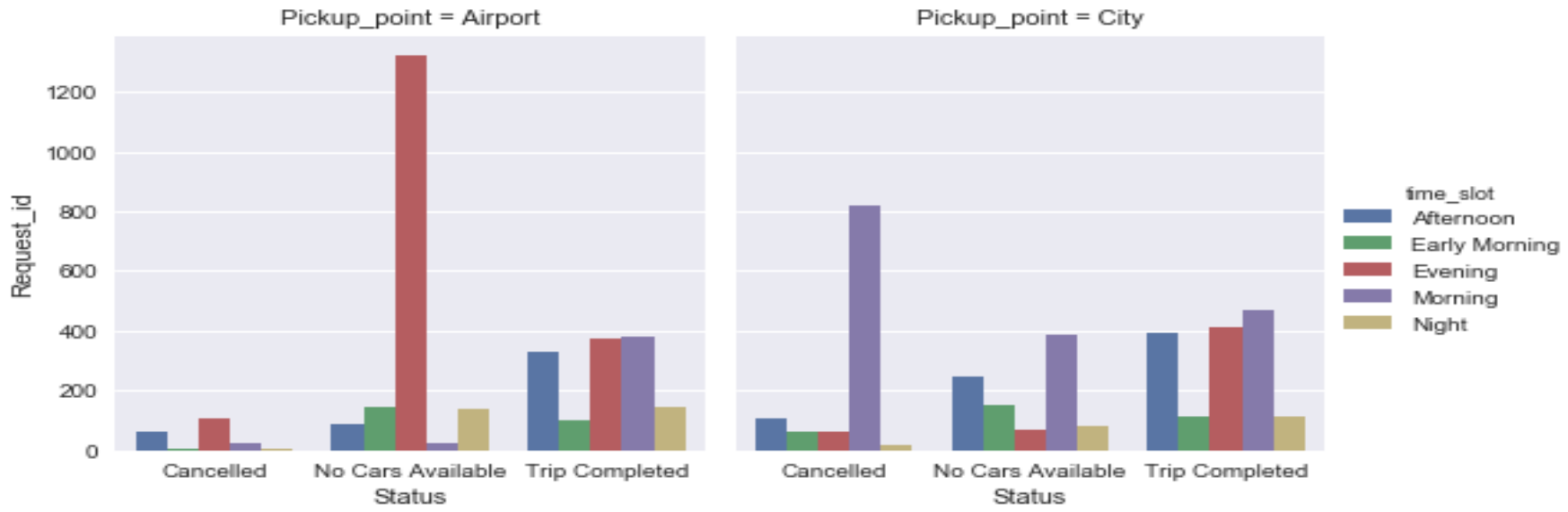
Problem 1 - Cancelled trips

- A large number of flights leave the airport during morning time slot and there are very less incoming flights in morning slot .A driver who reaches airport during that time has to spend idle time to pick a customer back to city. The driver could utilize this idle time for other trips if he chooses not to go to the airport. Otherwise he has to return back empty seated which is a waste of gas mileage for him. Due to this a large number of service requests were cancelled in morning resulting in huge supply demand gap.



Problem 2 - No Cars Available

- At the airport incoming flights are less during evening slot. As the outgoing flights are less, the cabs coming to the airport are also very less during that time. This is drastically reducing the availability of cabs at airport in the evening slot. As the incoming flights are more , the passengers are also more in the evening. These passengers are not getting sufficient cabs to leave the airport in the evening. This is leading to a huge supply demand gap at the airport in evening time slot.



Recommendations

- Recommendations for morning slot from city:
  - Reducing the percentage charged from cab drivers for utilization of Uber services for making a trip to the airport.
  - Charging more money from the customers for trips to the airport and rewarding the drivers accordingly.
  - Share this data with cab drivers and customers to better understand the issue.

- Recommendations for evening slot from airport:
  - Reducing the percentage charged from cab drivers for utilization of Uber services for making a trip to the city.
  - Charging more money from the customers for trips to the city and rewarding the drivers accordingly.
  - Share this data with cab drivers and customers to better understand the issue.