# UBER CASE STUDY

# SUBMISSION

Name: Meenu Jomi

# • Abstract

- The objective is to identify  and solve the most pressing problems for Uber has been facing.
- We have considered only the trips from the city to the airport and vice versa.
- The main aim is to find how many trips get completed or cancelled and also the situation in which no cabs are available.
- We also need to find the Gap between the supply and demand
- All these issues need to be decided by comparing against the drop and pick up time frames and the different days

# Problem solving methodology (EDA)

- Data Sourcing
- Data Cleaning
- Univariate Analysis
- Segmented Analysis
- Bivariate Analysis
- Derived Metrics

# • Data Sourcing / Data Cleaning

- The masked data was provided by Uber
- The removing of null values was skipped since the data is already cleaned and the null values are there because of the following reasons:
  - 'Driver id' has missing values since it is aligned to 'No Cars Available'
  - 'Drop timestamp' has missing values since the cab wasn't taken or available
- Few extra columns were made for the analysis by splitting the 'Request timestamp' and 'Drop timestamp'
- Also, different time slots were provided for them as shown.

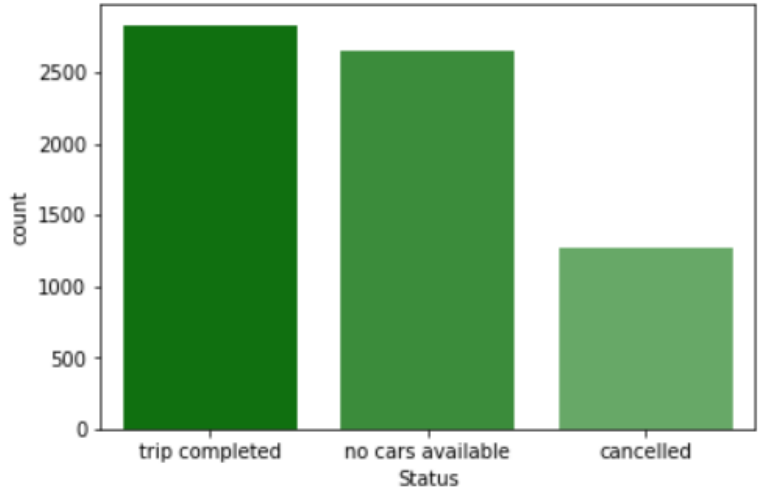
| Time Slot | Labels |
|---|---|
| 01 – 04 | Early Morning |
| 04 – 07 | Morning |
| 07 – 10 | Late Morning |
| 10 – 13 | Early Afternoon |
| 13 – 16 | Afternoon |
| 16 – 18 | Evening |
| 18 – 21 | Late Evening |
| 21 – 23 | Night |
| 23 – 01 | Late Night |

# • Univariate Analysis

The analysis was done in two different ways:

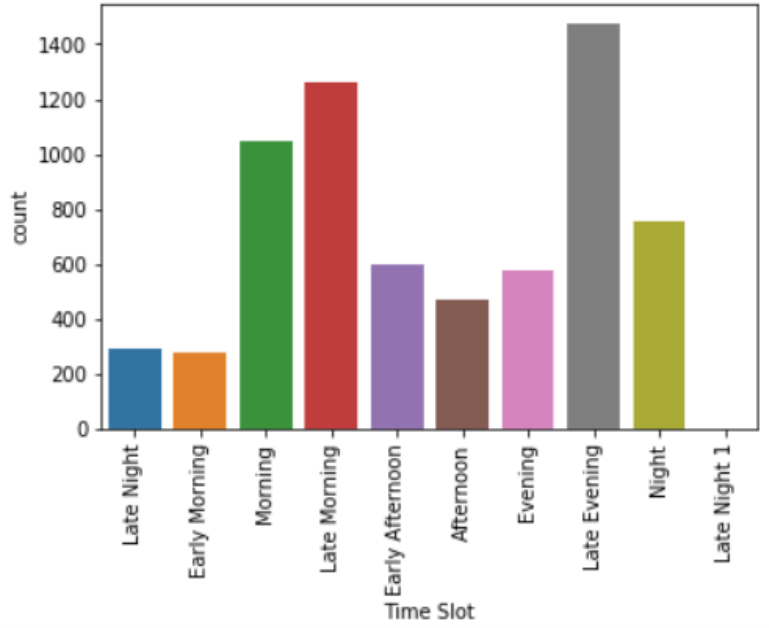
- Unordered Categorical

In this we saw different variables and how their count varied among each other. There wasn't much difference in most of it except in case of 'Status' as shown.




- Ordered Categorical

In this we saw how the timeslots given in the above slide showed changes. A similar result came for both as shown.

