Uber Supply-Demand Gap

Analysis is to identify the root cause of the problem, i.e. cancellation or non-availability of cars.

Business Objective for Analysis

➥ To increase the customer experience with the company by reducing the GAP between demand and supply of cabs

➥ This GAP can be generated by two ways :

- Cancellation done by a driver

- Non-availability of cars

Data Scientist Objective

➥ Receive the data and to identify the root cause of the problem (i.e. cancellation and non-availability of cars) and generate possible hypothesis and recommend corresponding ways to improve the situation

Data Understanding

➥ The dataset provided in form **.csv** file.

➥ It has six attributes associated with each request made by a customer:

- **Request id** : A unique identifier of the request
- **Time of request** : The date, time when 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

None: This analysis is limited to the trips to and from the airport and city

Received Data

Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
619	Airport	1	Trip Completed	11-07-16 11:51	11-07-16 13:00
867	Airport	1	Trip Completed	11-07-16 17:57	11-07-16 18:47
1807	City	1	Trip Completed	12-07-16 9:17	12-07-16 9:58
2532	Airport	1	Trip Completed	12-07-16 21:08	12-07-16 22:03
3112	City	1	Trip Completed	13-07-16 8:33	13-07-16 9:25
3879	Airport	1	Trip Completed	13-07-16 21:57	13-07-16 22:28
4270	Airport	1	Trip Completed	14-07-16 6:15	14-07-16 7:13
5510	Airport	1	Trip Completed	15-07-16 5:11	15-07-16 6:07
6248	City	1	Trip Completed	15-07-16 17:57	15-07-16 18:50
267	City	2	Trip Completed	11-07-16 6:46	11-07-16 7:25
1467	Airport	2	Trip Completed	12-07-16 5:08	12-07-16 6:02
1983	City	2	Trip Completed	12-07-16 12:30	12-07-16 12:57

Data Cleaning and Preparation

➥ Import required libraries and read in data received in python

➥ Change the data type of attributes into required one:

- Request timestamp (from string to date time)

- Drop timestamp (from string to date time)

➥ Checking of any duplicated rows present in dataset

- No duplicated rows found

➥ Derived new attributes based on need of the analysis:

- Request date : Date when the request has been made by customer

- Request time : Time when the request has been made by customer

- Time_slot   : Representing Day parting on the basis of hour of the day

(Note: Time slot division is shown in next slide….)

Time Slot Division

Time Slot	Hour of the day
Over Night	00:00 to 04:59
Early Morning	05:00 to 06:59
Morning	07:00 to 08:59
Late Morning	09:00 to 11:59
Afternoon	12:00 to 15:59
Early Evening	16:00 to 17:59
Late Evening	18:00 to 19:59
Night	20:00 to 23:59

Data Cleaning and Preparation Continuum

➡ Checking of NA values in dataset except where it can't defined (For e.g. Driver Id can't be defined for status of request No Cars Available)

➡ Formed Data Frame :

S.NO.	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request Date	Request time	Time_slot
0	619	Airport	1	Trip Completed	07-11-16 11:51	07-11-16 13:00	07-11-16	11:51:00	Late Morning
1	867	Airport	1	Trip Completed	07-11-16 17:57	07-11-16 18:47	07-11-16	17:57:00	Early Evening
2	1807	City	1	Trip Completed	07-12-16 9:17	07-12-16 9:58	07-12-16	9:17:00	Late Morning
3	2532	Airport	1	Trip Completed	07-12-16 21:08	07-12-16 22:03	07-12-16	21:08:00	Night
4	3112	City	1	Trip Completed	13-07-16 8:33	13-07-16 9:25	13-07-16	8:33:16	Morning
5	3879	Airport	1	Trip Completed	13-07-16 21:57	13-07-16 22:28	13-07-16	21:57:28	Night
6	4270	Airport	1	Trip Completed	14-07-16 6:15	14-07-16 7:13	14-07-16	6:15:32	Early Morning
7	5510	Airport	1	Trip Completed	15-07-16 5:11	15-07-16 6:07	15-07-16	5:11:52	Early Morning
8	6248	City	1	Trip Completed	15-07-16 17:57	15-07-16 18:50	15-07-16	17:57:27	Early Evening

# Plot representing most occurred type of status

➽ Data Preparation:

- Defining a new table representing the number of type status occurred

Formed new Table:

Status	Count
Trip Completed	2831
No Cars Available	2650
Cancelled	1264

Note: As the given time line is for only four days, frequency of any type of request not much change, so to understand how the different status is compared, total count of that status is plotted with each other.
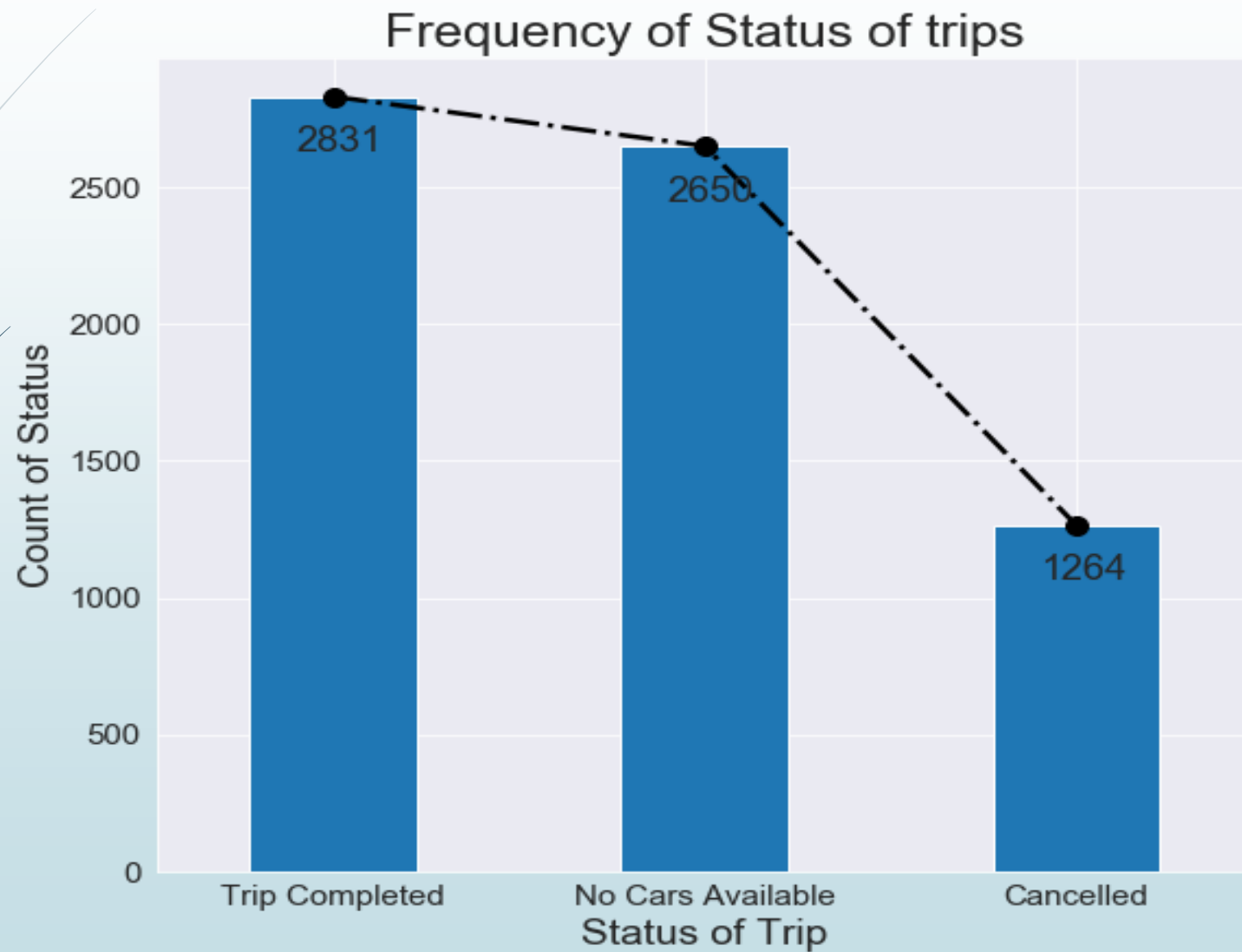
Plot

Insights from plot

➥ Plot Clearly shows that

the '**No Cars Available**' type of status is almost double than problem 'Cancellation' facing by the customers.

