# UBER SUPPLY DEMAND GAP

Shankhadeep Ghosal

# Data Exploration

Data Set contains the following column :

```
> uber_df<-read.csv("Uber Request Data.csv")
> str(uber_df)
'data.frame':   6745 obs. of  6 variables:
 $ Request.id       : int  619 867 1807 2532 3112 3879 4270 5510 6248 267
 $ Pickup.point     : Factor w/ 2 levels "Airport","City": 1 1
 $ Driver.id        : int  1 1 1 1 1 1 1 1 1 2 ...
 $ Status           : Factor w/ 3 levels "Cancelled","No Cars Available",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Request.timestamp: Factor w/ 5618 levels "11/7/2016 0:00",..: 81 266 1563 1252 2023 2790 3162 4389 5108 670 ...
 $ Drop.timestamp   : Factor w/ 2598 levels "11/7/2016 0:51",..: 79 201 923 754 1125 1439 1599 2097 2400 412 ...
>
```

uber_df_clean.csv.Pickup.point

There are six attributes associated with each request made by a customer:

- 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

# Problem Statement

The aim of analysis is to identify the root cause of the problem (i.e. **cancellation and non-availability of cars**) and recommend ways to improve the situation. As a result of your analysis, you should be able to present to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

# Data Cleaning and Preparation

Data issues :

| Request.id | Pickup.point | Driver.id | Status | Request.timestamp | Drop.timestamp |
|---|---|---|---|---|---|
| 0 | 0 | 2650 | 0 | 0 | 3914 |

Driver.id and Drop.timestamp are having NA values.

- After analysing the data set we can conclude that where the status is “Cancelled” or “No car available” for those records Drop.timestamp is having NA values. So , It is valid.
- Driver.id is unique. Thus we can’t replace NA values here.

Request.timestamp and Drop.timestamp are having inconsistent values (Dates and time)

- Converted all the values in a single data time format. Ex: 11/7/2016 11:51 to 2016-07-11 11:51:00

# Data Cleaning and Preparation

Derived Data :

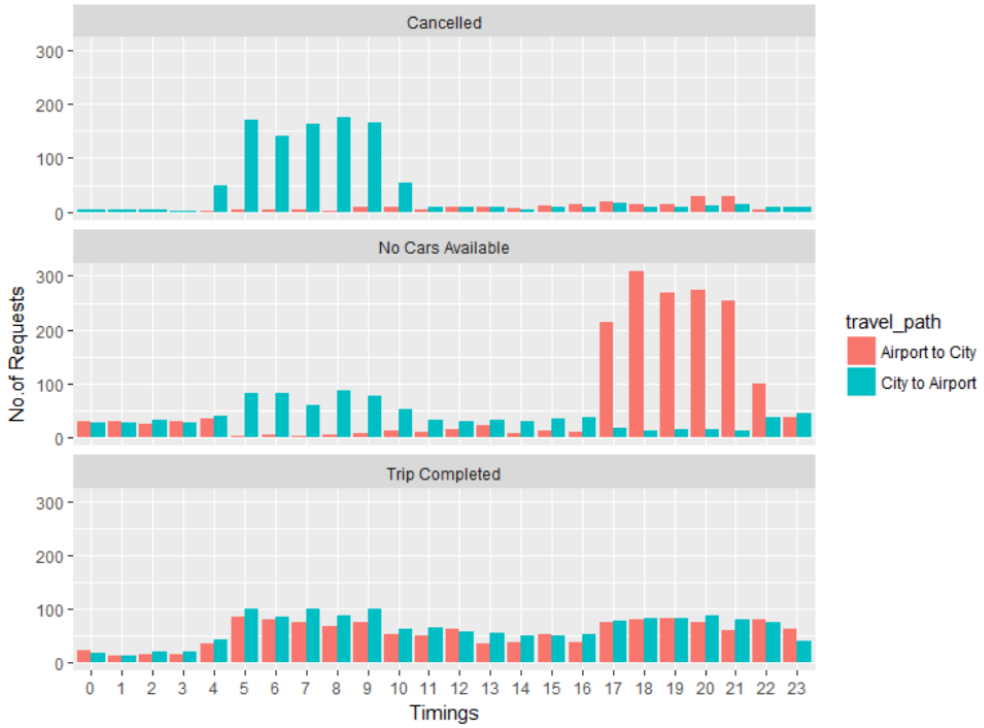
- From the Request.timestamp we have derived a new data field named time_day which categorized the available data to “Late Night”, “Early Morning”, “Morning”, “Afternoon”, “Evening”, “Night”
- From the Request.timestamp we have derived a new data field named Weekdays which categorized the available data to weekdays. Ex: Monday, Tuesday etc.
- From the Request.timestamp we have derived a new data field named Req_hour which categorized the available data to hours. Ex: 12,13 etc.
- From the Pickup.point we have derived a new data field name travel_path. Ex: If the pickup point is City then travel_path is City to Airport (According to the problem statement)

