UBER CASE STUDY
SUPPLY-DEMAND GAP

LAKSHYA AGARWAL AND SAYAK BHOWMICK

BUSINESS OBJECTIVES

- To identify the **root cause of the problem** (i.e. cancellation and non-availability of cars).
- Analyze data on basis of various pre-defined & derived parameters and make some hypothesis in basis of the findings.
- **Recommend** ways to **improve the situation**.

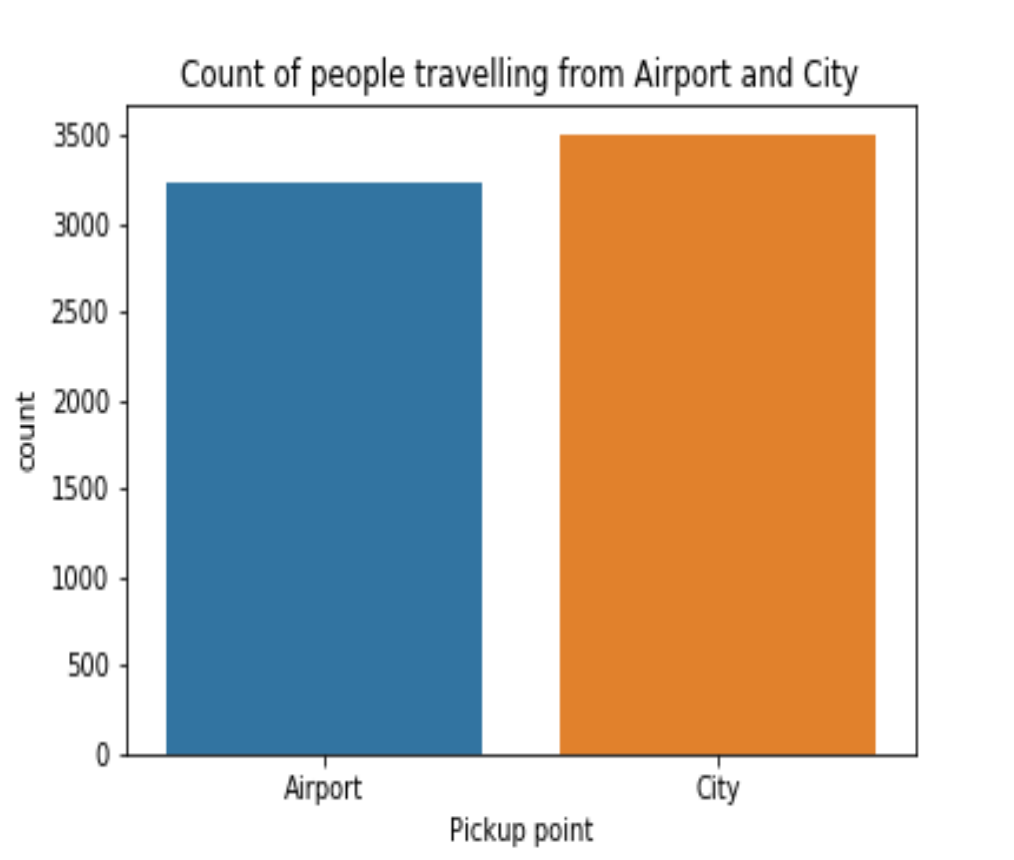
UNDERSTANDING THE DATA

- **Request ID** : A unique identifier of the request
- **Request timestamp** : The date and time at which the customer made the trip request
- **Drop timestamp** : The drop-off date and time, in case the trip was completed
- **Pickup point** : The point from which the request was made
- **Driver id** : The unique identification number of the driver
- **Status** : The final status of the trip, that can be either completed, cancelled by the driver or no cars available

