

# Assignment: Uber Supply-Demand Gap

Aim: Analyse why requests are being Cancelled or Cars are not available and investigate Supply-Demand of cabs at Airport and City.

Har Shobhit Dayal

# About The Data

- ▶ There are 6745 rows in data and 6 features/columns available.
- ▶ There are six features associated with each request made by a customer:
  1. **Request id:** A unique identifier of the request.
  2. **Time of request:** The date and time at which the customer made the trip request.
  3. **Drop-off time:** The drop-off date and time, in case the trip was completed .
  4. **Pick-up point:** The point from which the request was made.
  5. **Driver id:** The unique identification number of the driver.
  6. **Status of the request:** The final status of the trip, can be either completed, canceled by the driver or no cars are available.
- ▶ Request Id is unique for each row and there were no duplicate records found.
- ▶ Data Glimpse

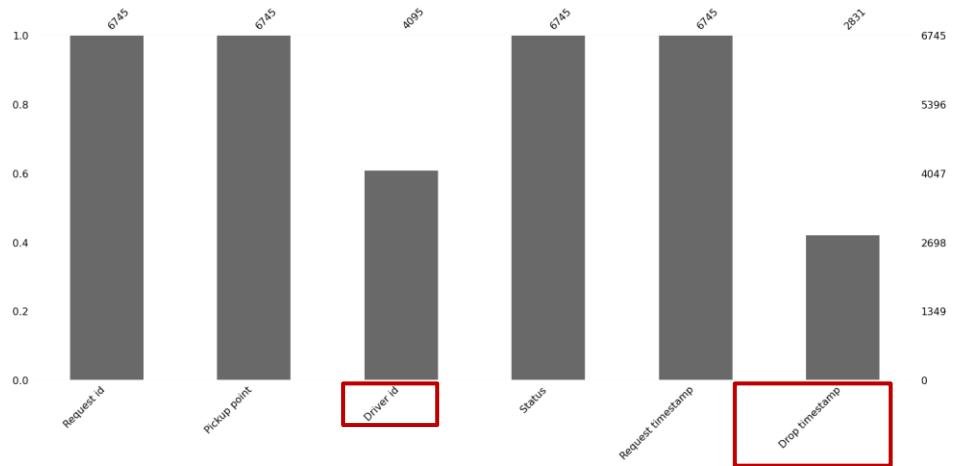
In [8]: `df_1.head()`

Out[8]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47

# Handling Missing Values

- ▶ Around 39% of values are missing for driver id column and around 58% are missing for drop timestamp.
- ▶ As visible from the below graph where around 4095 and 2831 are missing value counts for the respective columns.



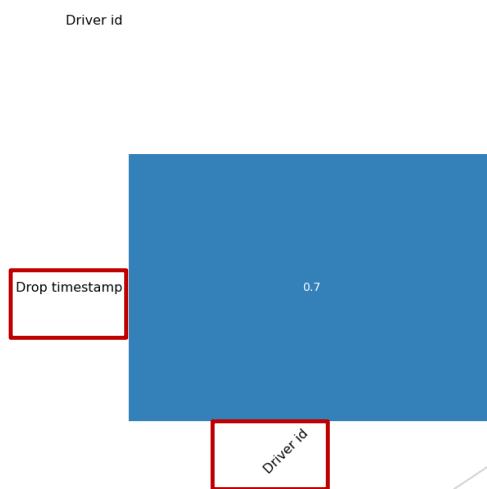
- ▶ The below plot appears blank(white) wherever there are missing values for any column.
- ▶ Driver Id and Drop timestamp have some common rows where the values are missing as depicted by the white region.
- ▶ It suggests that there might be some impact of a column which led to both Driver Id and Drop timestamp having missing values.



- ▶ From the correlation matrix on right we can see that there is a strong correlation(0.7) among the missing values for Drop time stamp and Driver Id.
  - ▶ This suggests that missing value type is of MAR(Missing at Random) and need to be handled.
- ▶ On further analysis it was found driver id is missing where Status column has values as 'No Cars Available' which is expected as no trip would be assigned in such case.
- ▶ For drop time stamp it was found that it is missing when Status column has values as 'Cancelled' or 'No cars available' which is expected as the trip never took place.