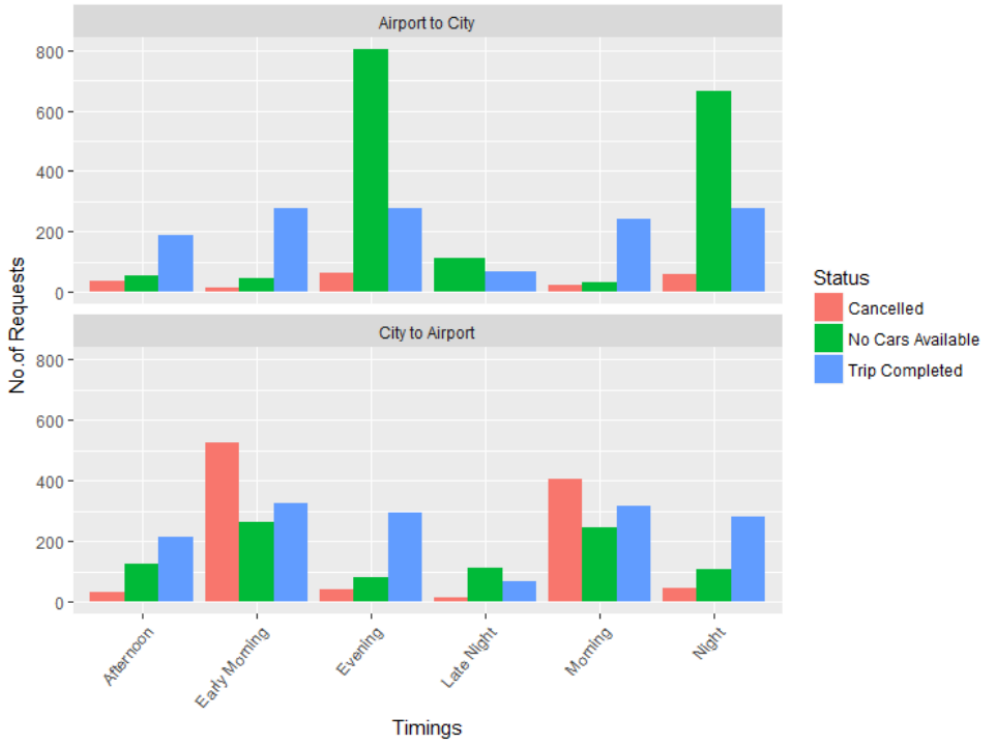
| Hour(24) | Category |
|---|---|
| 0 | Late Night |
| 1 | |
| 2 | |
| 3 | |
| 4 | Early Morning |
| 5 | |
| 6 | |
| 7 | |
| 8 | Morning |
| 9 | |
| 10 | |
| 11 | |
| 12 | Afternoon |
| 13 | |
| 14 | |
| 15 | |
| 16 | Evening |
| 17 | |
| 18 | |
| 19 | |
| 20 | Night |
| 21 | |
| 22 | |
| 23 | |

# Visual Identification of most pressing problem in Uber

Recognition of problems :