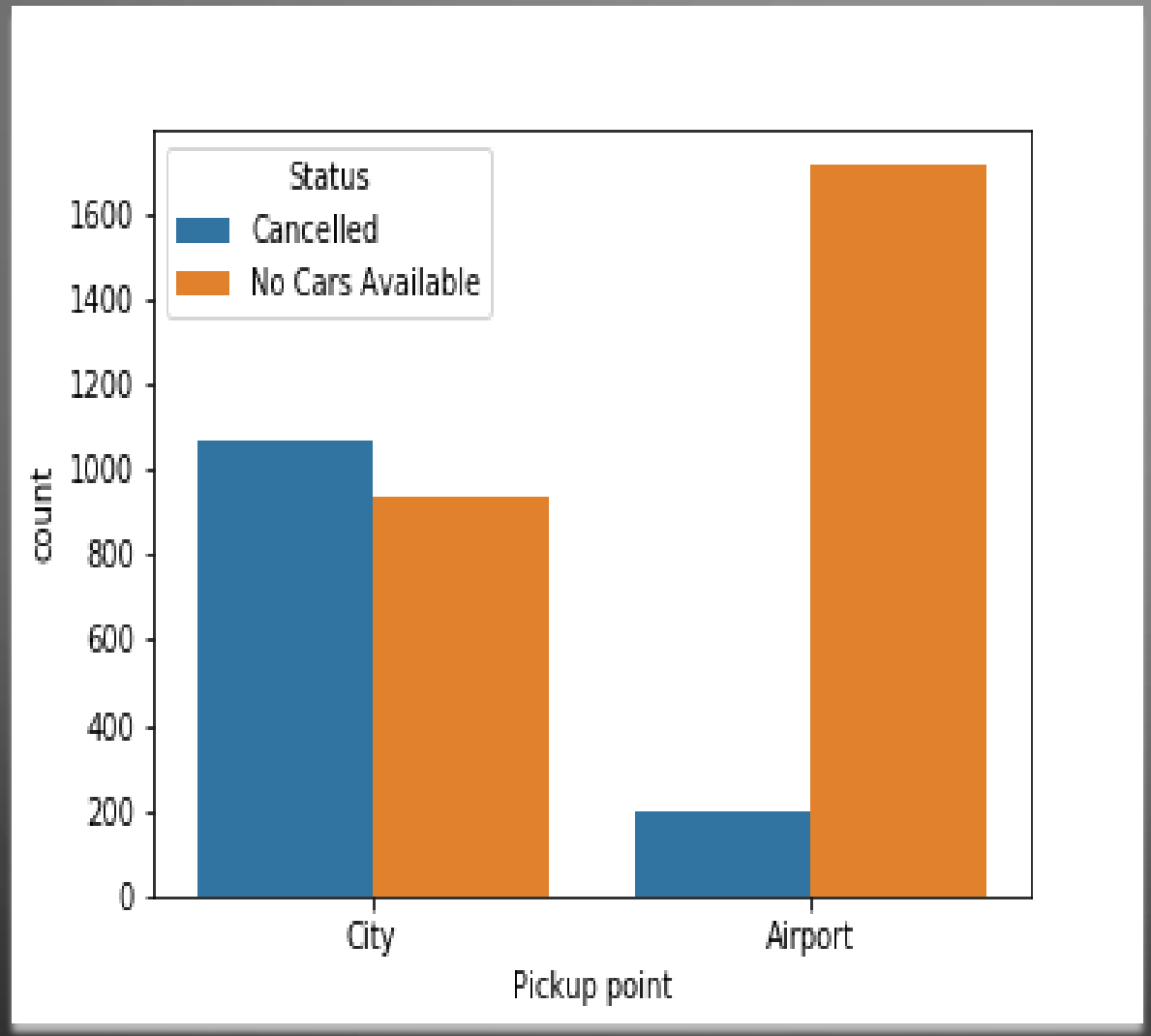
DATA CLEANING AND PREPARATIONS

- Only two columns named ‘Driver id’ and ‘Drop timestamp’ had null values which were very useful for analysis so there was no need to clean the data.
- Changed the datatypes of columns having date & time. (from object to datetime)
- Using various plots to understand the data in a better way.

UNIVARIATE AND SEGMENTED UNIVARIATE ANALYSIS ON PICKUP POINT

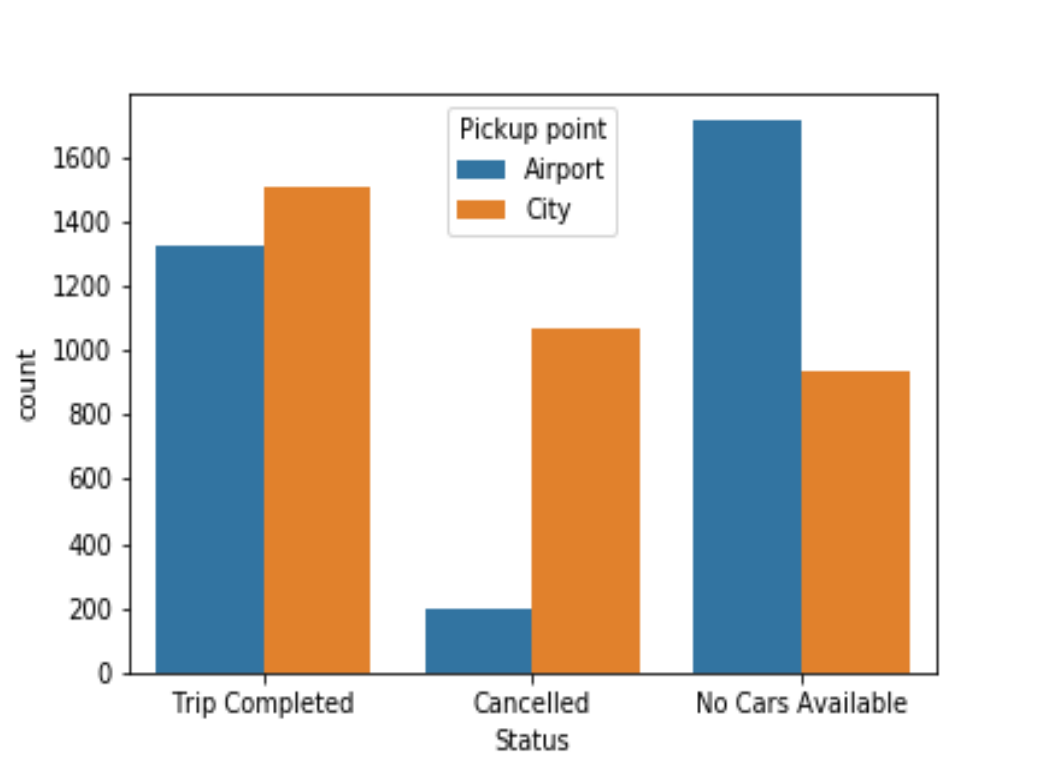


- These are the total number of requests made by the customers from both of the pickup points.
- The number of requests for both the pickup points are approximately same.
- Although, the requests made from City are a bit more than that of Airport.