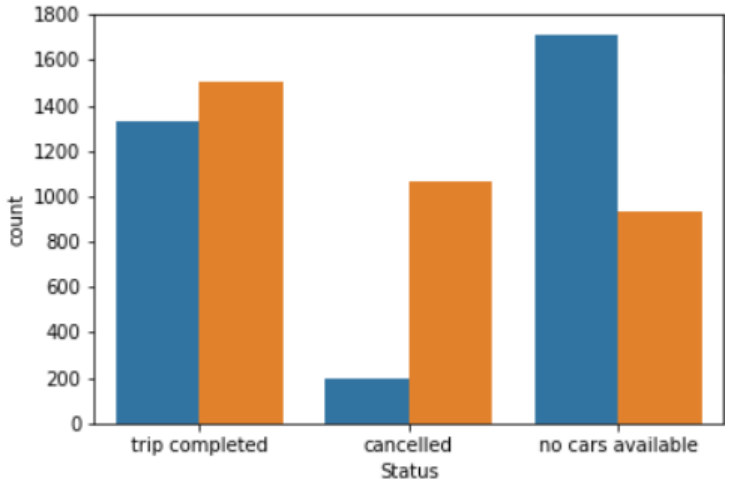


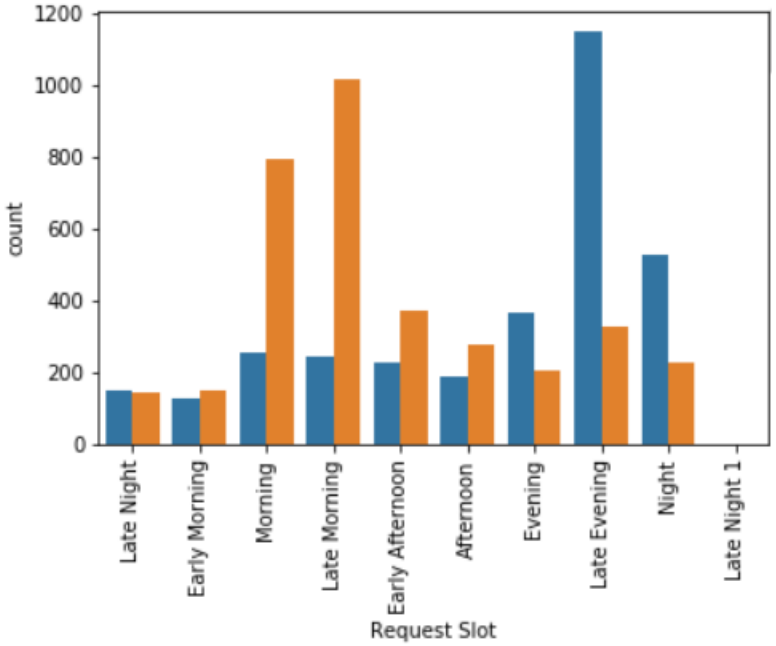

- *More required cabs are in the 'Late Mornings' and 'Late Evenings'*
- *Most of the trips get completed but 'No Cars Available' is almost the same. The no. of cabs cancelled are almost half of them.*

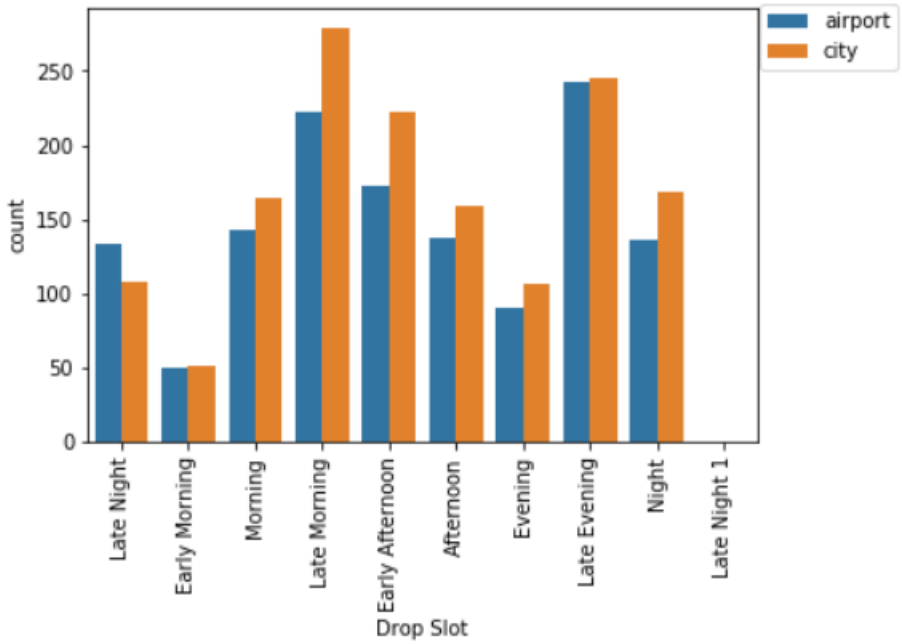
# • Segmented Univariate Analysis

In this grouped variables and plotted against other variables to find results.