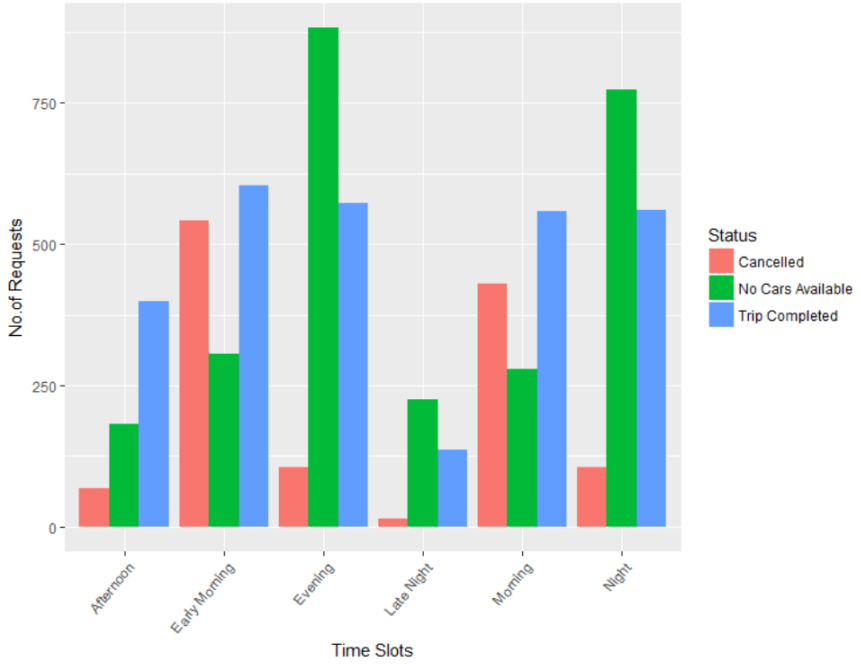
- Availability of car is far less in the evening and at night when pick up is airport.
- Cancellation is much higher in the Early Morning and Morning time when pick up point is City.

