Uber Supply-Demand Gap Analysis

By: Ingit Bhatnagar

Problem Statement

Uber is facing issues like - driver cancellation and non-availability of cars leading to a loss of potential revenue. As an analyst, you decide to address the problem by analyzing the sample data set from its warehouse.

Objective

The aim of the analysis is to identify the root cause of the problem (i.e. cancellation and non-availability of cars) and recommend ways to improve the situation. As a result of your analysis, you should be able to present to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

Data Given

The dataset contains data about the rides taken for roughly a few days from the City to the Airport and vice versa. Sample data:

	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

Analysis approach

1) Explore the data set

2) Data Handling and Cleaning

- Identify nulls, duplicates, garbage values

- Standardize data formats

- Fix data types

- Derive new meaningful columns if needed, etc

3) Draw insights to find problem areas

- Analyze individual variables

- Analyze the relationship/dependency between variables

- Find patterns/problems that may cause the issues

- Create visualizations to better demonstrate the insights

4) Supply Demand Gap Analysis

5) Suggest solutions to rectify the problems

Explore the Data Set

Below are the observations:

- Request Id (Integer): No nulls/duplicates

- Pickup point (String): No nulls/garbage values. Possible values: ["City", "Airport"]

- Driver id (Float): Is null where "Status" is "No Cars Available". Hence, it is "Missing not at random"(MNAR).

- Status (String): No nulls/garbage values. Possible values: ["Trip Completed", "No Cars Available", "Cancelled"]

- Request timestamp (DateTime): No nulls. Has multiple date formats like: "11/7/2016 13:00" and "13-07-2016 08:33:16"

- Drop timestamp (DateTime): Is null where "Status" is "No Cars Available" or "Cancelled". Hence, it is "Missing not at random"(MNAR). Has multiple date formats like: "11/7/2016 13:00" and "13-07-2016 08:33:16"

Data Handling and Cleaning

- Dropped "Request id" column

- Set null values in "Driver id" column as -1, and changed data type from float to int since Id is integer.

- Standardized date format of "Request timestamp" and "Drop timestamp" columns to "yyyy-mm-dd hh:mm:ss".

- Extracted "Date", "Time", "Day of the Week" and "Hour" from the timestamps for further analysis.

- Extracted the hours from the timestamp(which follows 24 hours clock format) and clubbed them as below(left and right inclusive):
  0-3 : Late Night
  4-6 : Early Morning
  7-11 : Morning
  12-16 : Afternoon
 17-20 : Evening
  21-23 : Night

  For instance,
  - ❖ If request time is "03:10"(3:10am), then the extracted hour is "3", and hence it will be considered Late night
  - ❖ If request time is "20:59"(8:59pm), then the extracted hour is "20", and hence it will be considered Evening

- The percentages referred to throughout the slide are a percentage of the total number of requests.