Most pressing problem to concerned about is 'No Cars Available'

Plot a Graph for understanding the most problematic type of request (city to airport/airport to city) and also the time slots in each type of request

➡ Data Preparation:

- Formed a new a data frame excluded status type 'Trip Completed'
- Modify of the table to represent the total count of category with index 'pickup point' and 'Time Slot'
- Adding the new attribute defining total count of incomplete trip in that category

➡ Formed Table:

Pickup point	Time_slot	Cancelled	No Cars Available	Total
Airport	Over Night	2	148	150
Airport	Early Morning	8	7	15
Airport	Morning	7	7	14
Airport	Late Morning	22	30	52
Airport	Afternoon	36	55	91
Airport	Early Evening	33	224	257
Airport	Late Evening	30	577	607
Airport	Night	60	665	725

Note : As Total incomplete trips either from airport (1911) or to airport (2003) are approximately equal. Therefore, A plot is formed in these two type of requests with respect to the Time slot

Plot : For Airport to City Request

Frequency of status of request w.r.t Time slot
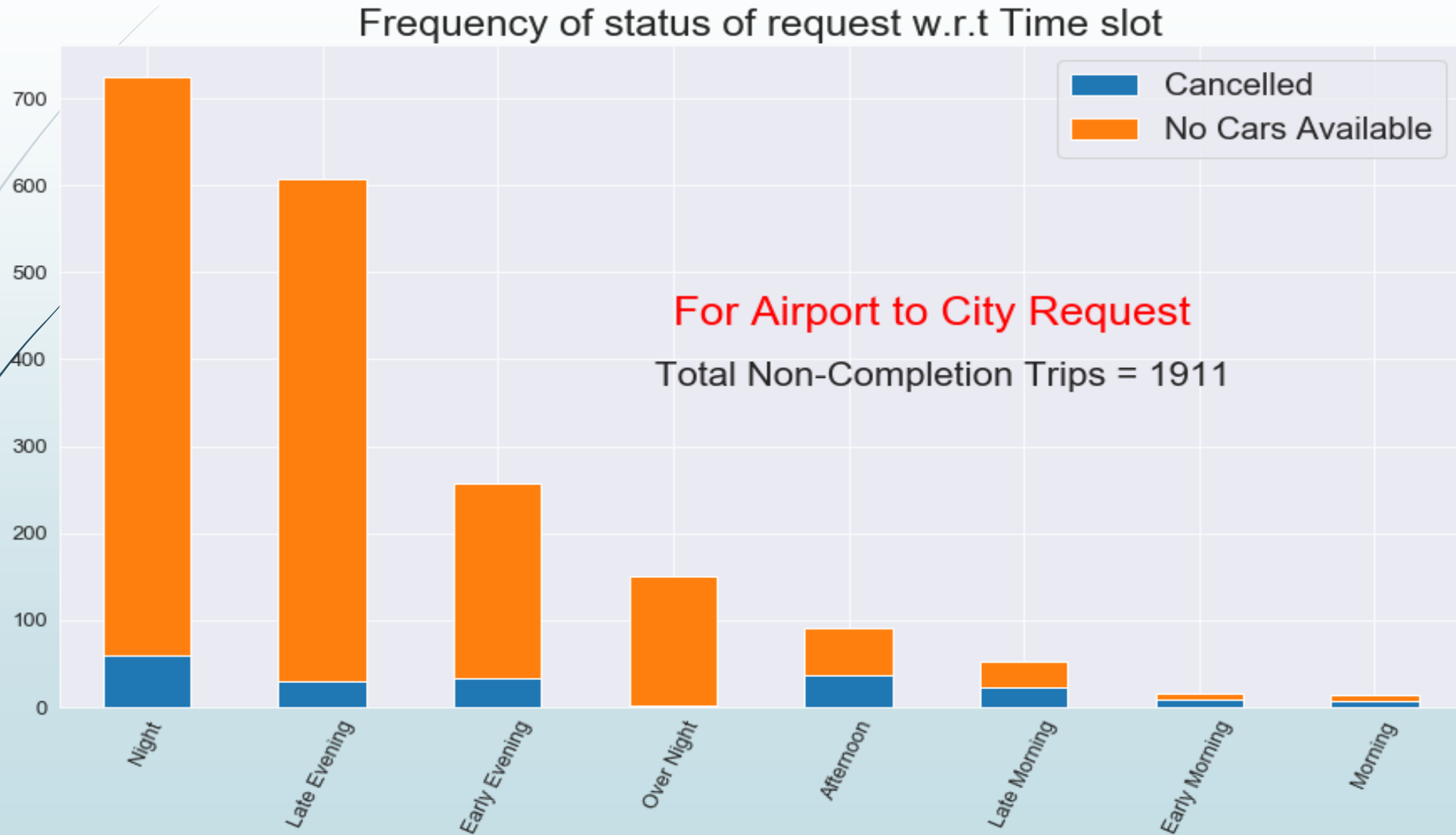
Plot : For City to Airport Request
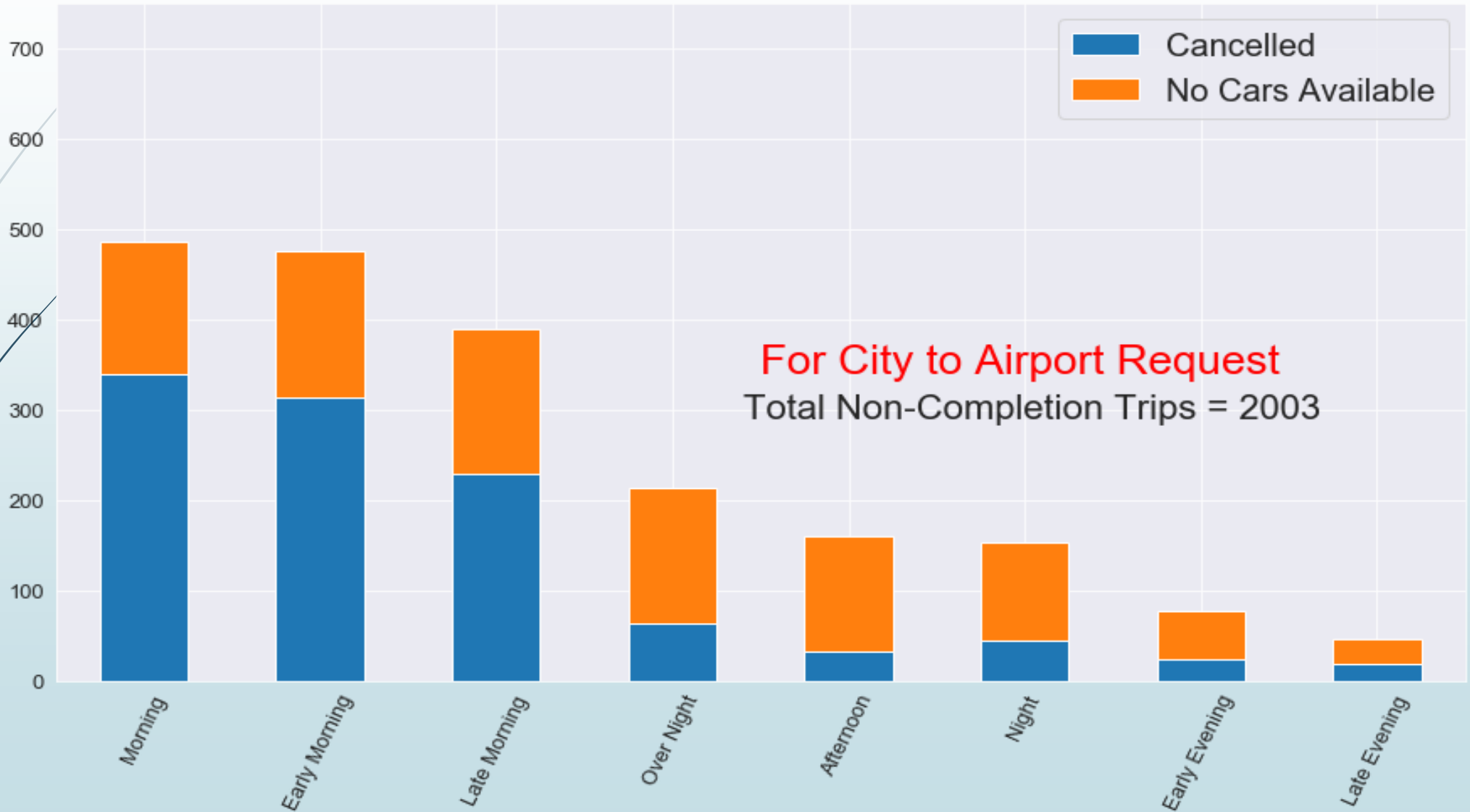
Insights From Plot

