UBER CASE STUDY SUPPLY-DEMAND GAP

Praveen Selvaraj

MSDS

Batch – C5

Data Understanding

In the given UBER data, we have 6 columns:

- Request id
- Pickup point
- Driver id
- Status
- Request timestamp
- Drop timestamp

	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

Timestamps are not in uniform format across both the columns
So, in the first step we are going to format timestamp columns

Data Formatting

- Data After Formatting Timestamps columns

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

Feature Engineering

- After formatting timestamp columns, we used them to derive new metrics:
- Request timestamp- Request Date, Request Time, Request Weekday, Request Hour
- Drop timestamp- Drop Date, Drop Time

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request Date	Request Time	Drop Date	Drop Time	Request Weekday	Request Hour
0	619	Airport	1.0	Trip Completed	2016-07-11 11:51:00	2016-07-11 13:00:00	2016-07-11	11:51:00	2016-07-11	13:00:00	Monday	11
1	867	Airport	1.0	Trip Completed	2016-07-11 17:57:00	2016-07-11 18:47:00	2016-07-11	17:57:00	2016-07-11	18:47:00	Monday	17
2	1807	City	1.0	Trip Completed	2016-07-12 09:17:00	2016-07-12 09:58:00	2016-07-12	09:17:00	2016-07-12	09:58:00	Tuesday	9
3	2532	Airport	1.0	Trip Completed	2016-07-12 21:08:00	2016-07-12 22:03:00	2016-07-12	21:08:00	2016-07-12	22:03:00	Tuesday	21
4	3112	City	1.0	Trip Completed	2016-07-13 08:33:16	2016-07-13 09:25:47	2016-07-13	08:33:16	2016-07-13	09:25:47	Wednesday	8

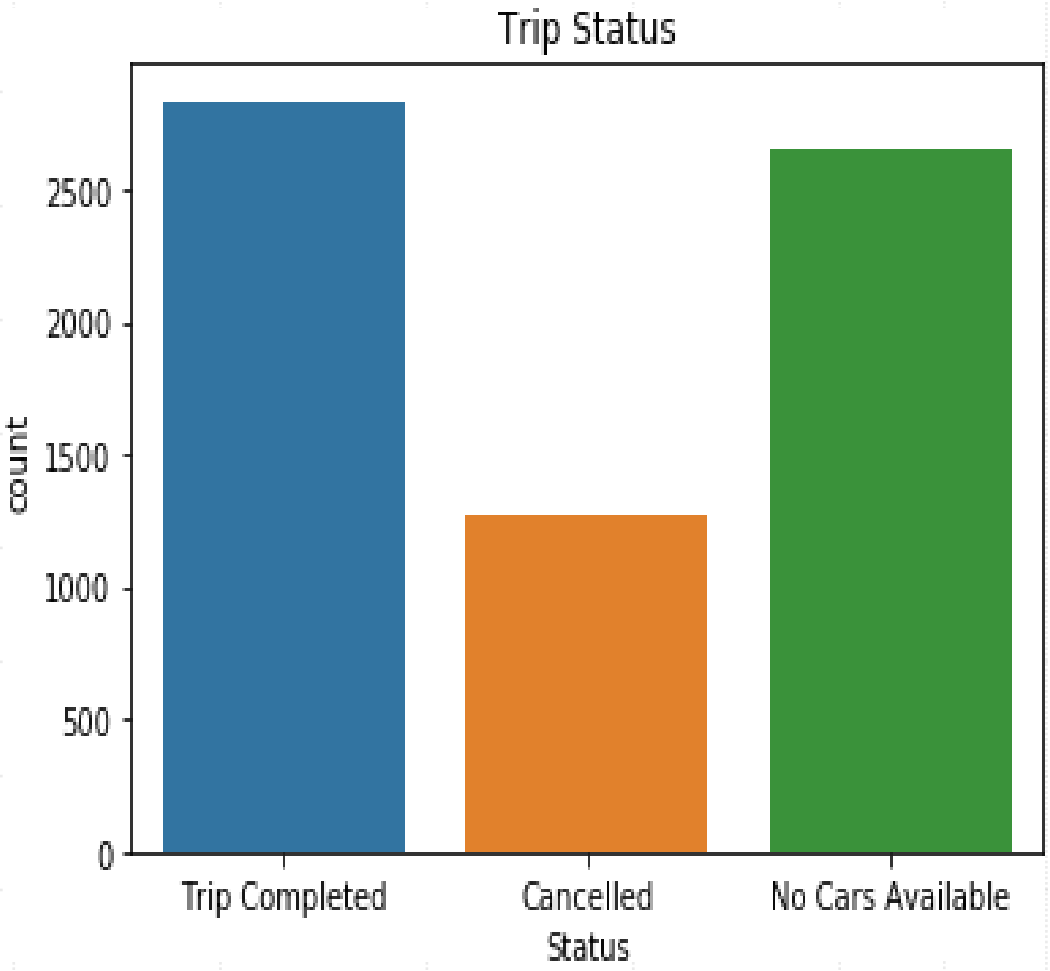
Data Analysis- Trip Status

As we can see from the graph:

Trip Completed	Cancelled Status	No Cars Available
2831	1264	2650

From the above table, we can observe:
Total Demand= 6745
Total Supply= 2831
Supply-Demand Gap= 3914

This shows only **42%** of total demand was met and there is a gap of **58%** of supply either due to trip cancellation or no cabs availability.