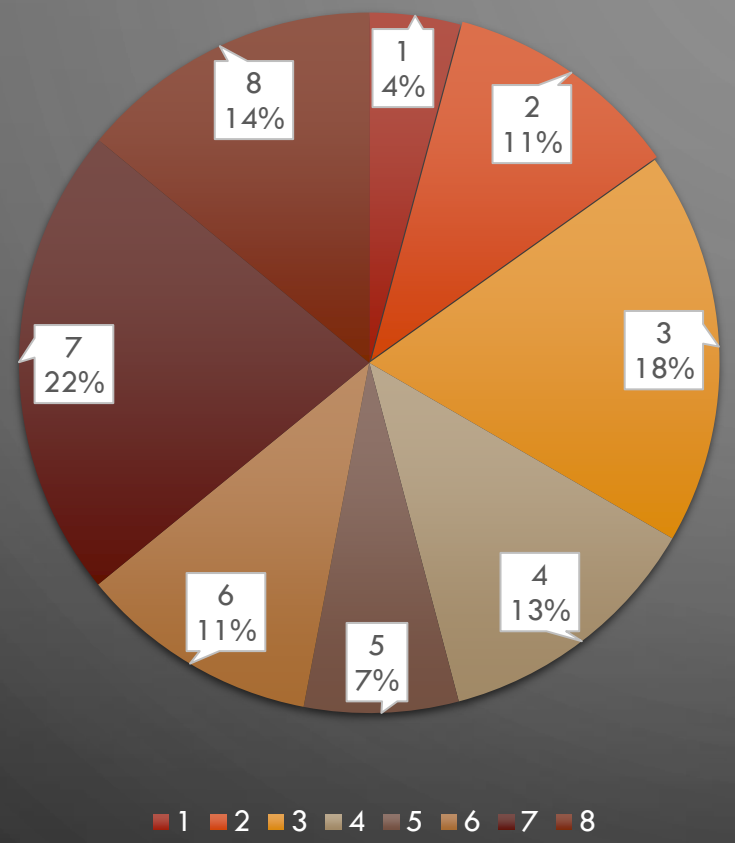
- Non availability of cars is a huge problem at Airport.
- Frequency of Cancellation and non availability is almost same at city.
- The condition in city sounds more problematic as the cancellation frequency is much greater than that of Airport.

SEGMENTED UNIVARIATE ANALYSIS ON THE STATUS OF CAB RIDES



- Trip completion is almost equal at both the pickup locations but a little bit more at City.
- The cancelled rides are pretty high in the City than that of Airport comparatively.
- Non-availability of cabs are too high at Airport than that in the City.
- Non-availability at the City is also not negligible as it is very close to the cancelled rides at City.