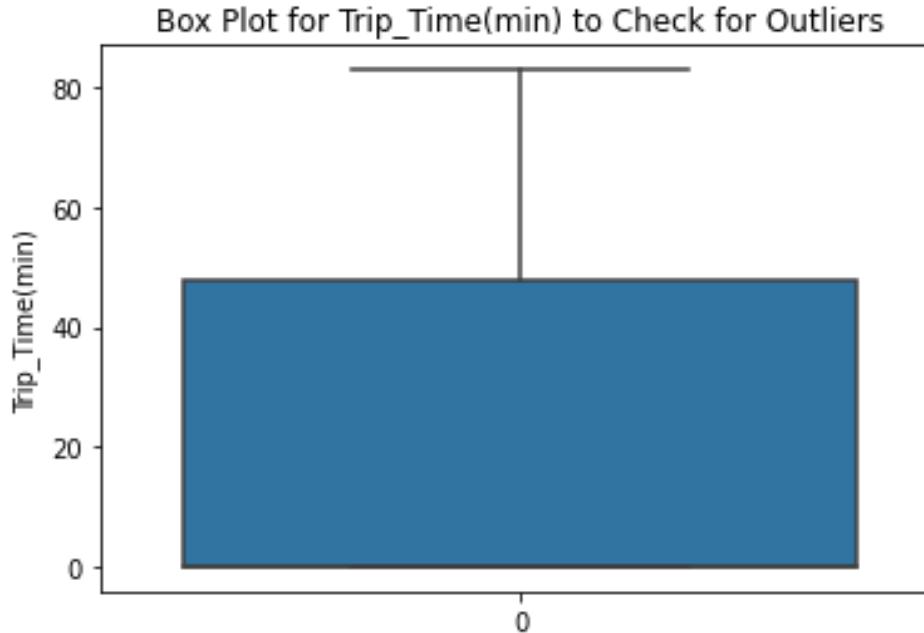
## Actions taken to handle missing values:

- ▶ Driver Id: Unique random driver id's were assigned where it was missing.
- ▶ Drop Timestamp: When this column was converted to datetime format it was assigned a value of Nat to indicate that no time is associated with this data.



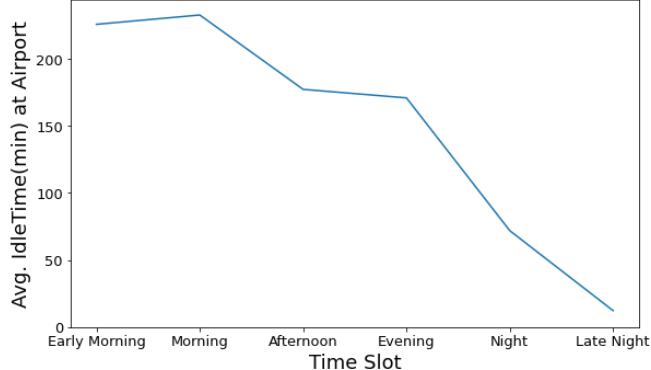
# Handling Outliers

- ▶ Trip time was calculated by taking the difference between the Drop time and Pickup time to see if there are any outliers in data.
  - ▶ If the trip time is more than a day it may indicate that there are outliers as usually the airport and city are not too far by.
  - ▶ It was found that no trip had taken more than a day to complete.
- ▶ There were no data points below ( $Q1 - 1.5 \times IQR$ ) or above ( $Q3 + 1.5 \times IQR$ ) suggesting there are not outliers.
  - ▶  $Q1$ -Quartile 1,  $Q3$ -Quartile 3,  $IQR$ (Interquartile Range =  $Q3 - Q1$ )
- ▶ From the box plot we can see that trip time ranges usually from 0 to 48 min and there are no values after the upper fence of 80 min which suggests there are no outliers as such.
- ▶ Majority of the rides were cancelled or No cars were available as almost 50% values have trip time of 0.