- "Demand" is considered 1 for all requests. Supply is considered 1 for requests where status is "Trip Completed", else 0. Gap is difference between Demand and Supply. Hence, if Gap=0 for a request, that means it was "Fulfilled", else "Not fulfilled"
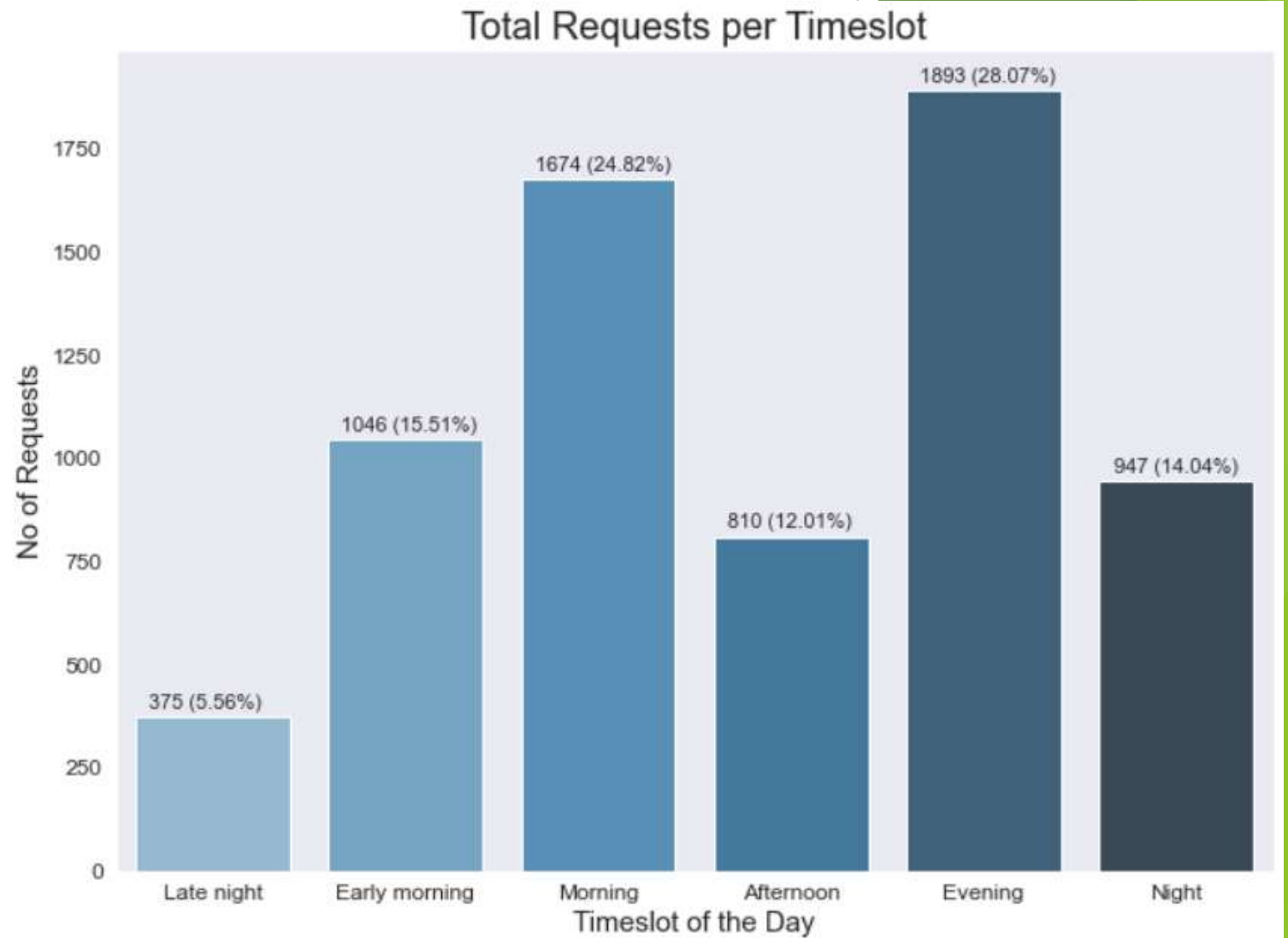