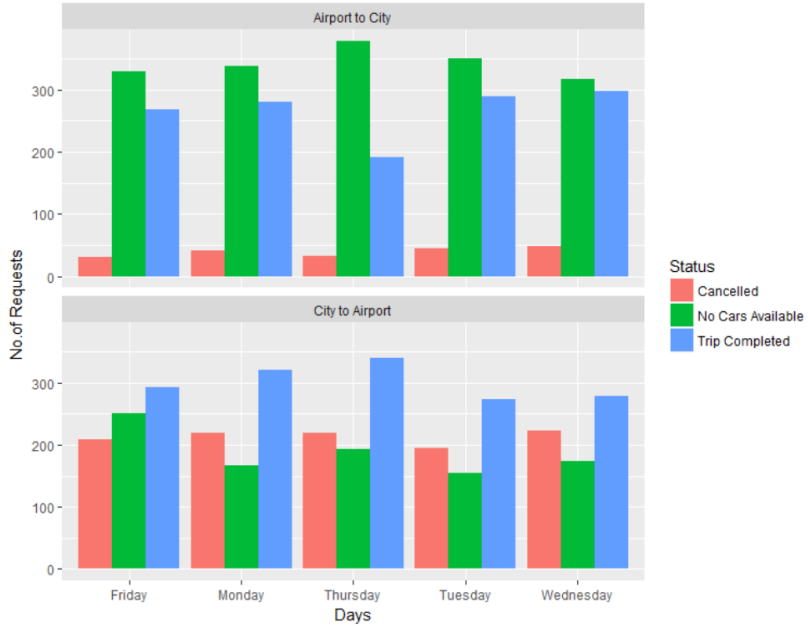


# Visual Identification of most pressing problem in Uber

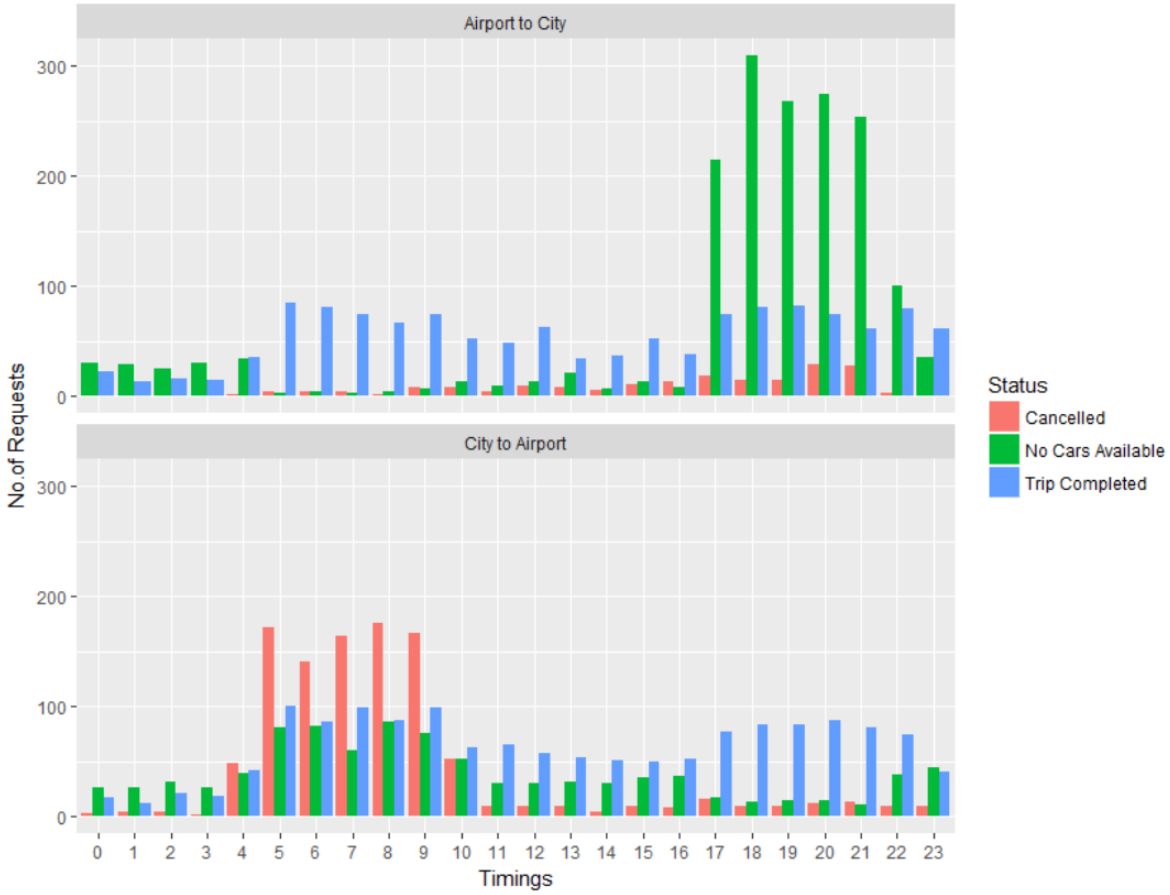
- Overall cancellation is higher in Early Morning.
- Availability of cars is much lower in the evening and at night
- Though through out the week the distribution of status is quite similar but we can observe that Cancellation of trip request is higher in the city-airport section, in the same time availability of car is consistently much lower in airport – city section.