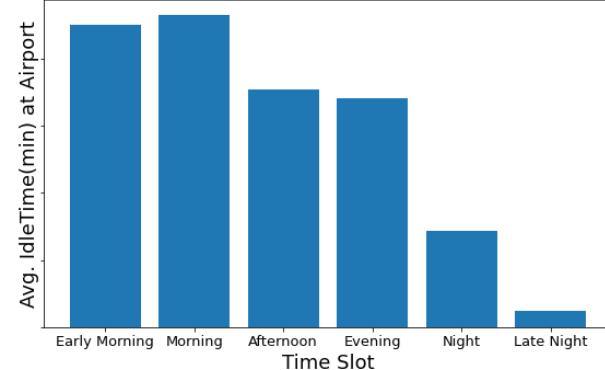


# Idle Time Of Drivers At the Airport

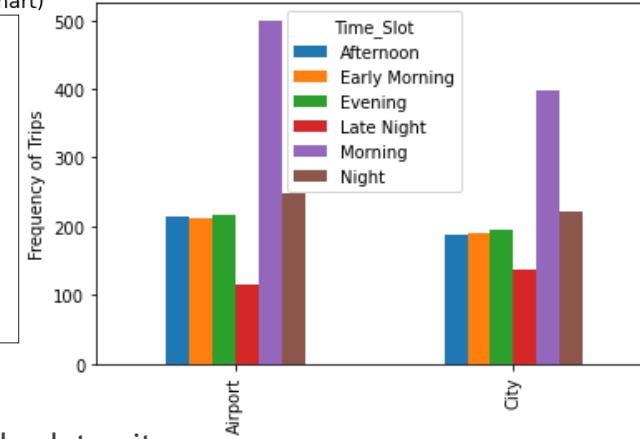
Avg. IdleTime(min) at Airport in Different Time Slots(Line Chart)



Avg. IdleTime(min) at Airport in Different Time Slots(Bar Chart)



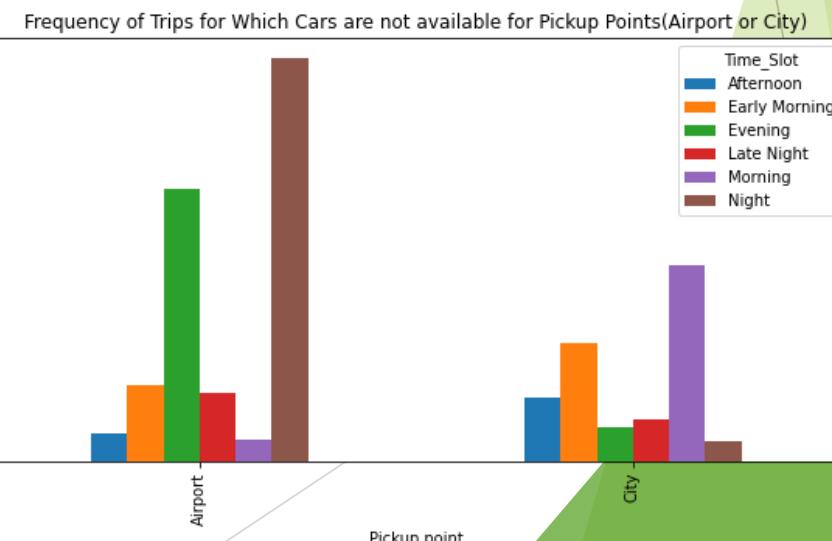
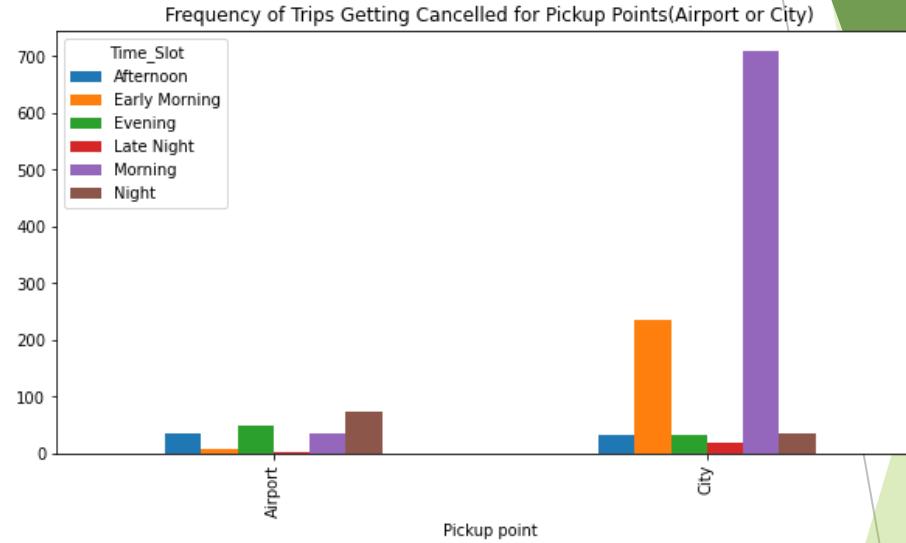
Frequency of Trips for Drop Points(Airport or City)