Data Analysis- Driver Count

Now lets have a look on the total number of drivers and trip request per day:

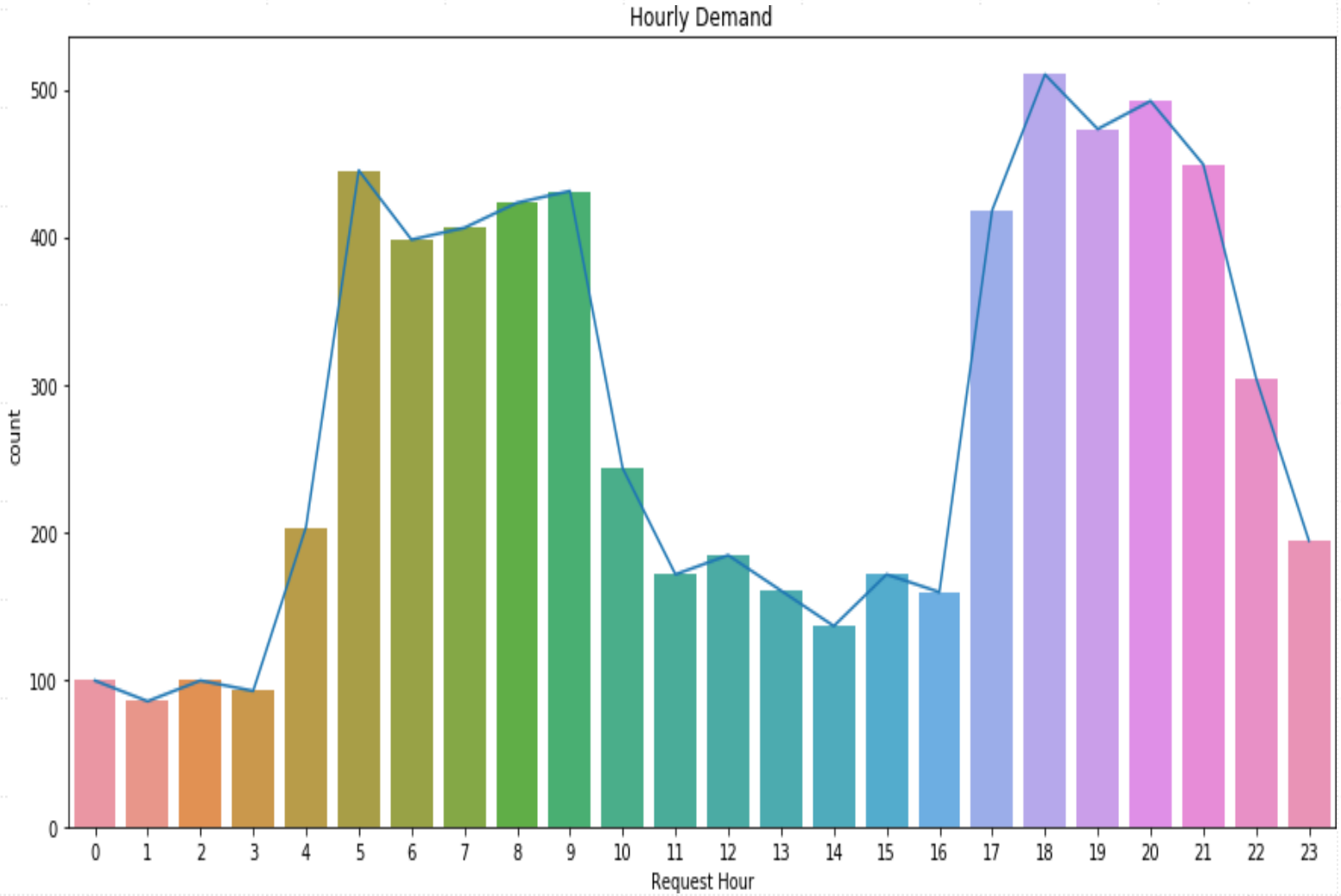
Number of Drivers	Average Trip Request Per Day	Average Trips Completed Per Day
300	1349	566.2

So, on average number of trip request per driver is approx. of **5**, but the trips completed by them on average is of **2**

Data Analysis- Hourly Demand

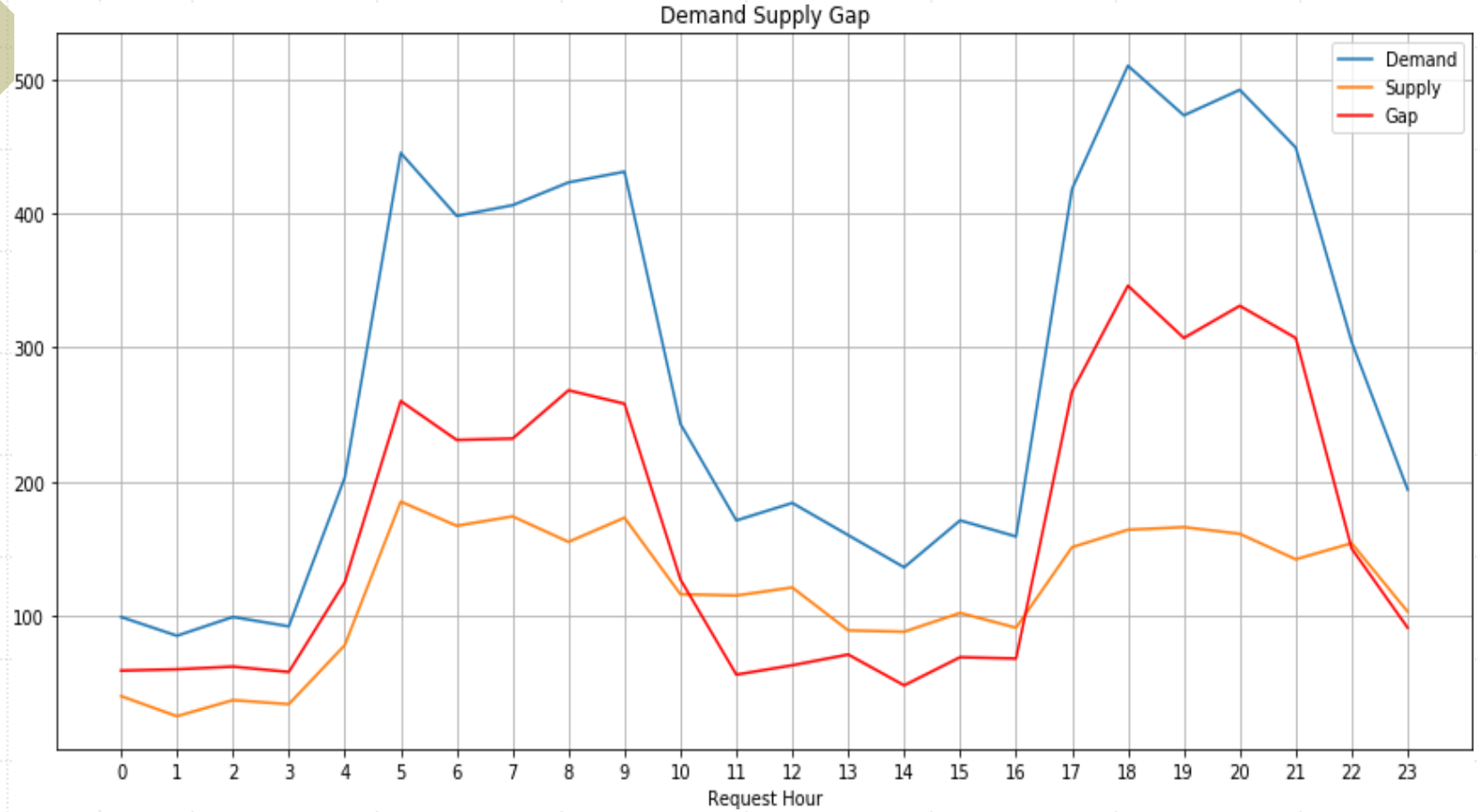
From the graph, we can observe peak-hours of demand:

- in morning
  - it is between **04:00-10:00**
- in evening
  - it is between **17:00-22:00**



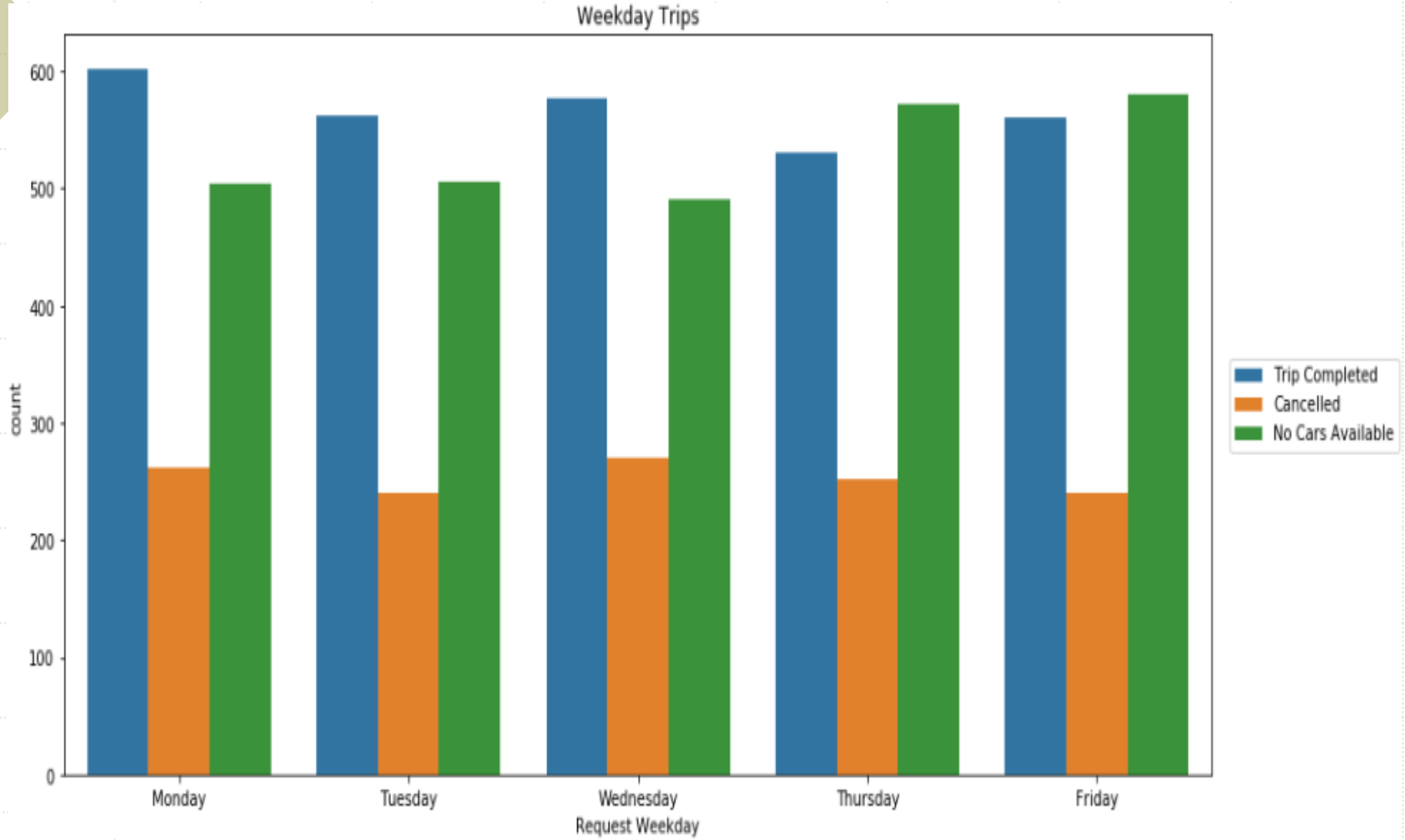
Data Analysis- Supply Demand Gap

Hourly Gap Between Supply and Demand



This graph shows the hourly trend of demand versus supply and the gap variation throughout the day. Demands remains constant between 100 to 200 with a sudden surge between 4 to 10 am in the morning and 5 to 10 pm at night.