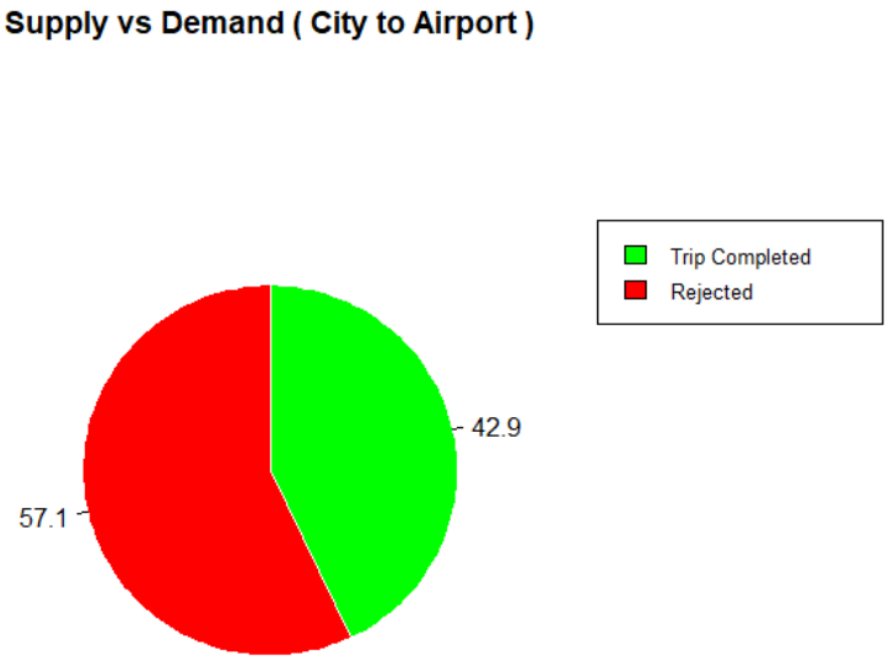
# Supply-Demand Gap Analysis

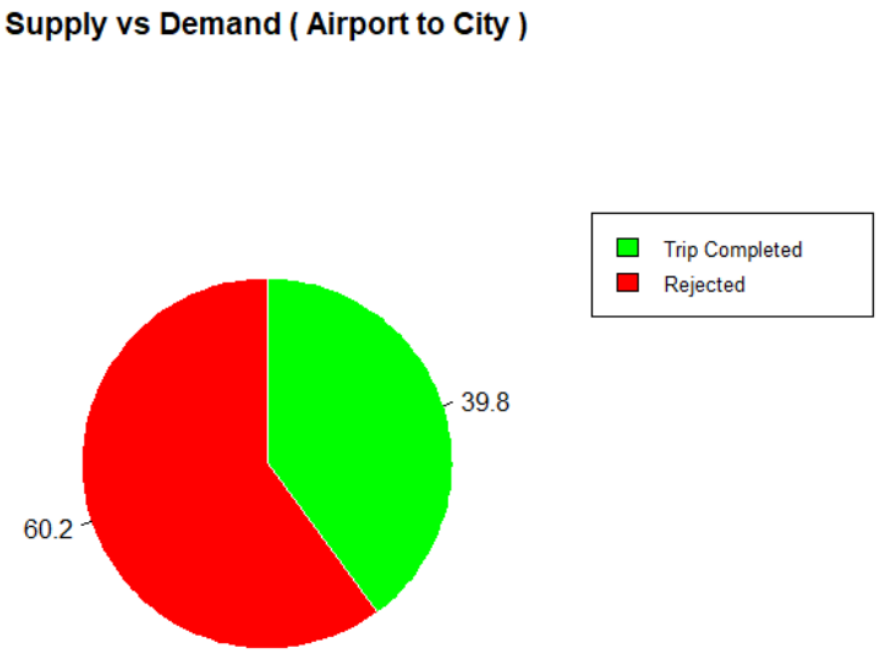
- If we carefully look into the visualization, we can see that demand of Uber car is higher from 17:00 PM to 22:00 PM at Airport, In the same time period the gap between demand and supply is the highest at airport. Notice that in that time slot, demand and supply gap is the lowest when pickup point is City.
- From 04:00 AM to 9:00 AM , in this time slot demand is comparatively higher when pick up point is City than Airport but cancellation of request is the highest in the city. So, we can conclude in this time slot drivers are not ready to go Airport which is creating supply-demand gap.