Univariate Analysis
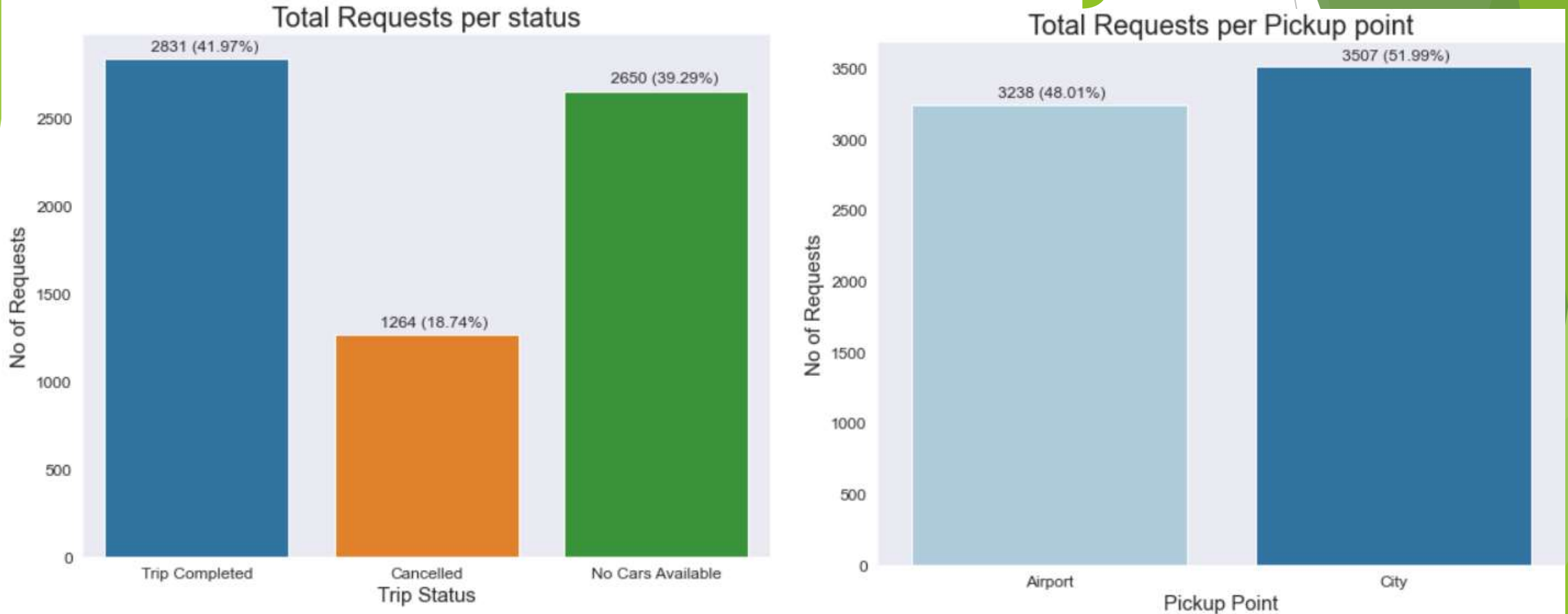
- **Total Requests per Status**: Only **41.97%** of the total requests are successfully completed. A major issue for unfulfilled requests is Cab unavailability(**39.29%**), while **18.74%** of requests are cancelled.