UNIVARIATE ANALYSIS ON THE 8 TIMESLOTS OF THE DAY

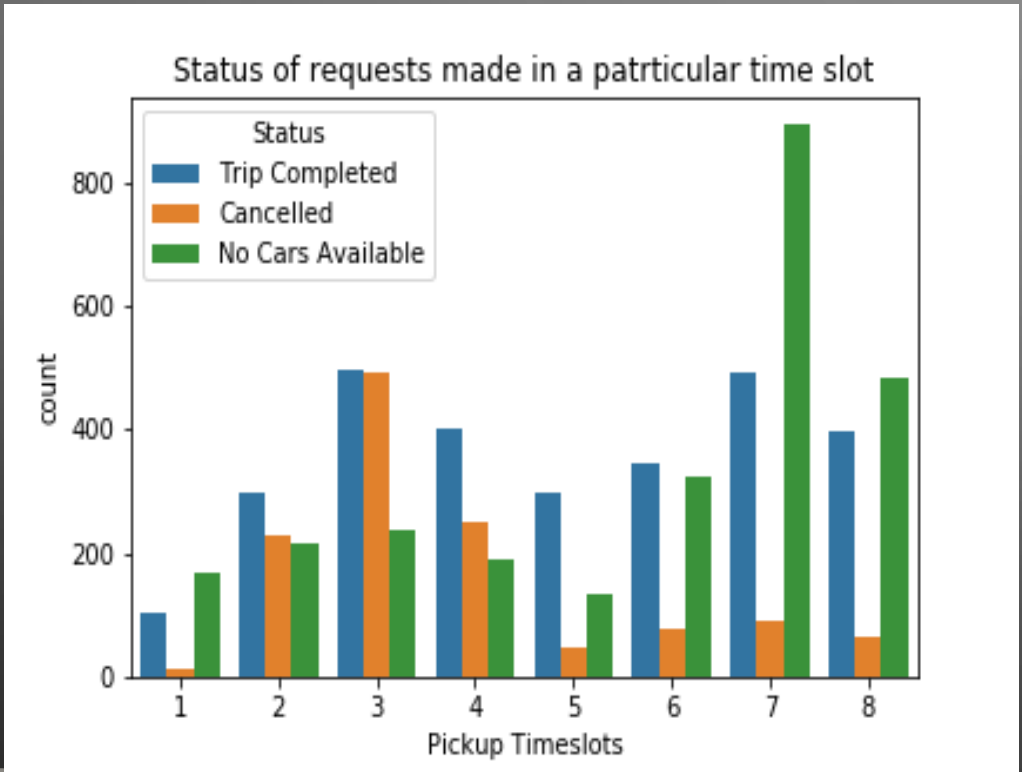


- Divided the 24 hours of the day in 8 timeslots.

Slot 1 : Midnight (00:00 - 02:59)
Slot 2 : Early Morning (03:00 - 05:59)
Slot 3 : Morning (06:00 - 08:59)
Slot 4 : Early Afternoon (09:00 - 11:59)
Slot 5 : Afternoon (12:00 - 14:59)
Slot 6 : Early evening (15:00 - 17:59)
Slot 7 : Late evening (18:00 - 20:59)
Slot 8 : Night (21:00 - 23:59)

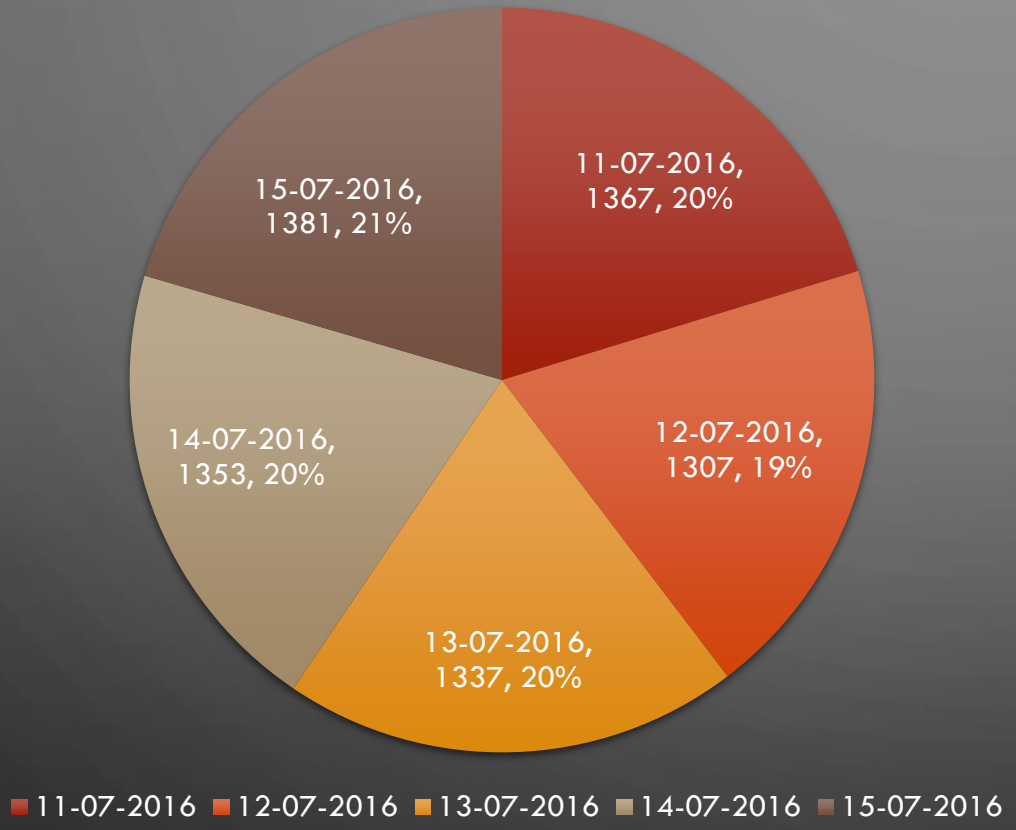
- 3 hours of the day allocated to each timeslot such as, 1 is the slot for 00:00 to 03:00
- **Least number of requests** came on the timeslots **1 and 5** i.e. **1% and 7%** respectively.
- Compare to other time slots **requests are pretty high** at slots 3 and 7 i.e. **06:00 to 09:00** and **18:00 to 21:00** . Which shows that demand is very high **during office reaching hours(18%) and returning from office hours(22%).**