# Supply-Demand Gap Analysis

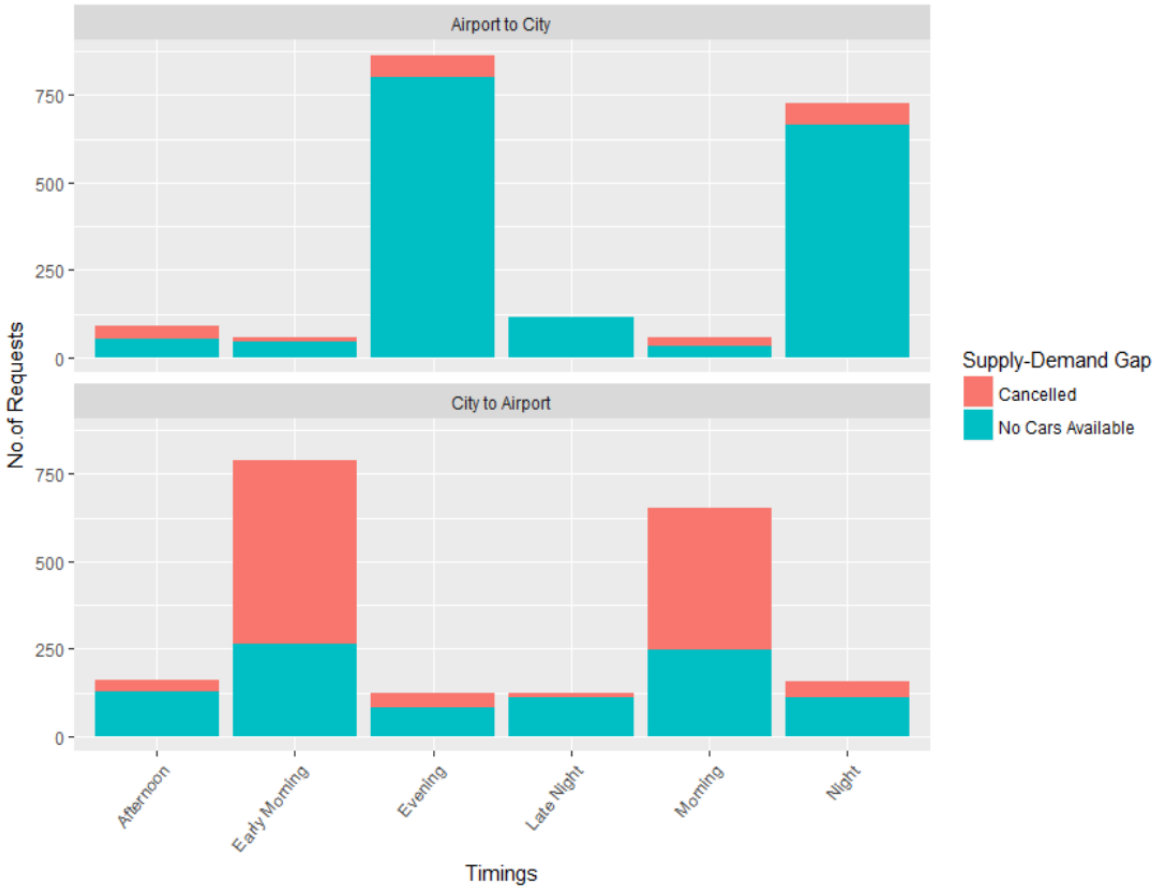
- Supply – Demand gap is higher when pick up point is Airport than the City. **60.2%** of total demand is not fulfilled because of either trip cancellation or non-availability of car.
  Trip Completed= Supply , Demand=Trip Completed+Rejected[Cancelled/No Car Available]



# Supply-Demand Gap Analysis

- In case of ***Airport*** ,**Supply-Demand Gap** is high in The ***Evening*** and at ***Night*** time slot.
- In case of ***City*** ,**Supply-Demand Gap** is high in the ***Early Morning*** and ***Morning*** time slot.