➥ Total Incompletion trips for:

- Airport to city : 1911
- City to Airport : 2003

➥ Total incompletion trips are almost **same** for both the type of requests

➥ For request **Airport to City**:

- **'Night'**, **'Late Evenings'** and **'Early Evening'** (time from 4 pm to 12 am) are **peak time** slots where the problem is **maximum**
- And in these identified time slots '**No Cars Available**' is emerging to be the main pressing problem where focused is needed

➥ For request **City to Airport** :

- Here **'Morning'**, **'Early Morning'** and **'Late Morning'** (time from 5 am to 12 pm) are **peak time** slots where the problem is **maximum**
- Here Both '**No Cars Available**' & '**Cancellation**' have being equally reasonable for the problem with '**Cancellation**' contribute a little higher percentage

Collective Insights

For both type of request 'No Cars Available' is sharing the major contribution (Also in count of total trip incompletion).

➥ Most Pressing Problem : '**No Cars Available**'

➥ As the total no. of incompletion request is same for both type of request, so both type of request need focused independently as the time slots when the either type of request facing problem is different.

Note: Even so, these time slots are different for both request, they are relatable as Airport to city major facing problem in night time and City to Airport facing problem in morning time. Now if we settle down the problem of night time of Airport to city, then Supply for City to Airport at morning time also increases and hence both can tackle at same time

# Gap between supply and demand

Data Preparation

➦ Create a Data frame representing the **Supply** needed in different time slots

➦ Create a Data frame representing the **Demand** needed in different time slots

➦ Merge this two data frame and create a new attribute name 'GAP' defining the **GAP** between Demand and Supply in different time slots

➦ Descend the whole data frame on the basis of GAP value

➦ Formed Table:

Time_slot	Demand	Supply	GAP
Night	1439	665	774
Late Evening	983	378	605
Over Night	578	279	299
Early Evening	577	299	278
Late Morning	845	656	189

Plot: GAP between Demand and Supply in different time slots
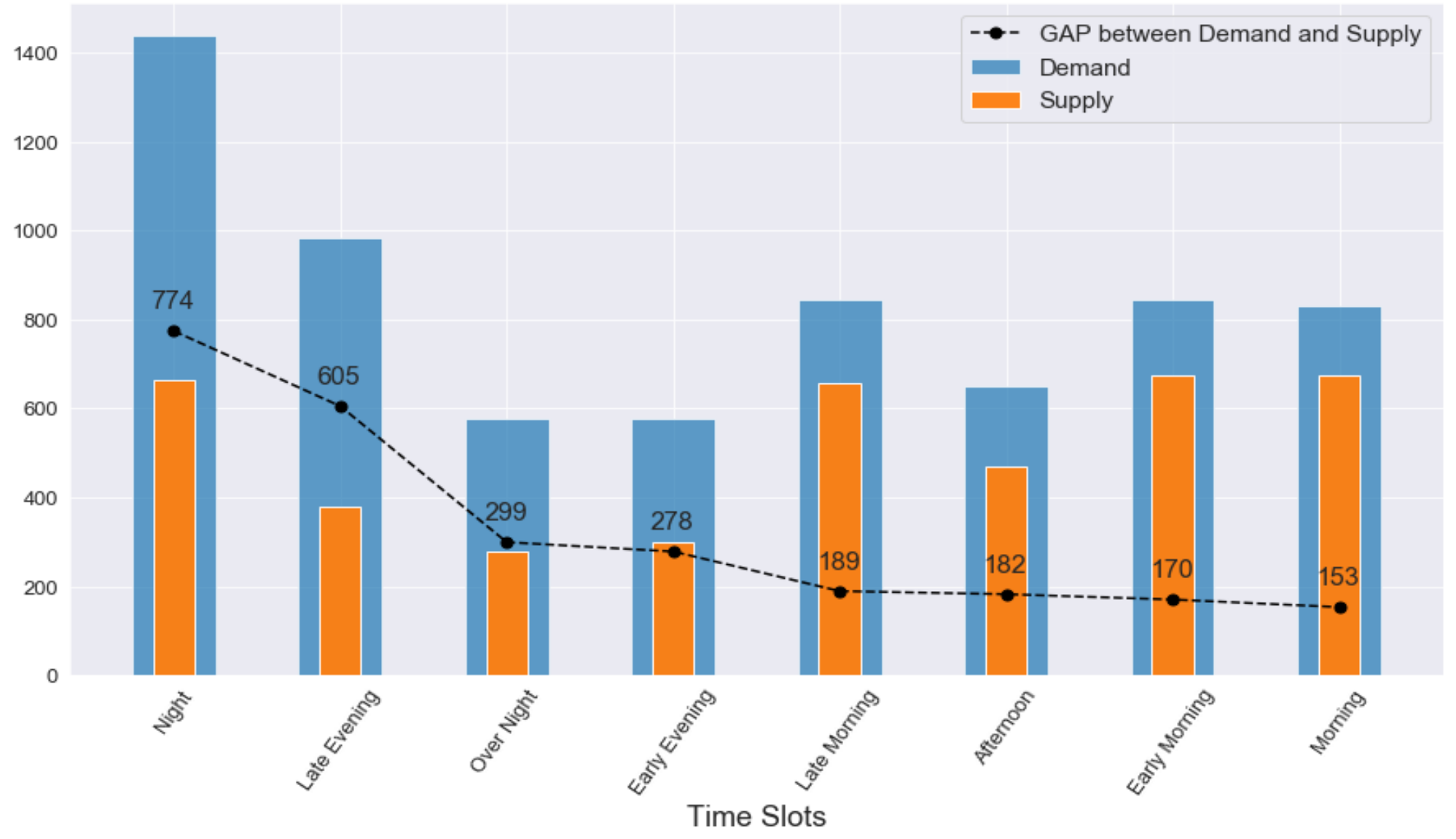
Plots From Insight