TOTAL REQUESTS MADE IN A PARTICULAR TIMESLOT VS STATUS



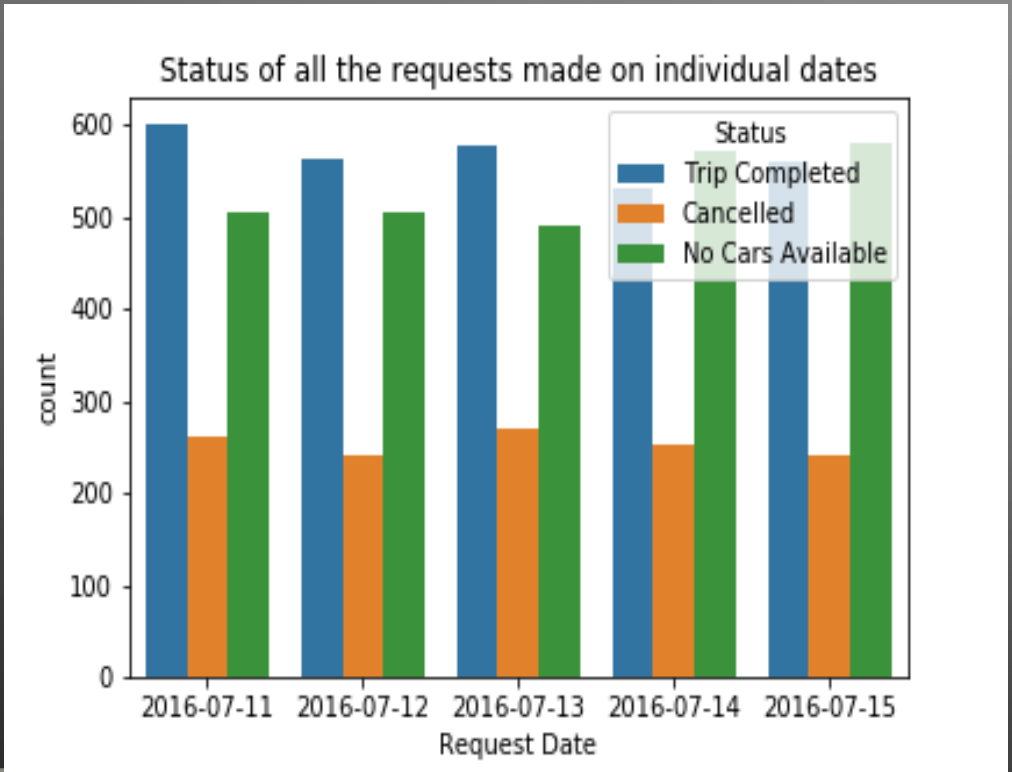
- As mentioned before, slots 3 and 7 shows the highest hike in the requests made.
- Trips with complete status are almost equivalent in both the half of the days.
- Cancelled trips are higher at the morning time or the day time (timeslots 2, 3 and 4)
- Non-availability of the cars are much higher at the night time (timeslots 6, 7 and 8) as compared to the morning.

TOTAL REQUESTS MADE ON EACH OF THE 5 CONSECUTIVE DAYS



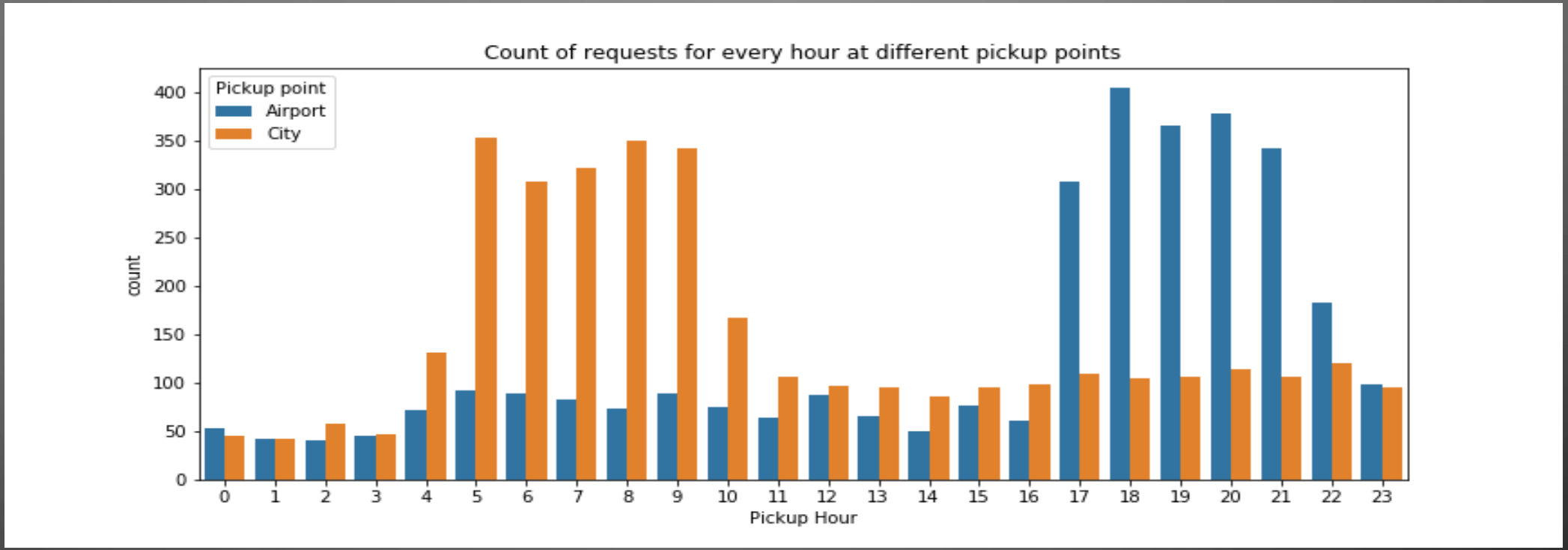
- Univariate analysis on the five consecutive dates of the month July, extracted from the Request timestamp of the data.
- We can see that the requests made everyday are almost same in numbers.
- All the five days are contributing **more than 1300 requests** to the dataset making up to approximately **20% of the requests each day.**