'Pickup point' is grouped in the below cases:



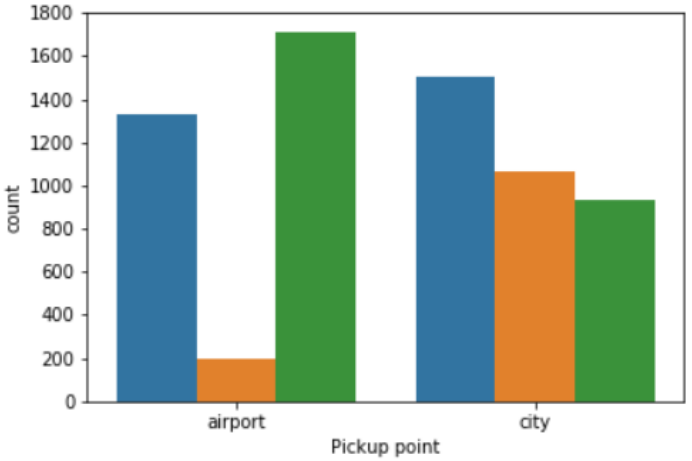



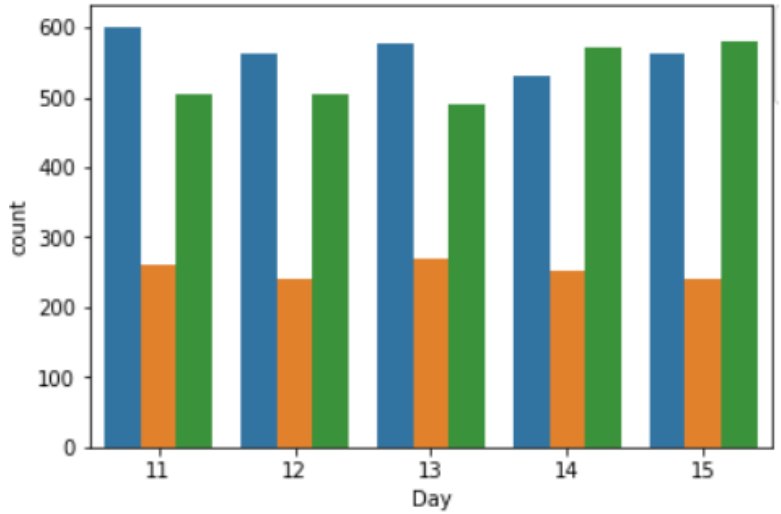

- *There are more 'No Cars Available' at the airport*
- *The cabs are usually 'Cancelled' at the city*
- *More requests of cab from the 'airport' during 'Late evenings' and 'Night' and from the 'city' during 'Mornings' and 'Late Mornings'*
- *More Drop offs at the 'airport' and at the 'city' during 'Late Mornings' and 'Late Evenings'*

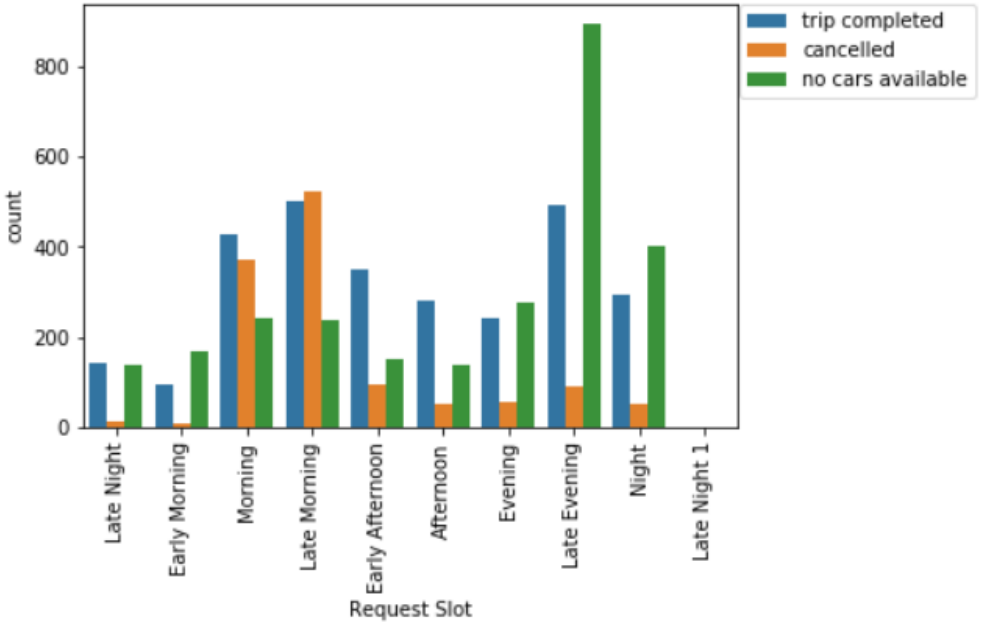
# • Segmented Univariate Analysis

In this grouped variables and plotted against other variables to find results.