- ▶ Idle time is the time the driver has to wait at the airport till he gets a ride back to city.
- ▶ Idle time was calculated for each driver for the different time slots i.e. Early Morning(12 am-6am),Morning(6am-12pm),Afternoon(12pm-4pm),Evening(4pm-7pm),Night(7pm-10pm) and late night(10pm-12am).
- ▶ It can be observed from the above graphs that during early morning and morning hour the avg idle time is high.
  - ▶ This might be due to the reason that in morning time most of the flights go from out of city. So, there might be more trips to airport rather than from airport to city.
- ▶ During Afternoon and Evening time slots as frequency of flights are more, we see that idle time is less as compared to morning time and drivers usually get their return journey.
- ▶ For Night and late night the idle time is very low as many drivers do not operate during that time, so the average idle time would be less in that case as there are not many trips to airport at night.
- ▶ From the Frequency of trips graph we can actually see that in morning time slot more trips are there to airport as compared to trips from airport to city which gives to rise of idle time.
  - ▶ This also matches with the conclusion that idle time is high during morning hours.
- ▶ So, there can by one hypothesis based on the findings of idle time is that during the morning hours idle time is much more as compared to rest of the day.

# Analysis on Trips which are Cancelled or Cars Are not Available

- ▶ Most of the trips which are cancelled by the drivers are from city to airport during morning hour as indicated by violet bar above under city label on the graph on right.
  - ▶ This coincides with the observation that idle time is more during early morning/morning hours which can lead to drivers cancelling the requests as they do not get return journey easily from the airport.
- ▶ For Afternoon and evening the trips getting cancelled are less as idle time was also less in this period.
  - ▶ The driver does not have to wait much in afternoon and evening to get next ride and he can earn more which leads to accepting the ride request.
- ▶ For trips from airport to city night time is usually when most of the rides are cancelled. Night time is usually after 9 or 10 pm which indicates the end of day for many drivers, so it can happen that they cancel these rides as it end of day for them.
- ▶ For pickup point airport a large proportion of cars are not available during the night time as observed from the graph on the right.
  - ▶ As during night time very few drivers operate this can lead to less number of cars available.
- ▶ For pickup point city a large proportion of cars are not available during the morning time .



# Supply and Demand Analysis

► From the line chart on right we can see that when destination is Airport, early morning/morning time slots have the highest number of trip requests.

- This can be due to various flights going out of the city in morning, so customers tend to go to airport during early morning/morning.