STATUS OF ALL REQUESTS MADE ON ALL FIVE CONSECUTIVE DATES



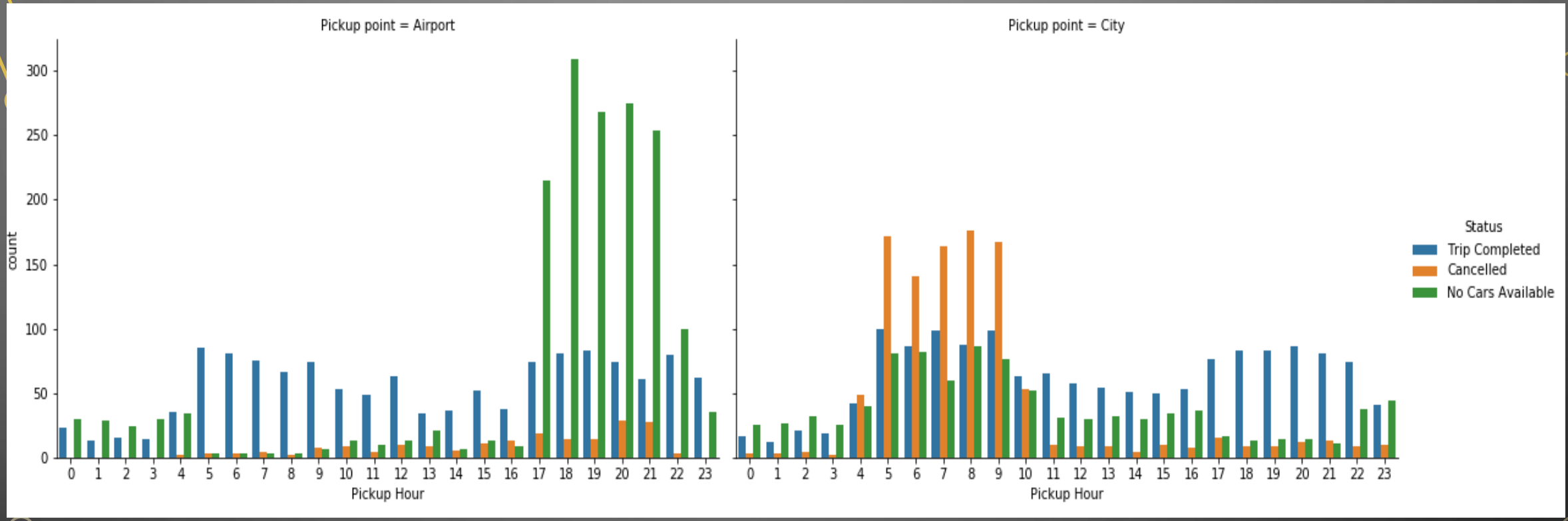
- Segmented univariate analysis on the Requests date column.
- Number of Trips completed are almost equal for everyday.
- The cancellation of the trips follows the same trend for everyday.
- The non availability of cars showed a bit of growth in last two days but the change is not that big.