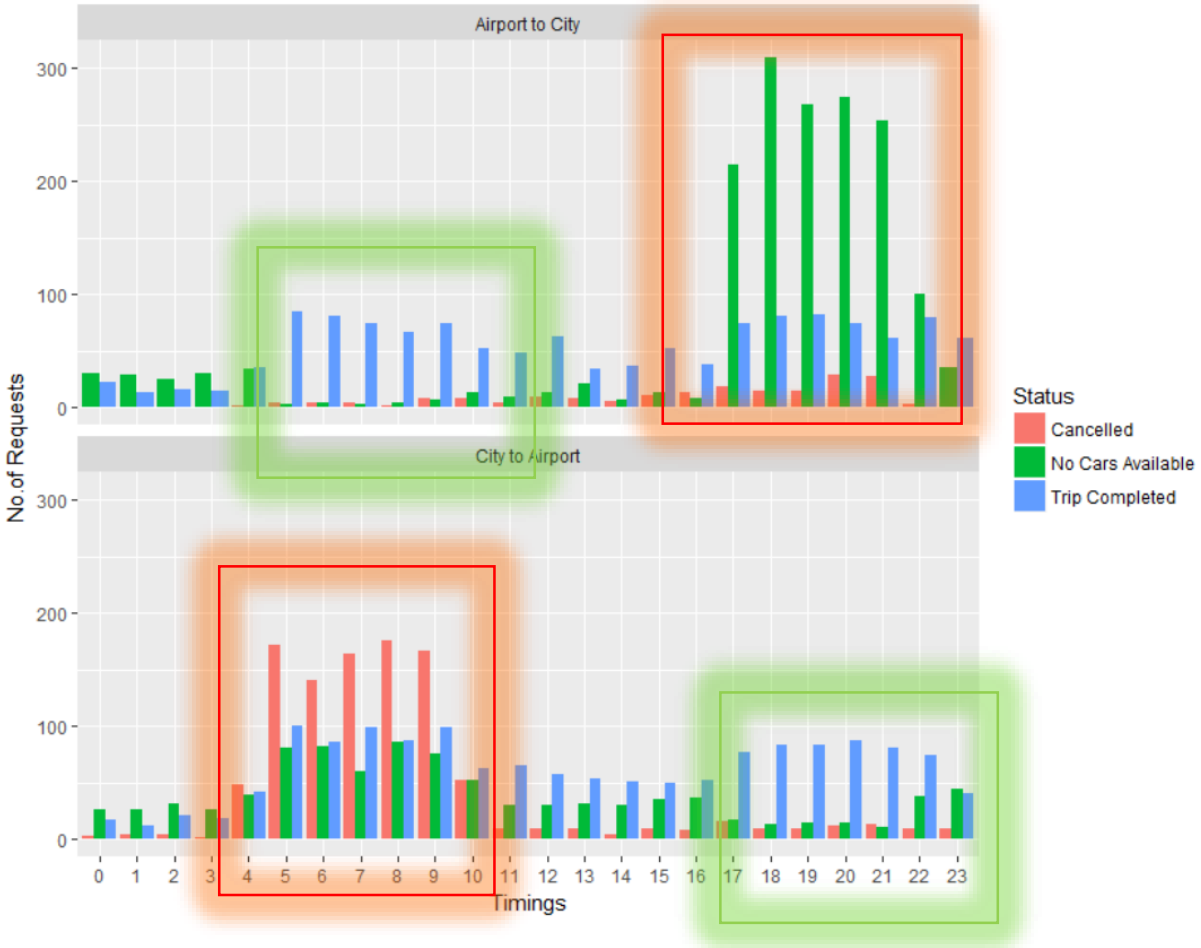

| time_day | Supply-Demand Gap | travel_path |
|---|---|---|
| Afternoon | 91 | Airport to City |
| Early Morning | 59 | Airport to City |
| **Evening** | **864** | Airport to City |
| Late Night | 114 | Airport to City |
| Morning | 58 | Airport to City |
| **Night** | **725** | Airport to City |

| time_day | Supply-Demand Gap | travel_path |
|---|---|---|
| Afternoon | 160 | City to Airport |
| **Early Morning** | **789** | City to Airport |
| Evening | 124 | City to Airport |
| Late Night | 125 | City to Airport |
| **Morning** | **651** | City to Airport |
| Night | 154 | City to Airport |

# Supply-Demand Gap - Reason

- We can observe by carefully analyzing the visualization that drivers are ready to go to Airport from City in the evening and night and in the same time availability of cars is less at airport. So these drivers are waiting at airport after traveling from city to airport for passengers. As the result in the late night and early morning time slot supply-demand gap is lowest at airport.
- In the early morning and morning timeslots drivers are not willing to go to airport as result cancellation is much higher in city area for city-airport trip request. Very general conclusion is in this time slot drivers are preferring intra city trips more than city-airport trips which is causing supply-demand gap

***Evening*** and ***Night***




***Early Morning*** and ***Morning***

# Recommendation

- Pre-booking of cabs if it is Late Night or Early Morning based on the pickup point and make the cab available based on the request.
- Make a mandate rule of highest no. of cancellation from driver's side if the trip is City to airport or vice-versa.
- Provide surprise extra allowance to the drivers for trips when supply-demand gap is increasing.
- When supply-demand gap is higher at airport, a possible solution is to mandate start point as airport for some cabs.
- Another innovative way is to start pool-car system under uber for personal cars. As a example a person who is having a car and returning from airport or going to airport can pick up some passengers with him. It will reduce supply-demand gap.