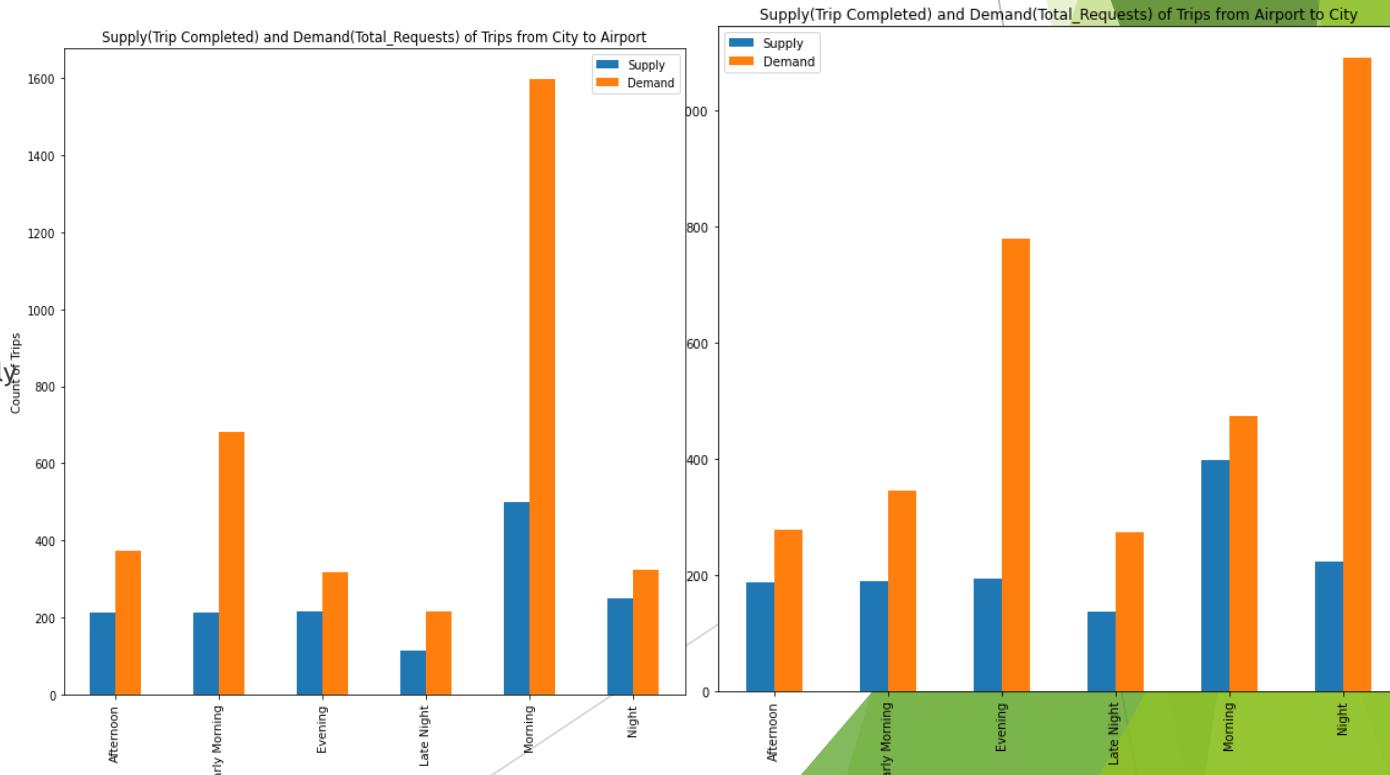
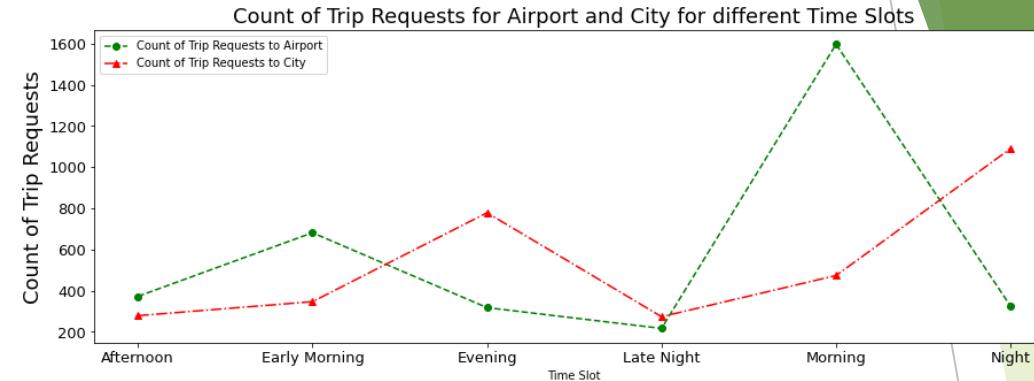
► At the same time where destination is city we can see that in early morning and morning time slots the trip requests count are not that much high.