TOTAL REQUESTS MADE FOR EVERY HOUR ON BOTH PICKUP POINTS



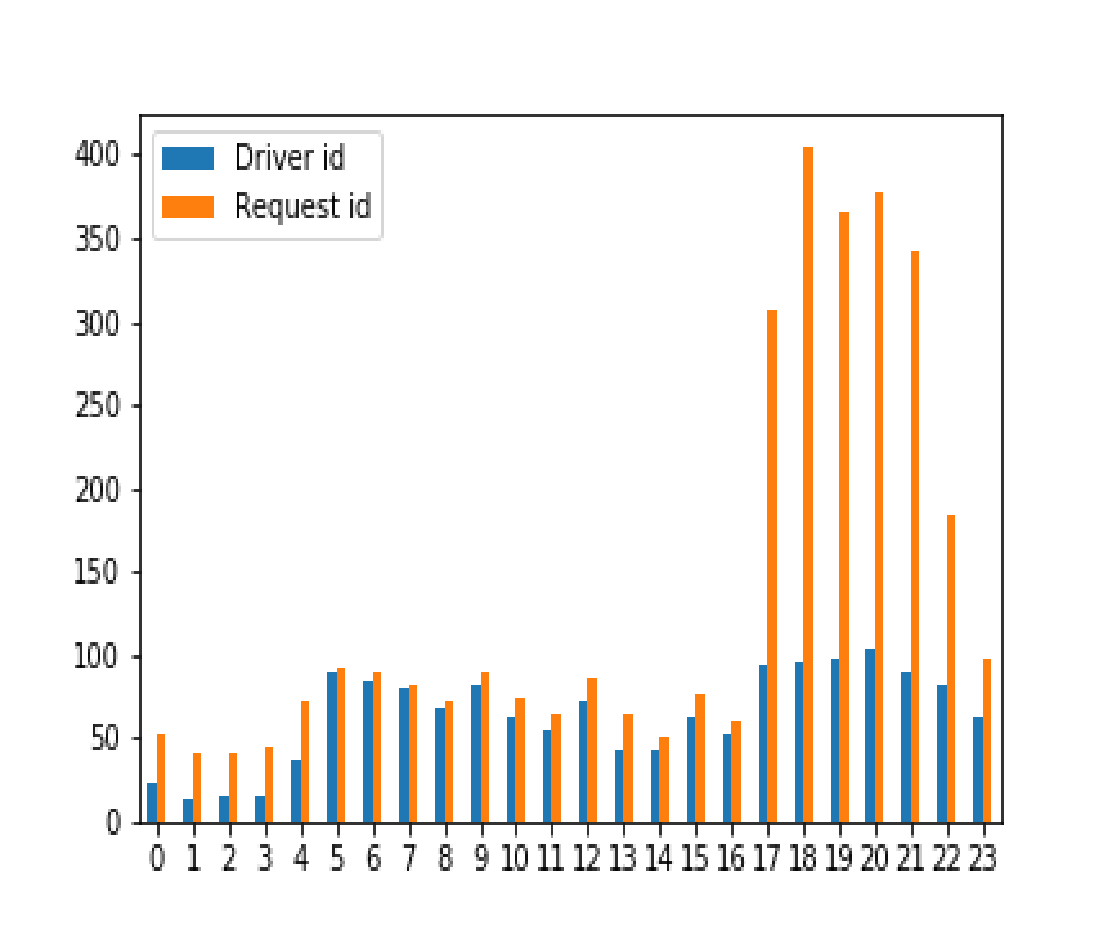
- The hike we saw in the total requests made in the morning were most of them coming from the City, which means that people are travelling from City to Airport mostly in the morning.
- Just like the previous insight, the hike in the night time was because most of the people at Airport were requesting for a cab.
- If we make this curve for both pickup points separately, they both will follow normalization curve.

STATUS OF ALL THE REQUESTS MADE EACH HOUR AT BOTH PICKUP LOCATIONS



- As the previous plot, this also follows normalization curve and we can see the hike in both plots.
- First plot shows that the request rate is very high from 17:00 to 22:00 in the night time at Airport, but most of the trips can't be completed because no cars were available.
- Second plot shows that morning requests are mostly from city and specially between 05:00 to 09:00, but trips can't be completed due to high cancellation rate.