‘Status’ is grouped in the below cases:



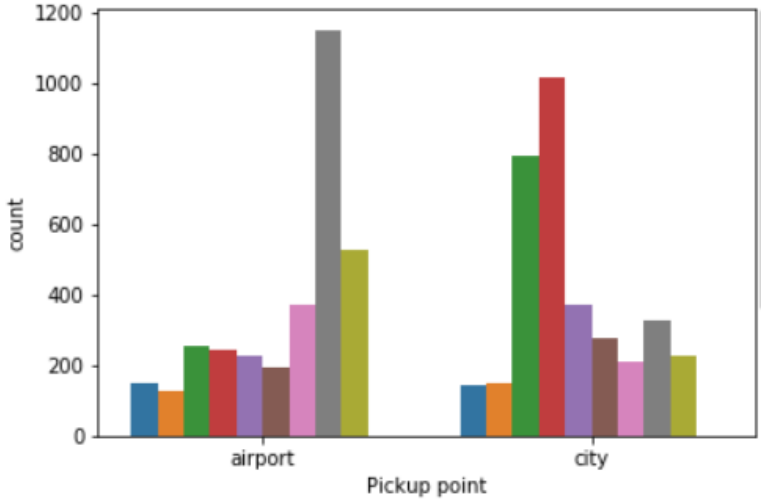



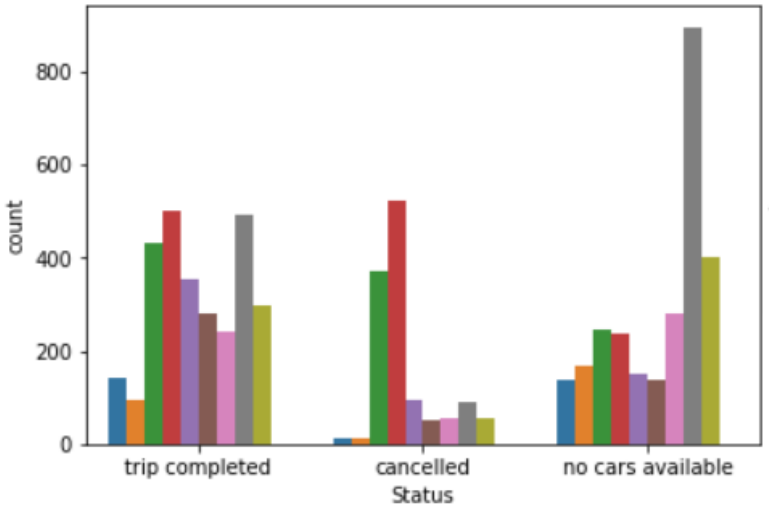

- *There are very few cabs located at the 'airport' and some do get 'Cancelled' as well.*
- *There are many trips from the city but around 55% of the cab are either unavailable or cancelled*
- *The number of 'Trips Completed' is almost equal to the number of 'No Cars Available' Status*
- *Most of the trips get Completed during 'Late evenings', 'Mornings' and 'Late Morning’*
- *Most of the trips gets Cancelled in the 'Mornings'*
- *There are very Few cabs available in the 'Late Evenings'*

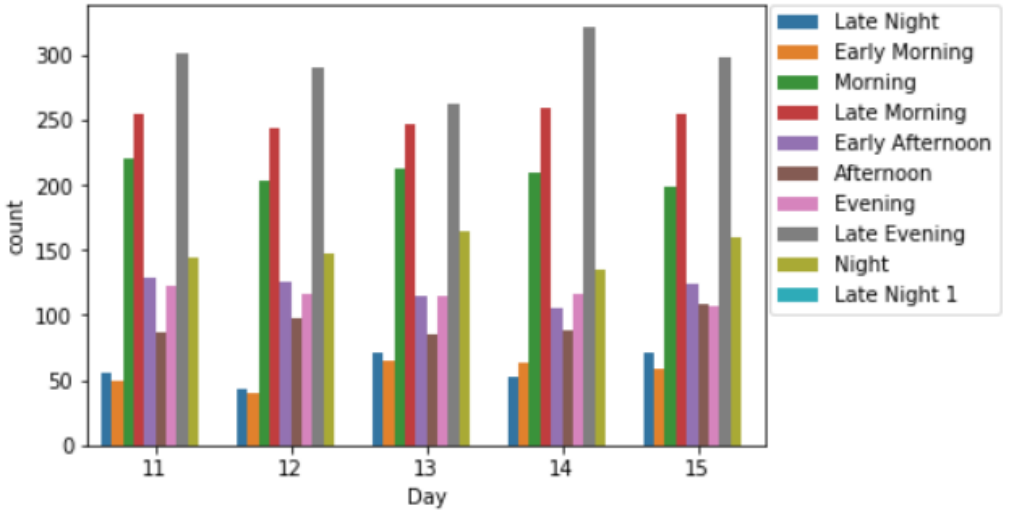
# • Segmented Univariate Analysis

In this grouped variables and plotted against other variables to find results.