- This can help in understanding why the idle time is more at Airport during the morning hour as there are less requests for trips from Airport to city.

► Trip requests are high for destination city during the night time as international flights operate during these hours which leads to more people going to city.

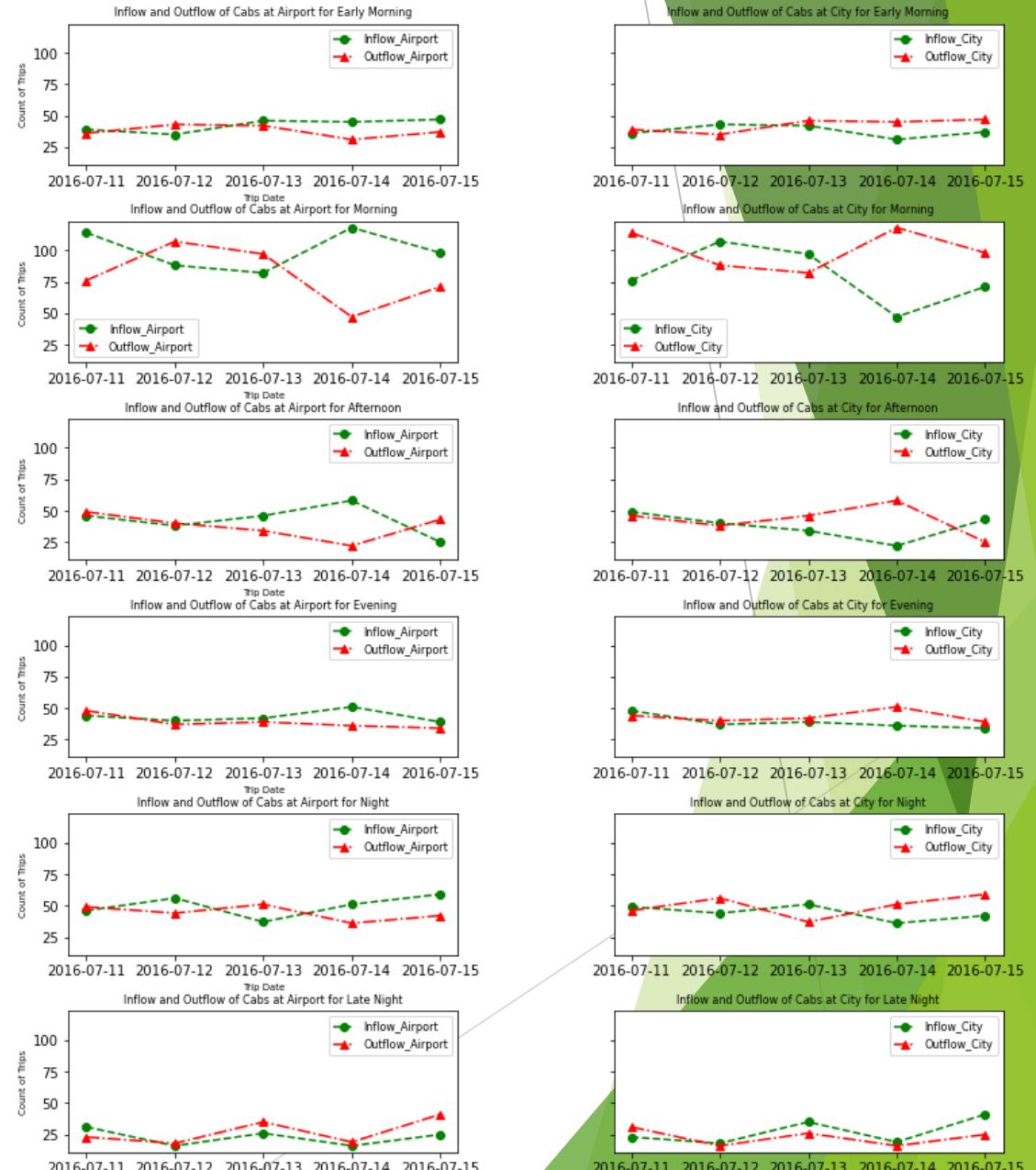
► From the bar graphs on right we can observe that:

- At **Airport night, evening and late night time** i.e. 4 pm - 12 am time slot supply is less but demand is more.
- At **City early morning and morning** i.e. around 3 am-12 pm time slots supply is less but demand is more.



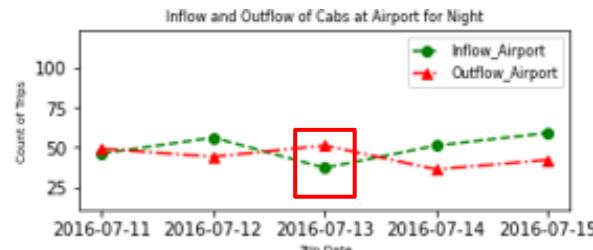
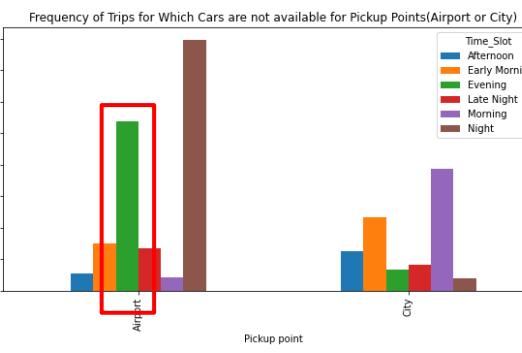
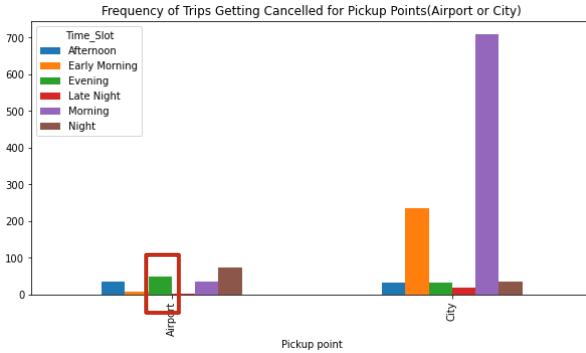
# Inflow and Outflow of Cars

- ▶ We can see from the graphs on right that if inflow is high at airport for a particular time slot then outflow is high from the city region which is the expected behavior.
- ▶ For early morning and morning time slots inflow of cabs to airport is more as compared to outflow from the Airport.
- ▶ For early morning and morning time slots inflow of cabs to city is less as compared to outflow from the city.
- ▶ For Afternoon time slot inflow of cabs to airport is more for some of the dates as compared to outflow from the Airport.
  - ▶ This behavior can be due to heavy traffic of flights going out from the city leading to customers booking cab trips to Airport which we earlier saw that total requests for afternoon is high for pickup point City.
- ▶ For Afternoon time slot inflow of cabs to city is less as compared to outflow from the city as the demand from City to airport is not high during this time slot.
- ▶ For Evening, night and late night time slots inflow of cabs to airport is less for some of the dates as compared to outflow from the Airport.
- ▶ For Evening, night and late night time slots inflow of cabs to city is more for some of the dates as compared to outflow from the Airport due to arrival of international flights at the Airport during this time slot.