- Plot shows the Demand and Supply in different time slots of a day with highlighted the GAP between them

- GAP between Demand and Supply is maximum during **'Night'** & **'Late Evening'** (From 6 pm to 12 am)

- On the average :

    Supply = 62% of the Demand (in all time slots)

Note: For the time slots after afternoon don't only have GAP maximum, also the Demand at that time is also very high.

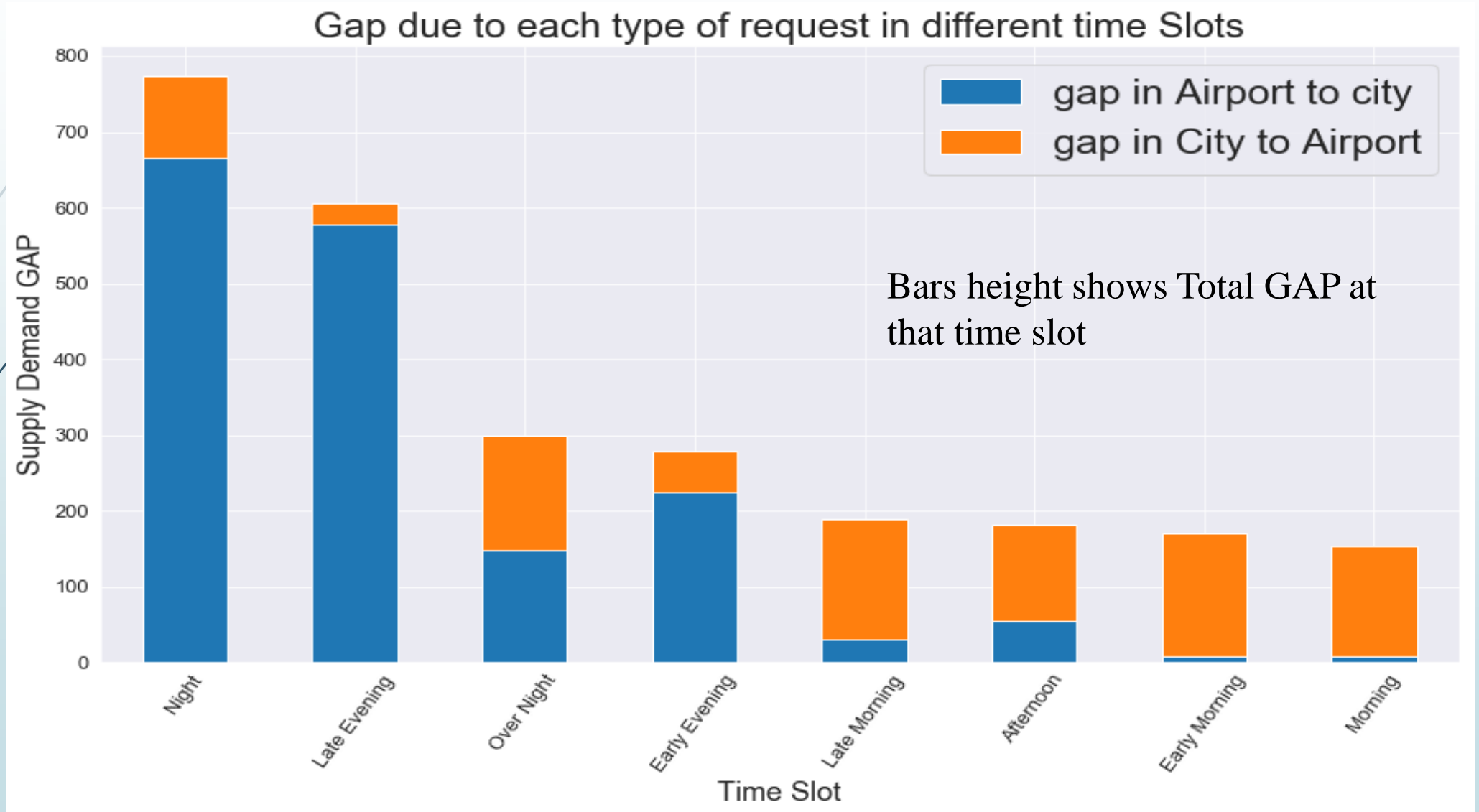# To identify the most severe type of request for which GAP is maximum