'Request Time' is grouped in the below cases:



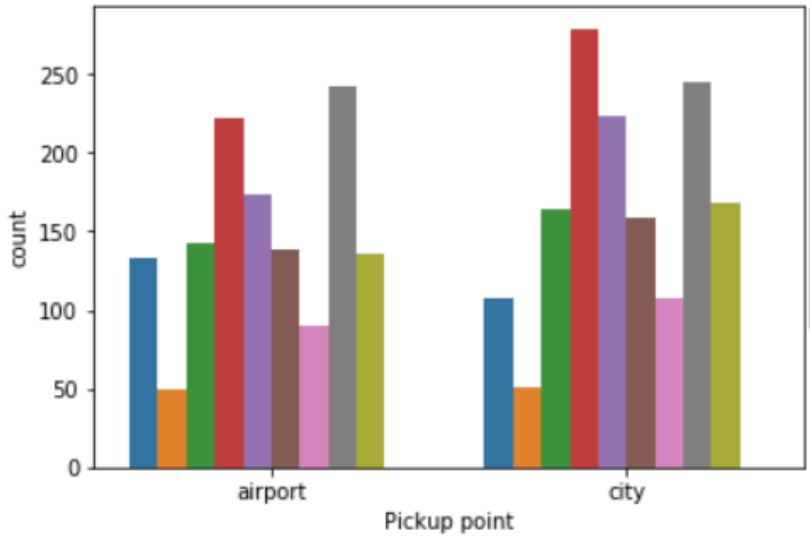



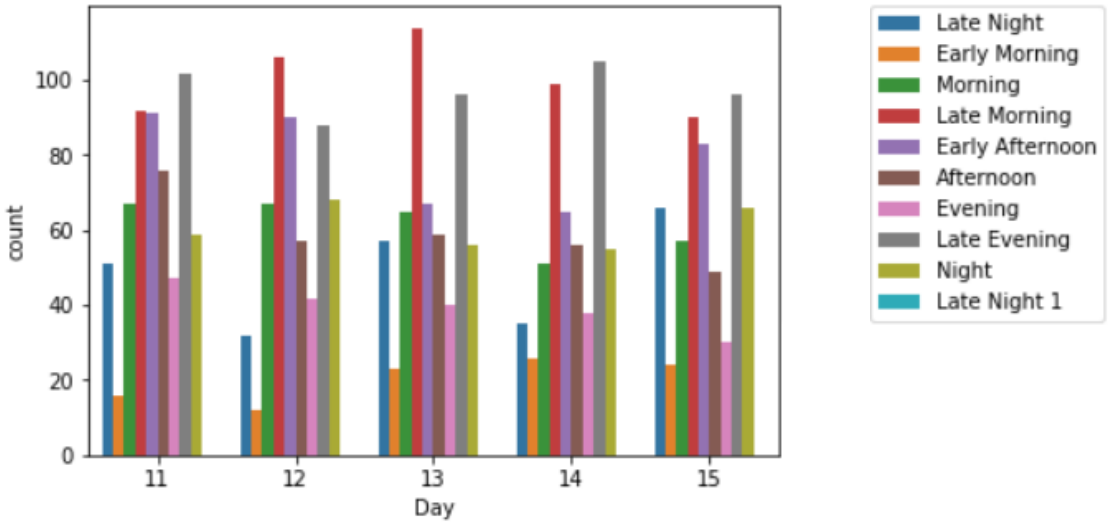

- *There are more trips from the 'airport' during 'Late evenings'*
- *There are more trips from the 'city' during 'Mornings'*
- *Many trips get Completed in the 'Early Morning', 'Mornings' and 'Evenings'.*
- *Most of the trips in the 'Late mornings' get 'Cancelled'*
- *There is a huge Unavailability of cabs in the 'Late Evenings'*
- *On each Day most of the cabs are taken in the 'Late Evenings', 'Mornings' and 'Late Mornings'*

# • Segmented Univariate Analysis

In this grouped variables and plotted against other variables to find results.

‘Drop Time’ is grouped in the below cases:

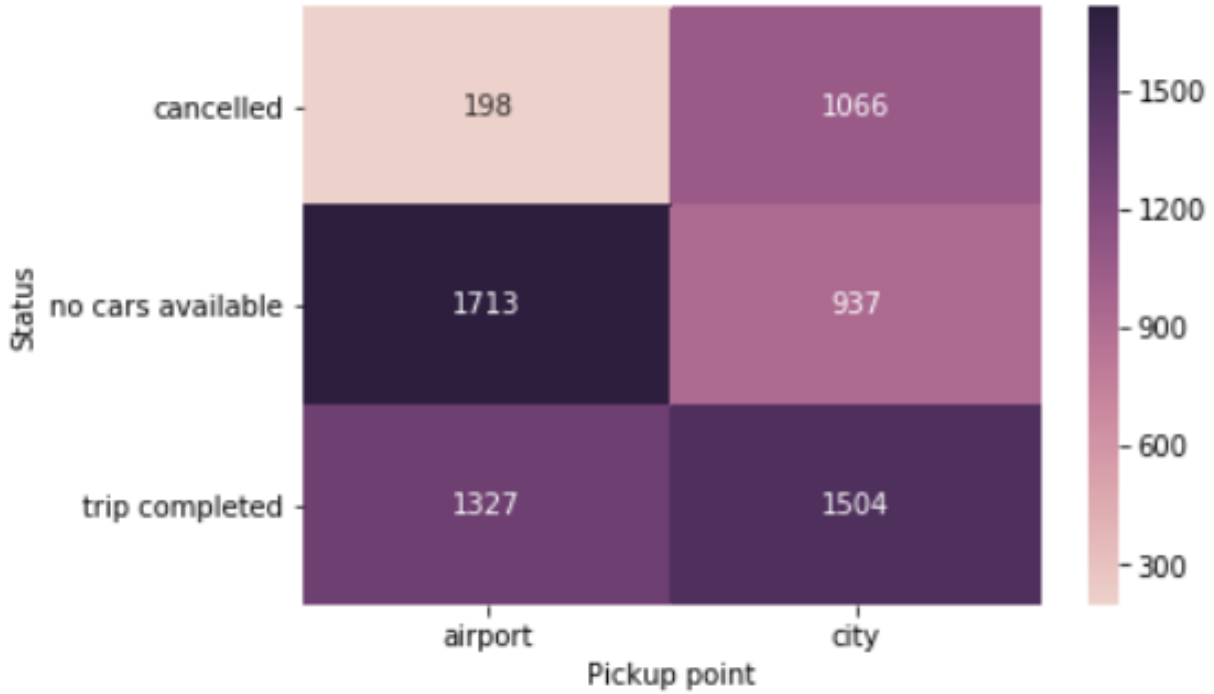


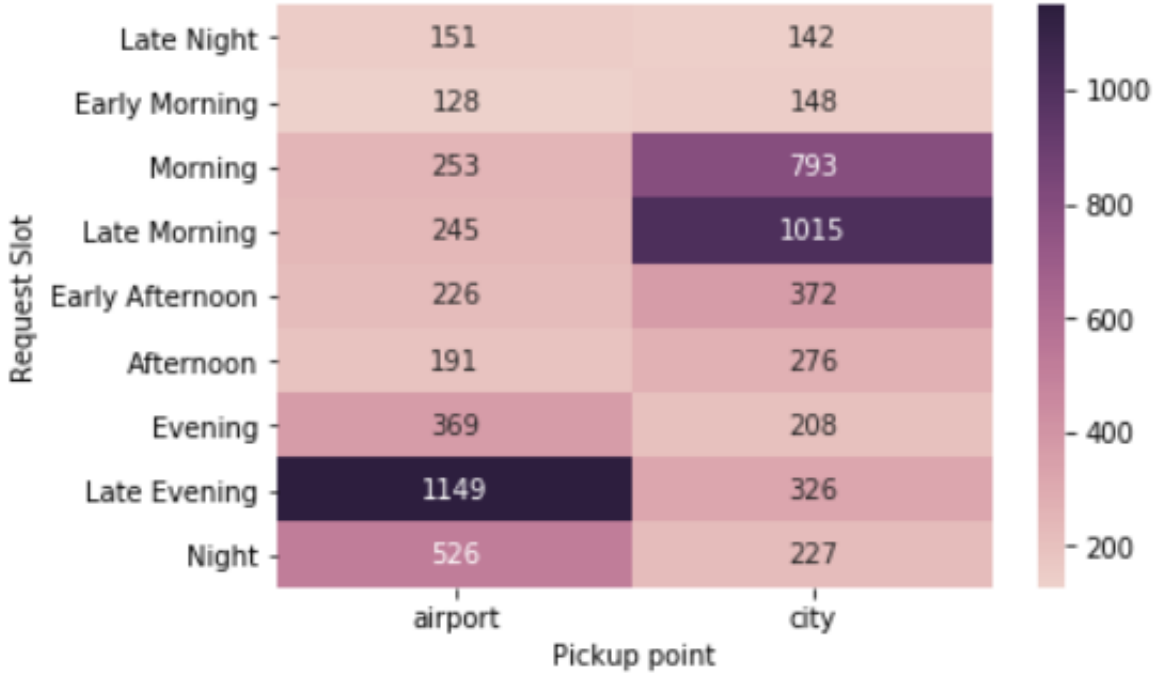




- *In the 'Late Morning' there is a huge number of Drop offs. Then it gradually decreases till 'Evening'. There is a huge increase in Drop offs at 'Late Evening' and later it decreases till the next 'Early Morning'. It carries on the same way from there. The scenario is the same for 'Airport' and 'City‘*
- *In the 'Late Morning' there is a huge number of Drop offs. Then it quickly decreases till 'Evening'. There is a huge increase in Drop offs at 'Late Evening' and later it decreases till the next 'Early Morning'. It carries on the same way from there. The scenario is the same in each Day*