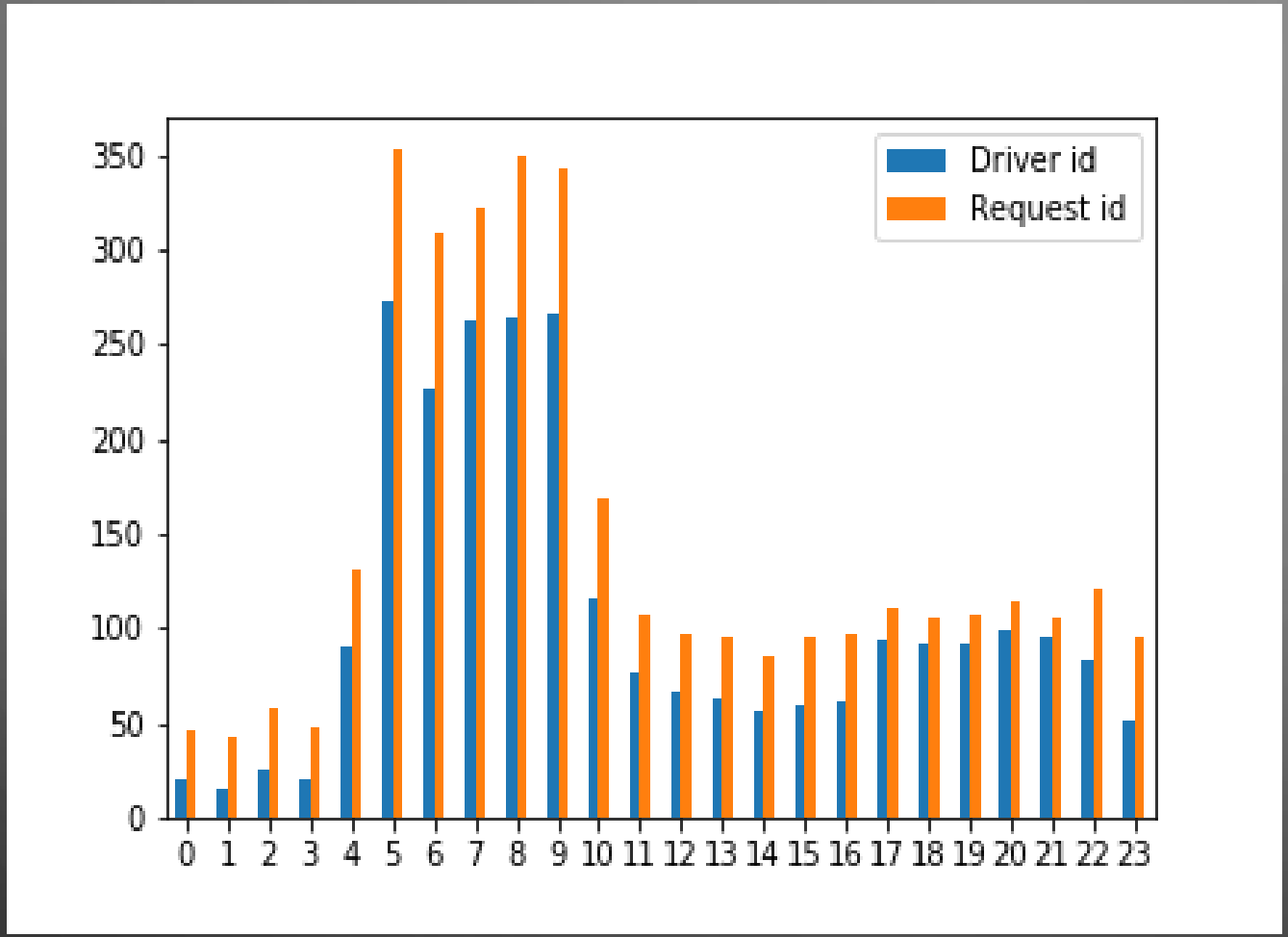
SUPPLY DEMAND GAP FOR EACH HOUR ON PICKUP POINTS



Supply and Demand gap plot of Airport.

Analysis

- Follows the normalization curve and the **hike** is between 17:00 to 22:00.
- The **demand** for the cabs are **very high at night** (an average of **approx. 350** every hour)
- The cabs or drivers available at the busy time are very low comparatively due to No cars Available. (an average of **100** approx.)



Supply and Demand gap plot of City.

Analysis

- Also follows the normalization curve and the **hike** can be seen between **05:00 to 10:00**.
- The **demand** for cabs **are very high in the morning**. (an average of approx. **300**)
- The cabs drivers available in the **morning time** are much **lower** and the supply demand gap is more due to high cancellation by drivers.

CONCLUSIONS MADE AFTER THE ANALYSIS

- Problematic requests are almost equal at both the pickup locations.
- Everyday, there are equal amount of requests made between the two pickup points.
- Morning and the late evening time are when we get to see most of the traffic and other slots get very less traffic.
- Majority of the requests made at night are from Airport and maximum of the times, there are ‘No cars Available’.
- Most of the requests made in the morning time are from City and majority of them gets either ‘Cancelled’ by the drivers. City also have a good amount of Non-availability rate.

POSSIBLE REASONS FOR THE SUPPLY DEMAND GAP

FROM AIRPORT TO CITY

- There is a huge supply demand gap at night due to the high rate of non availability of the cabs at the location.
- The cabs starting their journey from city in the afternoon is very less, so the available cabs at Airport in the night time are very less, as compared to the amount of requests.

FROM CITY TO AIRPORT

- This route has higher cancelation rate by the drivers and non availability of cabs in the daytime.
- The cabs starting their journey from Airport in the midnight is very less, so the available cabs at City in the early morning time are very less and due to early morning most of the drivers cancels the trip.

RECOMMENDATION FOR THE PROBLEM

From City to Airport (05:00-09:00)

- Fare should be dynamic i.e. as request increases, fare should also increase.
- More cabs should be made available at the location
- Add a pre-booking/extra charges option for early morning cabs, making it easier for both the parties to manage the booking and balance the supply-demand.
- Cancellation charges should be applicable for both (Driver and customer)

RECOMMENDATION FOR THE PROBLEM

From Airport to City (17:00-21:00)

- Fare should be dynamic i.e. as request increases, fare should also increase.
- More cabs should be made available at the location
- Add a pre-booking/extra charges option for early morning cabs, making it easier for both the parties to manage the booking and balance the supply-demand.

CONCLUSION

It is clear from the above graph that during the time period of 17:00 to 21:00, at Airport there is a large number of non availability. But at the same time, the trips are getting completed from city to airport with a huge peak and vice-versa (05:00 to 09:00) which makes it clear that there among other reasons, one of the reason for this situation is lower availability of cabs.

And to solve this problem, the company needs to provide cabs at that location to have a well managed and profitable business.

THE END ☺