➡ Data Preparation:

- Create Data Frame Demand and Supply representing Demand and Supply of the cars from and to the airport in different time slot

- Merge these two data frame in such a way that new data frame representing the GAP in Demand and Supply in cars needed to and from the airport

➡ Formed Table:

Time_slot	gap in Airport to city	gap in City to Airport	GAP
Night	665	109	774
Late Evening	577	28	605
Over Night	148	151	299
Early Evening	224	54	278
Late Morning	30	159	189

Plot

Gap due to each type of request in different time Slots

Bars height shows Total GAP at that time slot

Insights from plot

➽ The time slots where the GAP between the Demand and Supply is maximum (For e.g. 'Night', 'Late Evening') has a major contribution of the request from **'Airport to City'**

Conclusion : **Airport to City** request type need more focused for after '**Afternoon**' time until '**Morning**'

Note: Whereas time slots having less GAP (For e.g., Early Morning, Morning) is majorly due to the request type of **'City to Airport'**

To identify the Driver who Cancelled the trip most

➥ Data preparation:

- Count of Cancellation trips for each Driver individually and arrange them in descending order

➥ Insight from formed data:

- Max. No. of Cancellation by any driver : 12

- Min. No. of Cancellation by any driver :

- Average No. of Cancellation by drivers :

➥ **List of Driver ID having Cancellation Count More than 8:**

➢ 84, 54, 206, 142, 267, 166, 114, 138, 210, 27

Hypothesis

Major GAP between Demand and Supply is in request '**Airport to city**' after **Afternoon** time slot,

There can be possible reasons:

- Flights incoming are more than outgoing after afternoon time (As we see at these time slots Demand is also very high)

- People coming to airport from city are usually use there own cars which decrease the supply of cars at the airport

- Drivers of cars tend to stay in city rather than to make 1 long trip to airport, wait there for few hours and then get back (As we see Driver are tend to cancel the request from city to airport very often)

- Low No. of cars available (As we have see 'No Cars available' is major reason for any type of request which results in shortage of supply

Recommendations

➥ Launch more cars to meet the 'No Cars Available' type of status

➥ Talk to drivers who cancelled trips more often and try to understand the reason behind it

➥ Increase the increment of drivers for trips through and out to the airport

➥ By taking IN and OUT flights data of the airport, Supply cars directly to airport when the Demand is on its peak (like in night time)

Thank You…