Data Analysis- Weekday Status Count

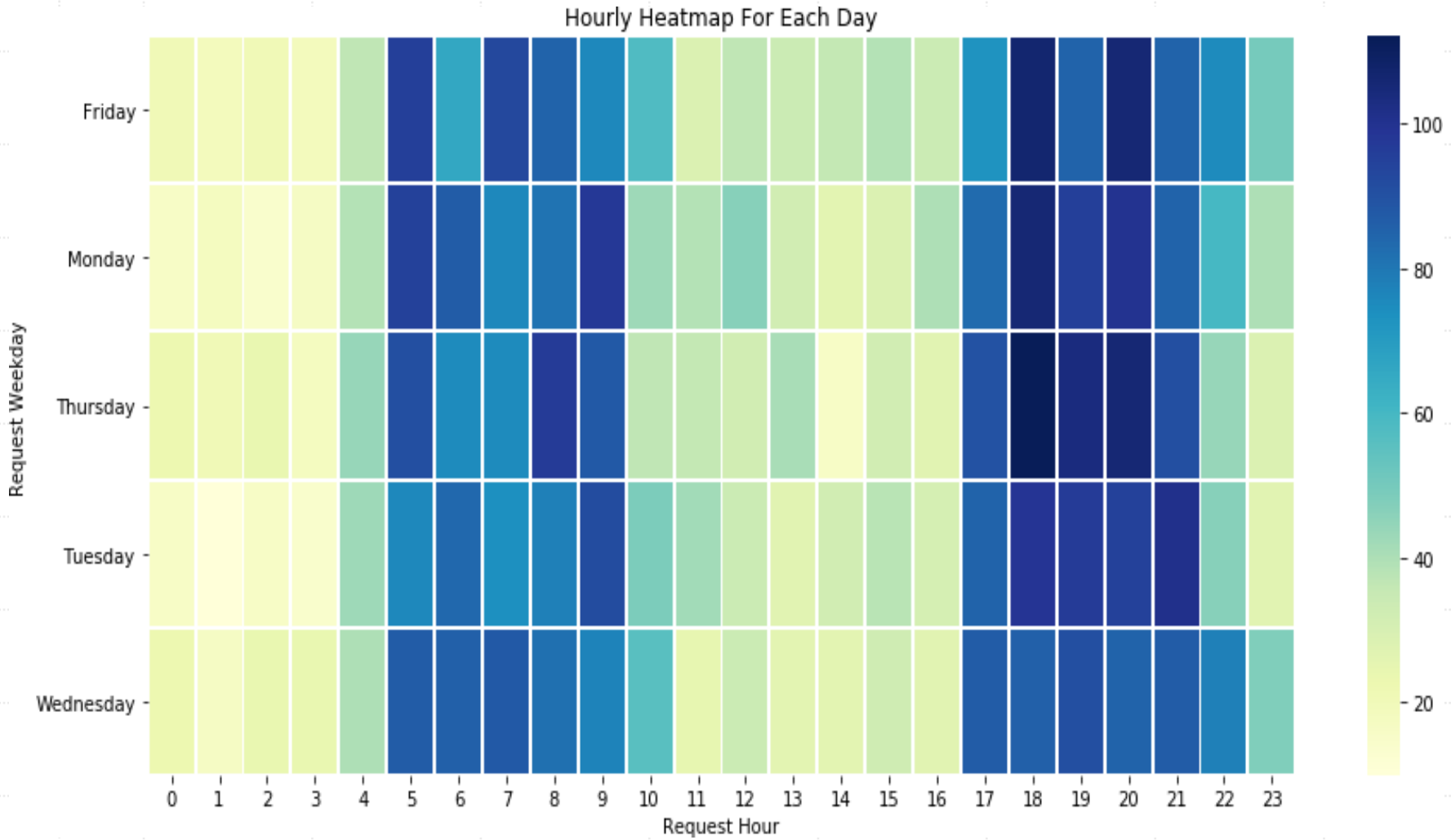


Daily Trip Analysis

From the graph we can observe:

- Number of trips completed is higher on Monday and least on Thursday
- Number of trips cancelled on Wednesday is higher
- Number of no cars availability is higher on Thursday and Friday and slightly lower on other days

Data Plotting- Hourly Demand Heatmap



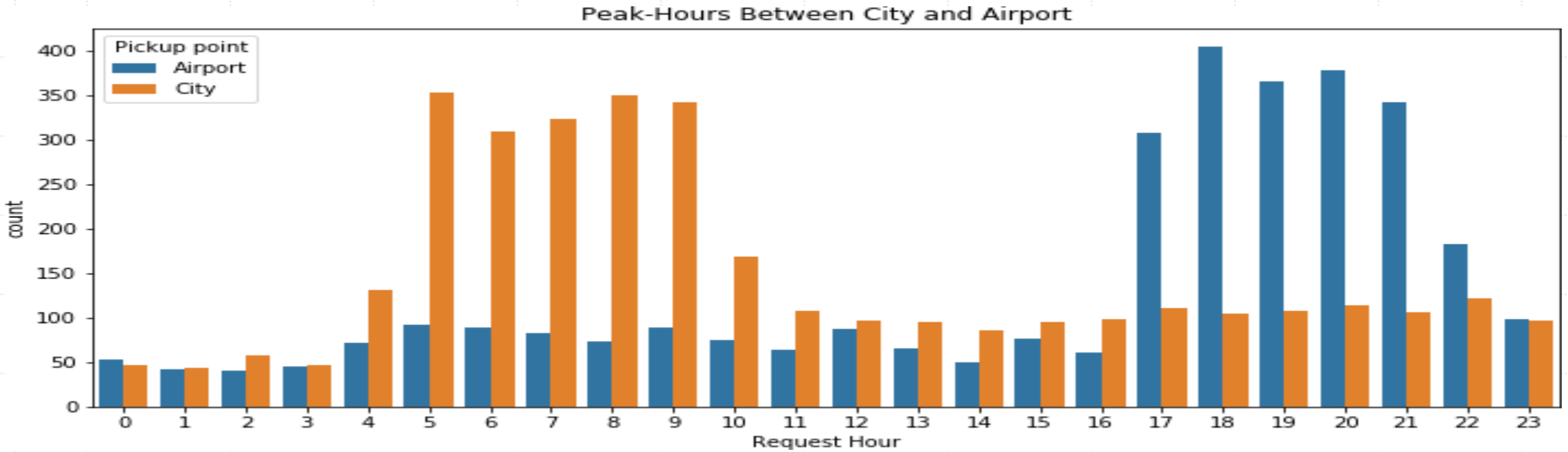
Heatmap for showing hourly demand for each day.

We can observe peak-hour for each day is between 04:00 to 10:00 in the morning and 17:00 to 22:00 in the evening and night.

Demand is very high on Monday, Thursday and Friday between 18:00 to 21:00.