- **Total Requests per Pickup point**: The number of requests is almost even from the City and the Airport.

Total Requests per Day

Total Requests per Hour of the Day
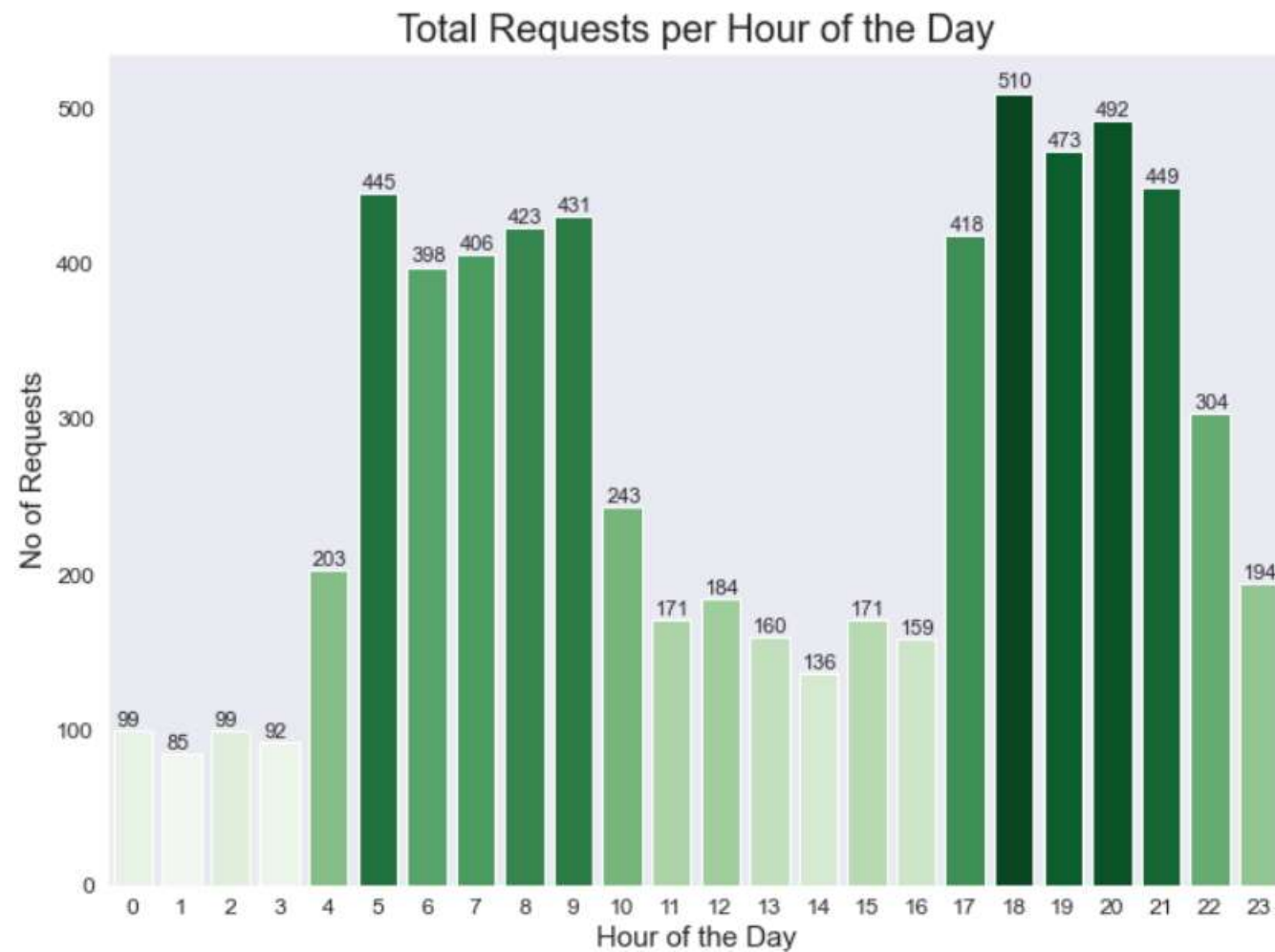