# • Bivariate Analysis

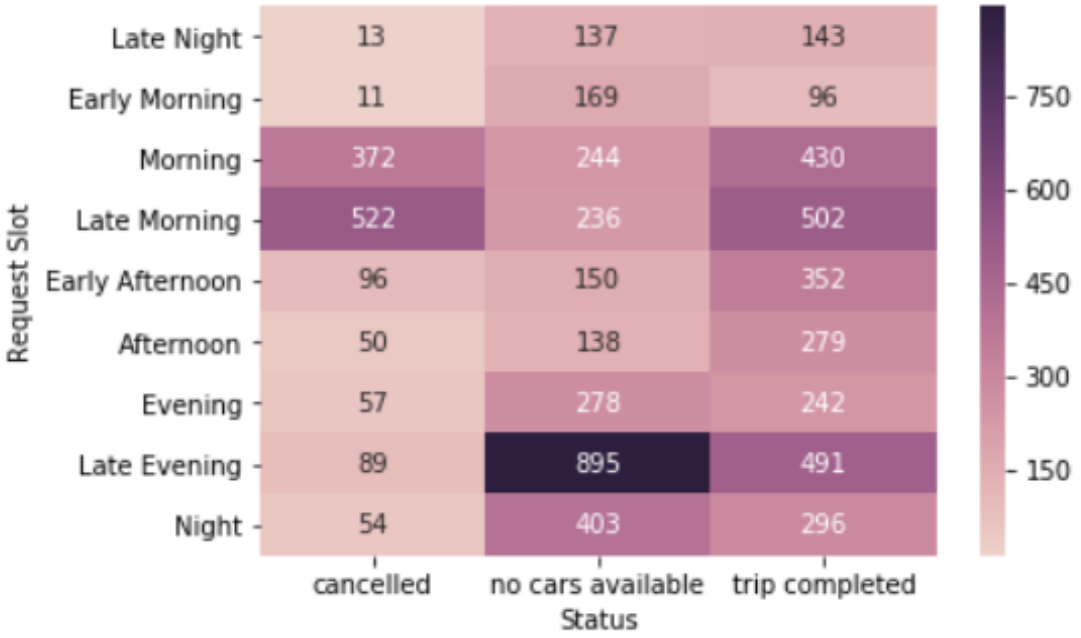
Two Variables are taken against each other to understand the correlation

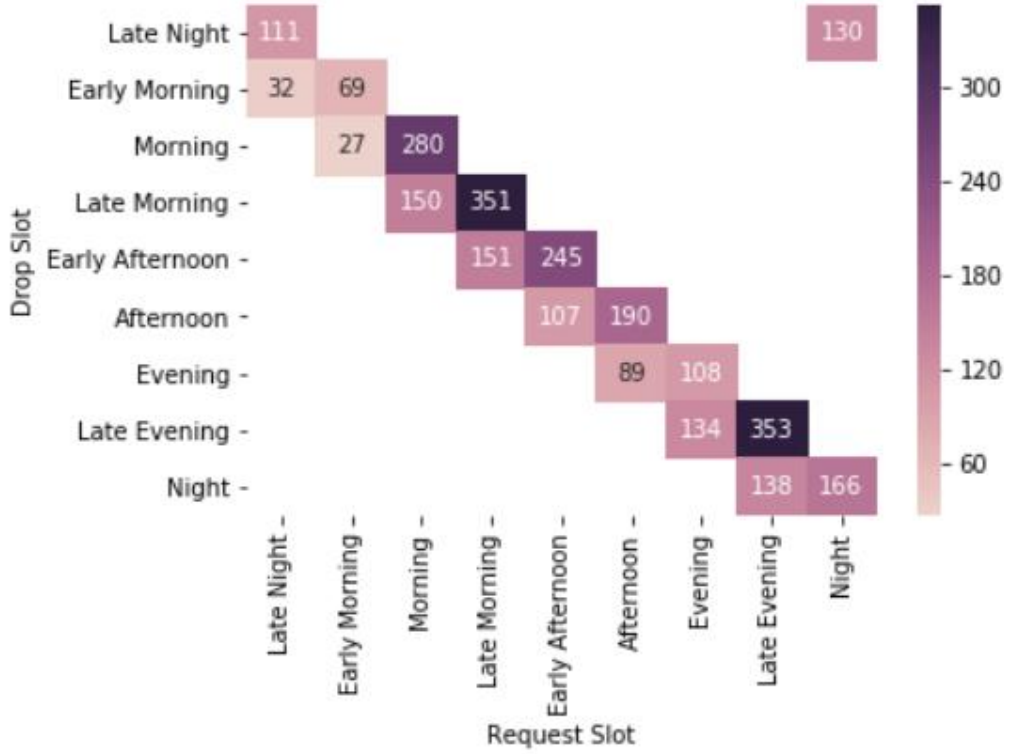



- *It is clear that cars are Unavailable at the 'Airport' and most of the trips either get Cancelled or Completed from the 'City'*
- *There is a huge demand for cabs from the 'Airport' in 'Late Evenings' and from the 'City' in the 'Late Mornings'*