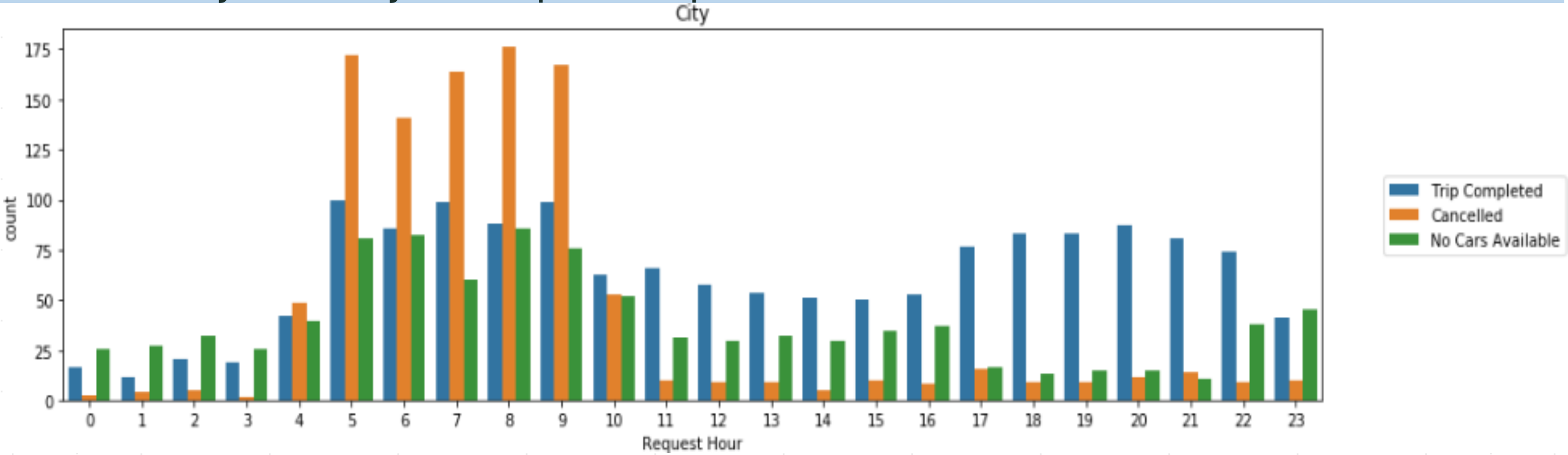
Data Analysis- Pickup Point Based Demand

Insight of peak-hours for trips from City and Airport



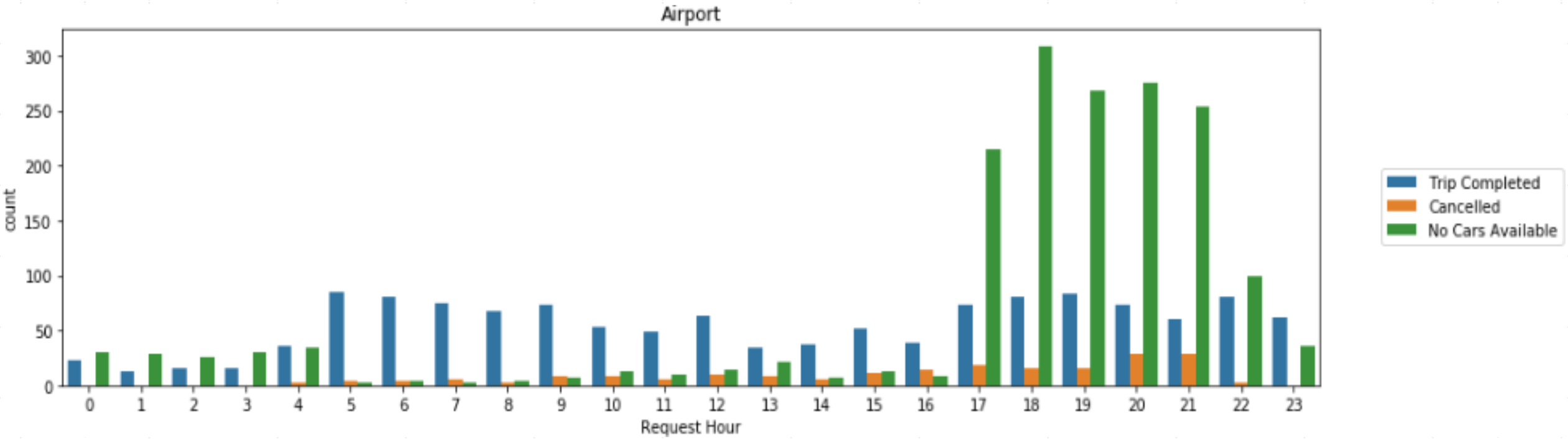
From this graph we can clearly observe, demand is high for the trips from **City-Airport** between **05:00 to 10:00** in the morning, while for the trips from **Airport-City**, the demand is high between **17:00 to 22:00** at night

Data Analysis- City to Airport Trips



- Demand is extremely high between **05:00** to **10:00** in the morning for city to airport trips
- Cancellation of trip request is also very high in the same time interval.