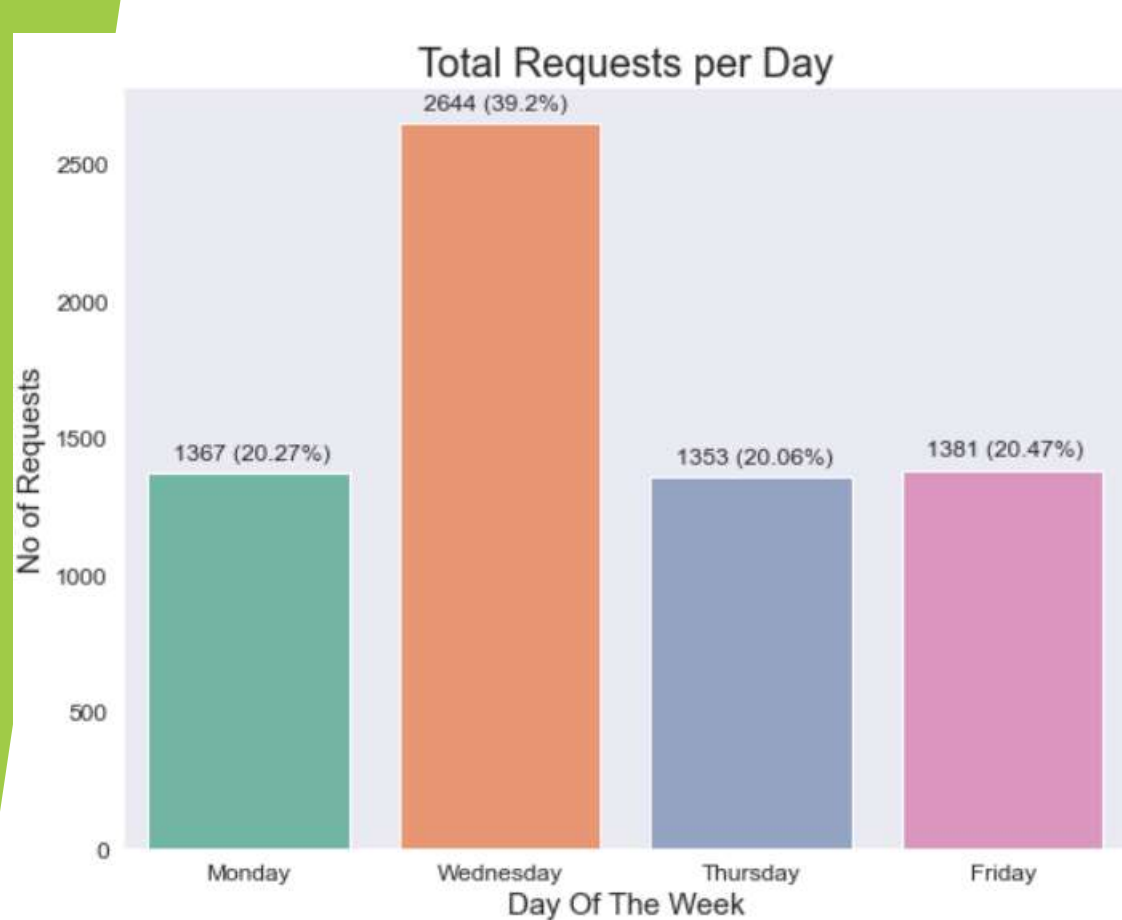
- **Total Requests per Day**: Requests are high on Wednesday, but since the data set is limited to one week, we can't conclude it in general. Hence, ignoring it for now.