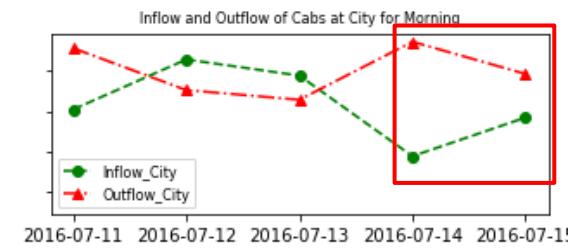
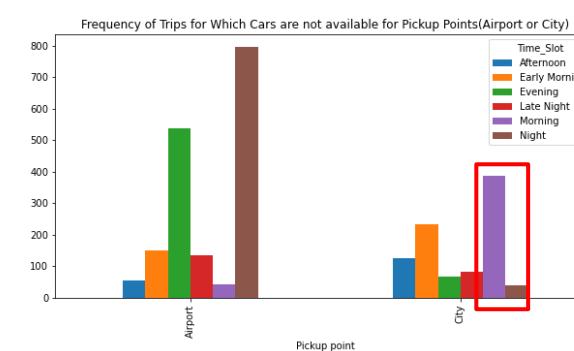
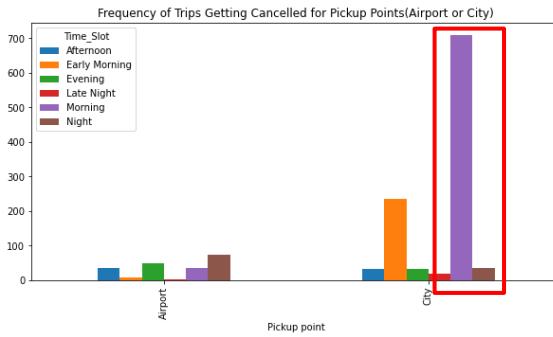
# Why there is a Gap b/w Supply and Demand

- At Airport night, evening and late night time i.e. around 4 pm - 12 am time slot supply is less but demand is more.



- This can be due to the reason that inflow of cabs to airport is less during this time frame which leads to the status of 'Cars not available' leading to less of supply.
- The idle time in evening and night time is also significant at Airport, which can lead to drivers not accepting the rides as they do not get a return journey easily. This can lead to trips getting 'Cancelled' which increases the gap between supply and demand.

- At city night, early morning and morning i.e. around 3 am - 12 pm time slot supply is less but demand is more.



- This can be due to the reason that inflow of cabs at city is less during early morning and morning which leads to the status of 'Cars not available' leading to less of supply.
- Most of the trips get Cancelled during morning time where pickup point is City, as indicated by the above bar graph which leads to a gap in supply and demand.

# Ways to Solve Supply-Demand Issue



- ▶ Actions that can be taken to resolve the supply and demand issue:
- ▶ As inflow of cabs is more at Airport than required during early morning and morning time, we can instruct some drivers to go to city around 3 am-12 pm so that inflow of cabs at the city increases and the supply-demand gap is reduced.
- ▶ As inflow of cabs is more at city than required during evening, night and late night time we can instruct some drivers to go to airport between 4pm -12 am so that inflow of cabs at the airport increases and the supply-demand gap is reduced.
- ▶ Provide some incentives for drivers for airport trips during peak time.
- ▶ Impose a penalty for cancellation of requests by the drivers. A threshold to be set for maximum cancellations that can be done in a day.
- ▶ Flight patterns can be analyzed and same can be communicated to the drivers so that idle time for the drivers is reduced and less number of trips get cancelled.
  - ▶ For e.g. if at morning between 9-10 am there are lot of incoming flights so there might be large number of cab requests.
  - ▶ It can be communicated to drivers to reach airport around 8:45 am as flight traffic is to be expected and it would easy to get a ride.
  - ▶ This would reduce idle time and also help in supply-demand problem.