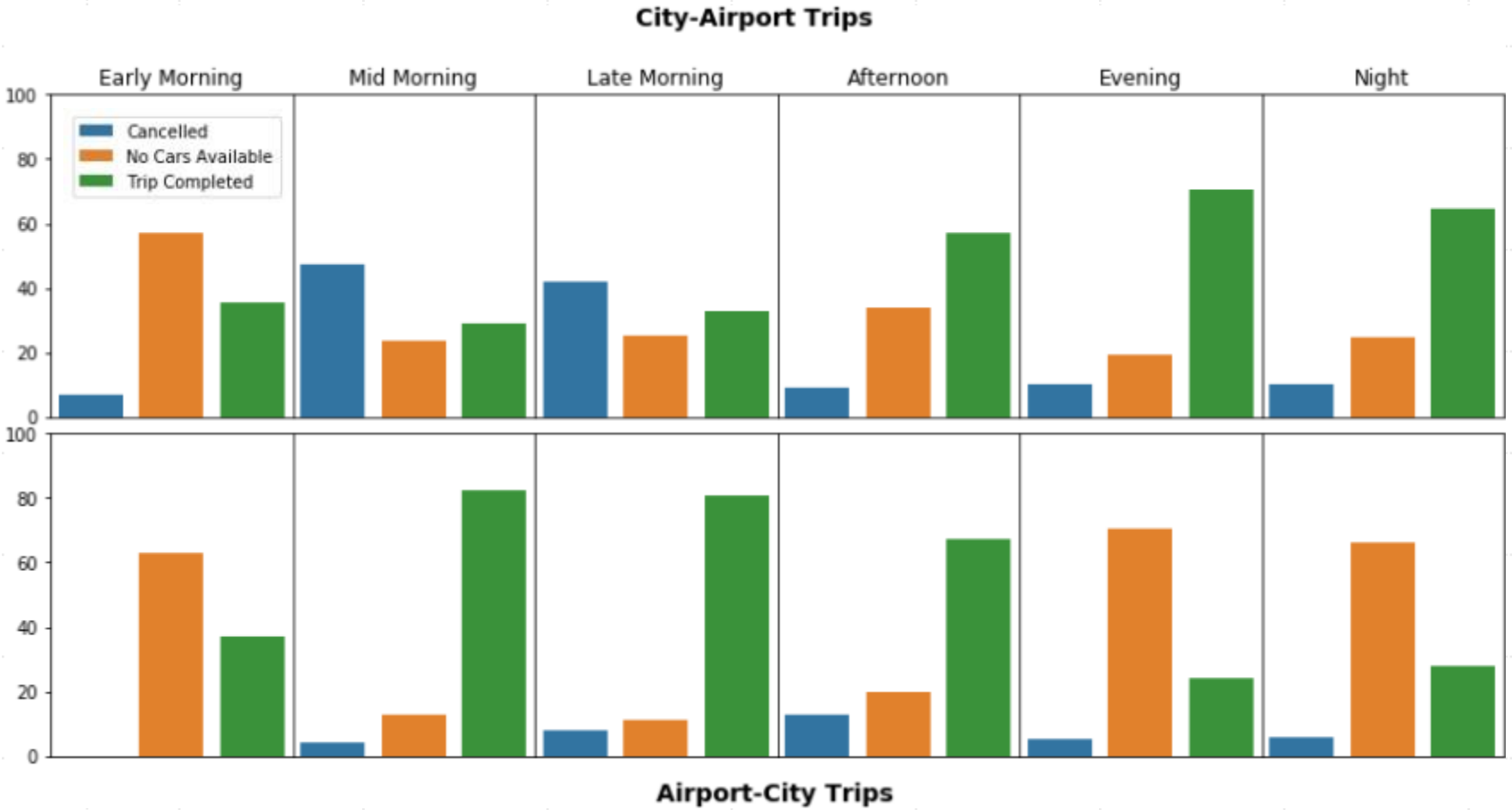
Data Analysis- Airport to City Trips



- Demand is extremely high between **17:00** in the **evening** to **22:00** in the **night** for airport to city trips
- **Availability of the car** is the main issue for supply-demand in this time interval.

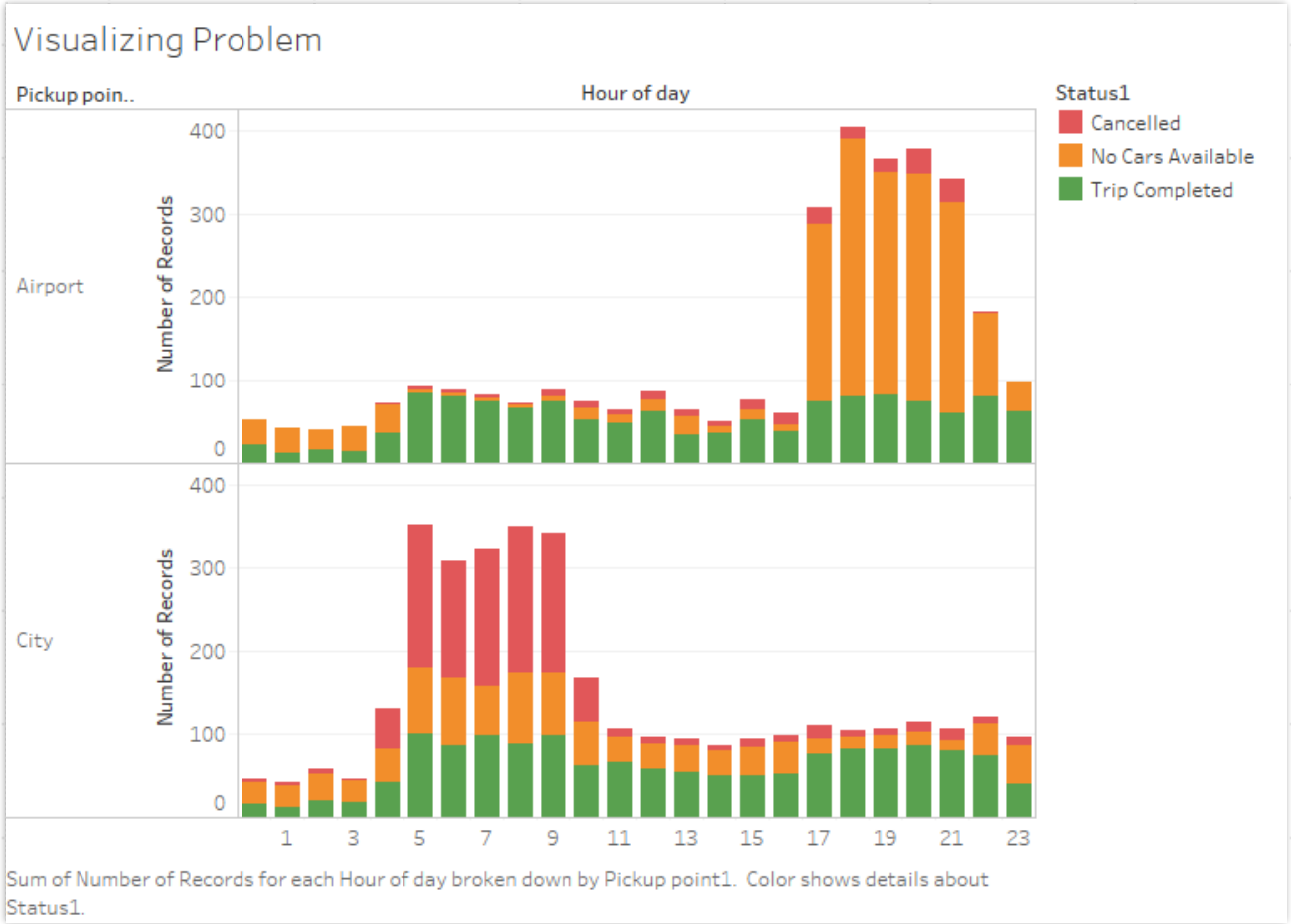
Data Analysis- Based on Time Slots and Pickup Point