# • Bivariate Analysis



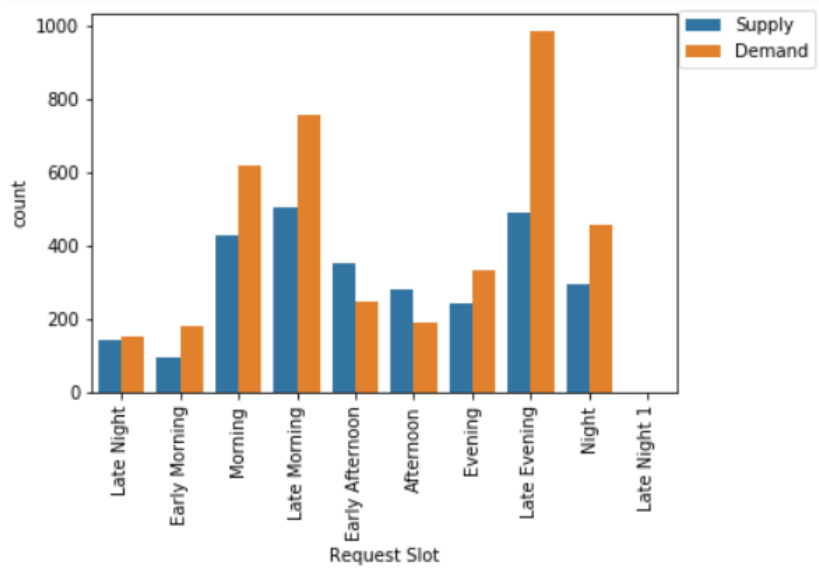

- *Most of the trips get Completed in the 'Mornings' and 'Late evenings'.*
- *There is a huge unavailability of cars in the 'Late Evenings'.*
- *Many cars get Cancelled in the 'Late Mornings'.*
- *It is visible that most of the trips get Requested and Dropped at the same time frame. Some of them move to the next time slot. This might be because the distance is huge between 'Airport' and 'City' or at the extremes of the created time slot*

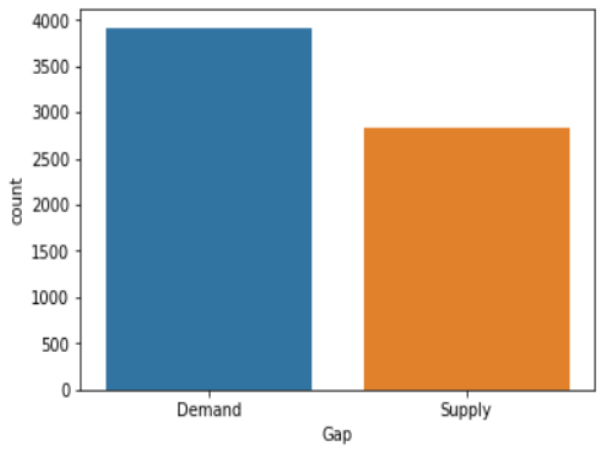
# • Derived Metrics

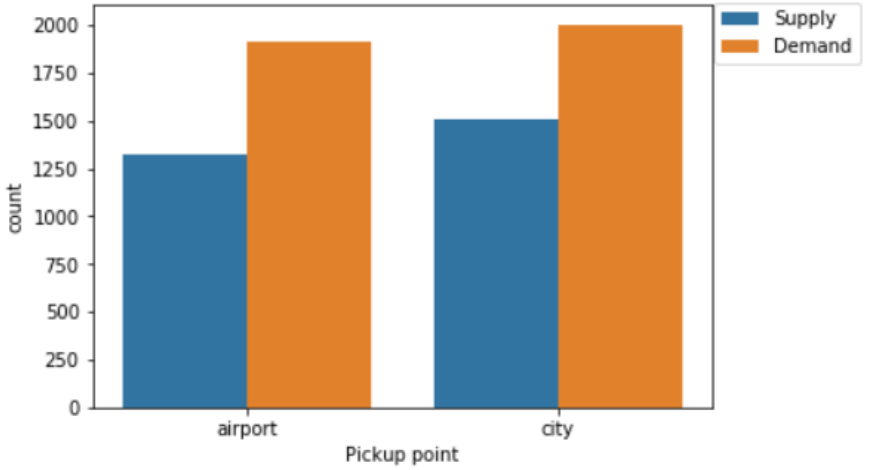
**Business Driven Variable:**

The data was divided into two:

- 'Supply' – When the status was 'Trip Completed'
- 'Demand' – When the status was 'No Cars Available' and 'Cancelled'




- *There is a huge 'Demand' and the 'Supply' is very low*
- *The highest Gap is in the 'Late Evening'*
- *It is clear that the difference is the most in 'Airport' although there is much greater demand in the 'City'*

# • Conclusion

| Issues | Solution |
|---|---|
| Many trips from the city but around 55% of the cab are either unavailable or cancelled. | Extra charge should be given to the driver who takes up cabs from the city. |
| The number of 'Trips Completed' almost equal to 'No Cars Available' Status. | More cabs should be owned by Uber and some of it can be cabs just dedicated for airport services. |
| Requirement is during 7am to 10am and 6pm to 9pm | Extra charges or badge should be given to the drivers who do trips during this time. |
| Most trips from the 'airport' during 6pm to 9pm and there is a huge unavailability. | There must be dedicated cabs just for airport services during this time at the airport. |
| Most trips from the 'city' during 7am to 10am and most of the trips gets Cancelled. | The trip to the airport at this time should be given extra charge to the drivers that encourage them to accept rides. |
| It is visible that some Requests and Drops are at different time frame. This might be because the distance. | Nothing can be done to distance. But the ones that are very far can be costed more thus the driver will also be charged extra. |
| It is visible that there is a huge 'Demand' and the 'Supply' is very low | Encouraging the drivers and increasing the number of cabs will provide a solution for this problem. |