Trip Analysis in Different Time Slots



Time-Slot Name	Time-Interval
Early Morning	00:00:00 – 03:59:59
Mid Morning	04:00:00 – 07:59:59
Late Morning	08:00:00 – 11:59:59
Afternoon	12:00:00 – 15:59:59
Evening	16:00:00 – 19:59:59
Night	20:00:00 – 23:59:59

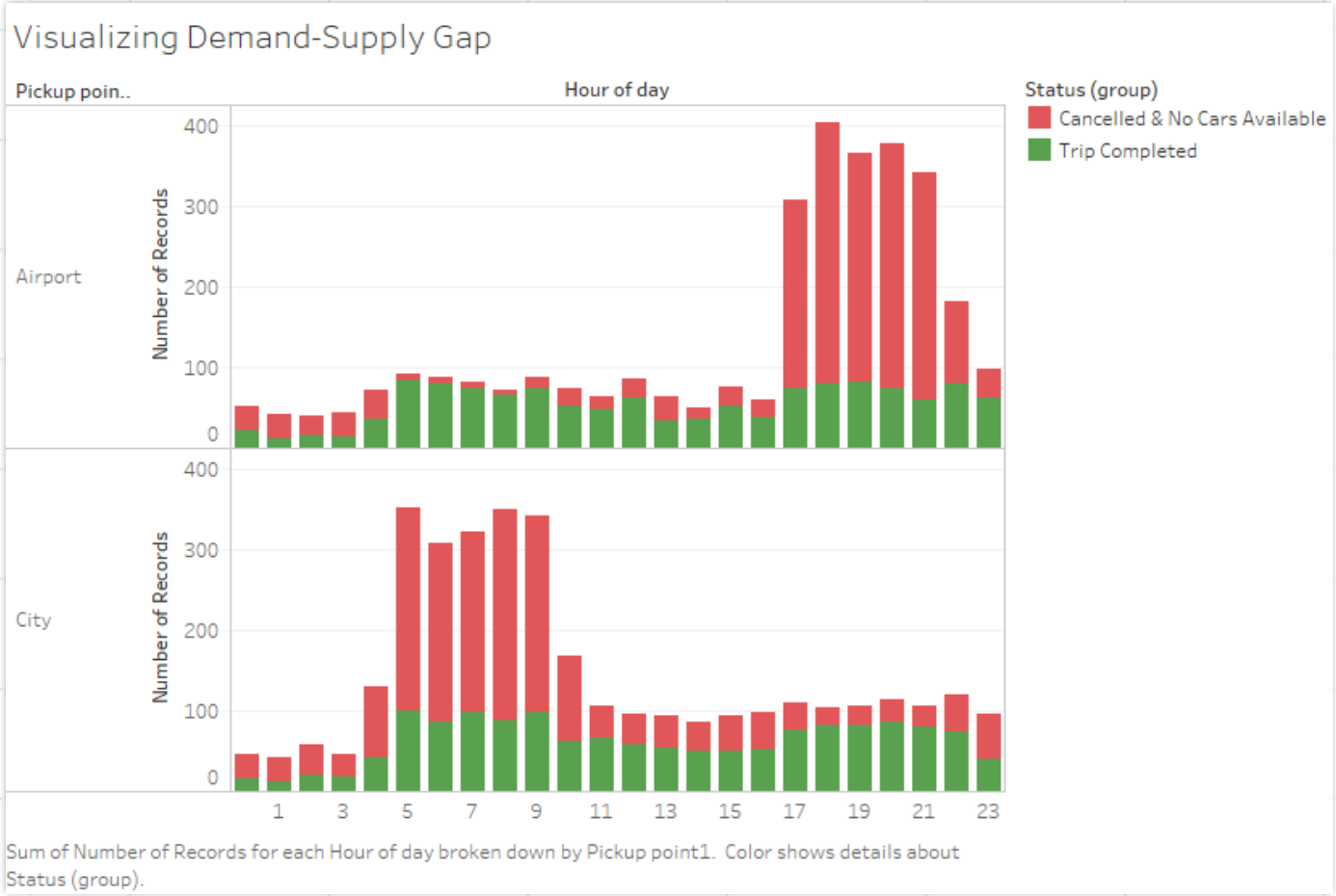
Problem Visualization



Upon observing plots, we conclude:

- Uber is facing '**No Car Available**' issue at Airport in evening and night (**5pm to 12am**).
- Uber is facing cab '**Cancelled**' issue in City during early morning and morning (**4am to 10am**)
- Uber is facing `**No Car Available**' issue in City during early morning and morning (**4am to 10am**)

Demand–Supply gap



Upon observing plots, we conclude:

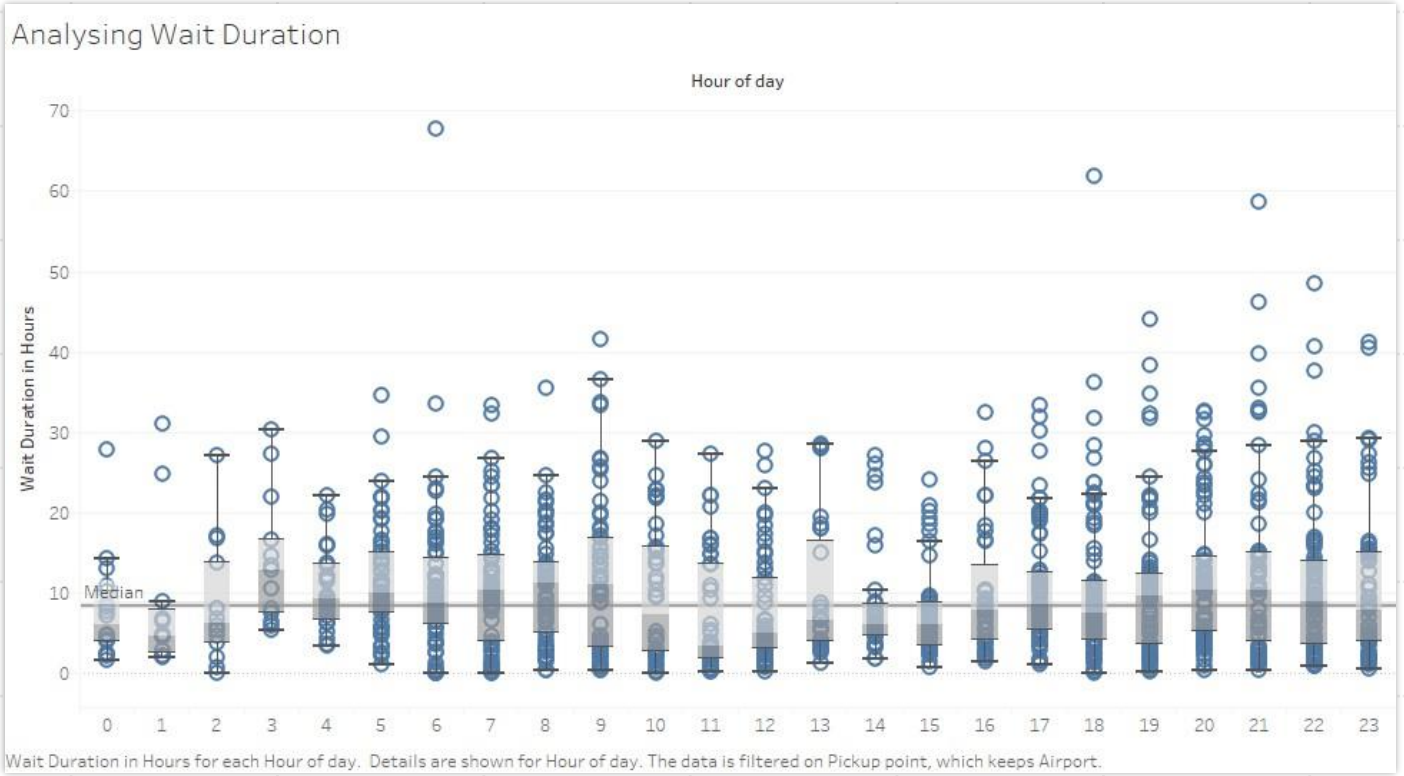
- Uber is facing high demand supply gap at Airport during **evening and night (5pm to 12am)**
- Uber is facing high demand supply gap in City during **early morning and morning (4am to 10am)**

Is traffic & Trip Duration causing issue?

Upon observing plots, we conclude:

- Even though Trip frequency is varying, the Trip duration is not varying much throughout the day.
- Thus, **Trip Duration and traffic are NOT causing problems faced by Uber**.

Is Driver wait duration causing issue?



Upon observing plots, we conclude:

- We have comparatively high wait time at Airport during **early-morning, morning and also during evening, night**.
- Cab drivers may not prefer to take bookings in city and travel instead of taking booking for the airport and keep waiting there.
- Thus, **Driver Wait Duration could be causing problems faced by Uber**.

Inferences- Based on Time Slots and Pickup Point

Inferences	Early Morning	Mid Morning	Late Morning	Afternoon	Evening	Night
City To Airport	High supply demand gap due to less number of cabs available	Number of cancellations is higher	Cancellation is slightly higher then the trips completed	Supply is good in comparison to other time slots and can be improved if there are more cabs	Demand supply gap is least in this slot	Number of trips completed is highest in this slot and less number of cancellations
Airport To City	Cabs availability is the main issue in this slot to fulfill the supply-demand gap as there is no cancellation	Number of trips completed is high in this slot and less no. of cancellations	Number of trips completed is high in this slot and less number of cancellations	Number of trips completed is high in this slot and less number of cancellations	Cabs availability is the main cause in this time slot	Cabs availability is the main cause in this time slot

Final Results 1- Problem Statements

- No cars available is the main issue for supply demand gap as we observed from “Trip Status” graph. **39.2%** of total request was not fulfilled due to non-availability of cars
- Number of drivers are less in comparison to daily demand which on average each driver can serve
- Sudden surge in demand between 4 to 10 am in the morning and 5 to 10 pm at night.
  - In morning from City to Airport trips
  - In evening from Airport to City trips
- Cancellation of trips are higher between 4 to 10 am time slot in City to Airport trips resulting in high gap between supply and demand. The difference between demand and supply is 1205. Out of 1677 requests, **48.9%** is due to cancellation of trip.
- Cabs non-availability is the main issue between 5 to 10 pm time slot in Airport to City trip requests for the high supply-demand gap. The difference between demand and supply is 1427. Out of 1800 requests, **73.9%** is due to non-availability of car.

Final Results 2 – Problem Statements

- For addressing **early morning and morning 'Cancelled' rides for Airport**
  - A penalty for driver canceling Airport booking thrice a day
  - Provide incentive to drivers for Airport Rides / Airport Wait Duration

- For addressing **'No Car Available' issue at Airport in evening and night**
  - Airport Rides can be given a weight of 1.5 ride count
  - Exempt drivers with Airport Rides (in evening and night) from daily minimum rides

- For addressing **'No Car Available' issue in City during early morning and morning**
  - Provide incentive to drivers for Airport Rides / Airport Wait Duration

Final Results- Recommendations

- Hiring more part-time drivers to overcome the non-availability of cars problem during peak-hours.
- Increasing profit margins for drivers so that they don’t cancel the trip for city to airport in peak morning hours.
- Surge pricing when the demand is high to increase revenue while maintaining transparency.
- Increase demand at the airport through marketing initiatives so that drivers don’t have to wait for a longer time.
- Uber can pay drivers to come without passengers from the airport if they are not getting and pickups.

Thank You!