- **Total requests per Hour of the Day**: Requests are very high between **5-9(5am to 9am)[31.17%]** and then fall suddenly between **10-16(10am to 4pm)[18.14%]**, and then surge again between **5-10(5pm to 10pm)[39.2%]**. The requests are the least between **0-3(12am to 3am)[5.55%]**
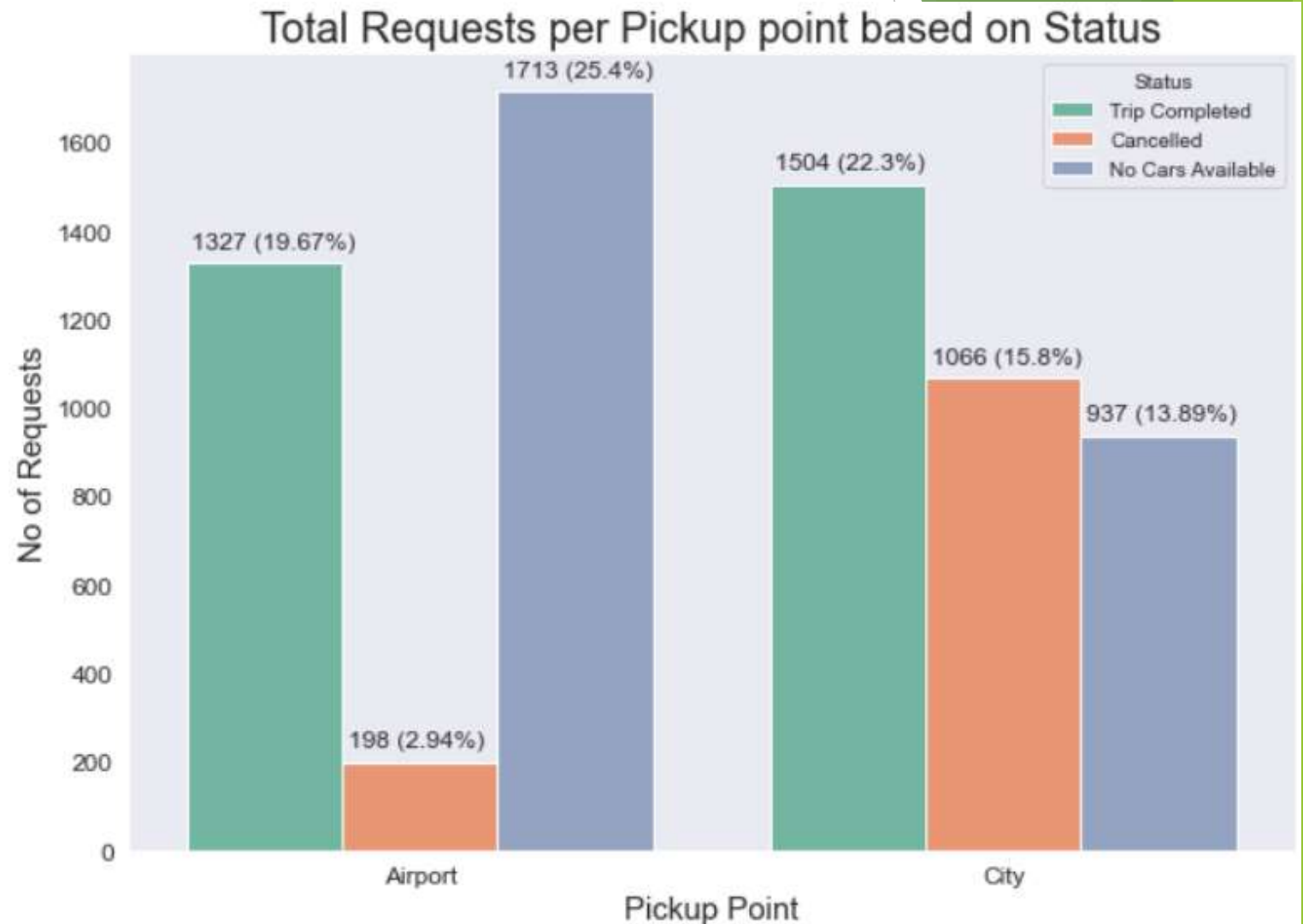
Total Requests per Timeslot: After grouping the requests in timeslots we can say that the requests start increasing from the **Early morning(15.51%)** reaching a peak in the **Morning(24.82%)**, after which there is a sudden decline in the requests all **Afternoon(12.01%)**. Then there is another peak in requests during the **Evening(28.07%)** which drops at **Night(14.04%)**. The requests are minimum during **Late night(5.56%)**

Total Requests per Timeslot

Bivariate Analysis

Total Requests per Pickup point for each Status:

- A major number of unfulfilled trips at the **Airport are due to "No Cars Available"(25.4%)**, while in the **City, about 15.8% of requests are "Cancelled"**, while **13.89% of requests are unsuccessful due to "No Cars Available"**

- The drivers from the City might not prefer going to the airport, due to which there are huge cancellations, and hence fewer drivers available at the Airport.


Total Requests per Pickup point based on Status

Total Requests per Timeslot based on Pickup Point:

- The graph indicates that there is a huge number of requests **from the City to the Airport during** the **Early morning(11.76%)** and **Morning slots(19.13%)**, while a huge number of requests from the **Airport to the City during** the **Evening slots(21.6%)**.

Total Requests per Timeslot based on Pickup point
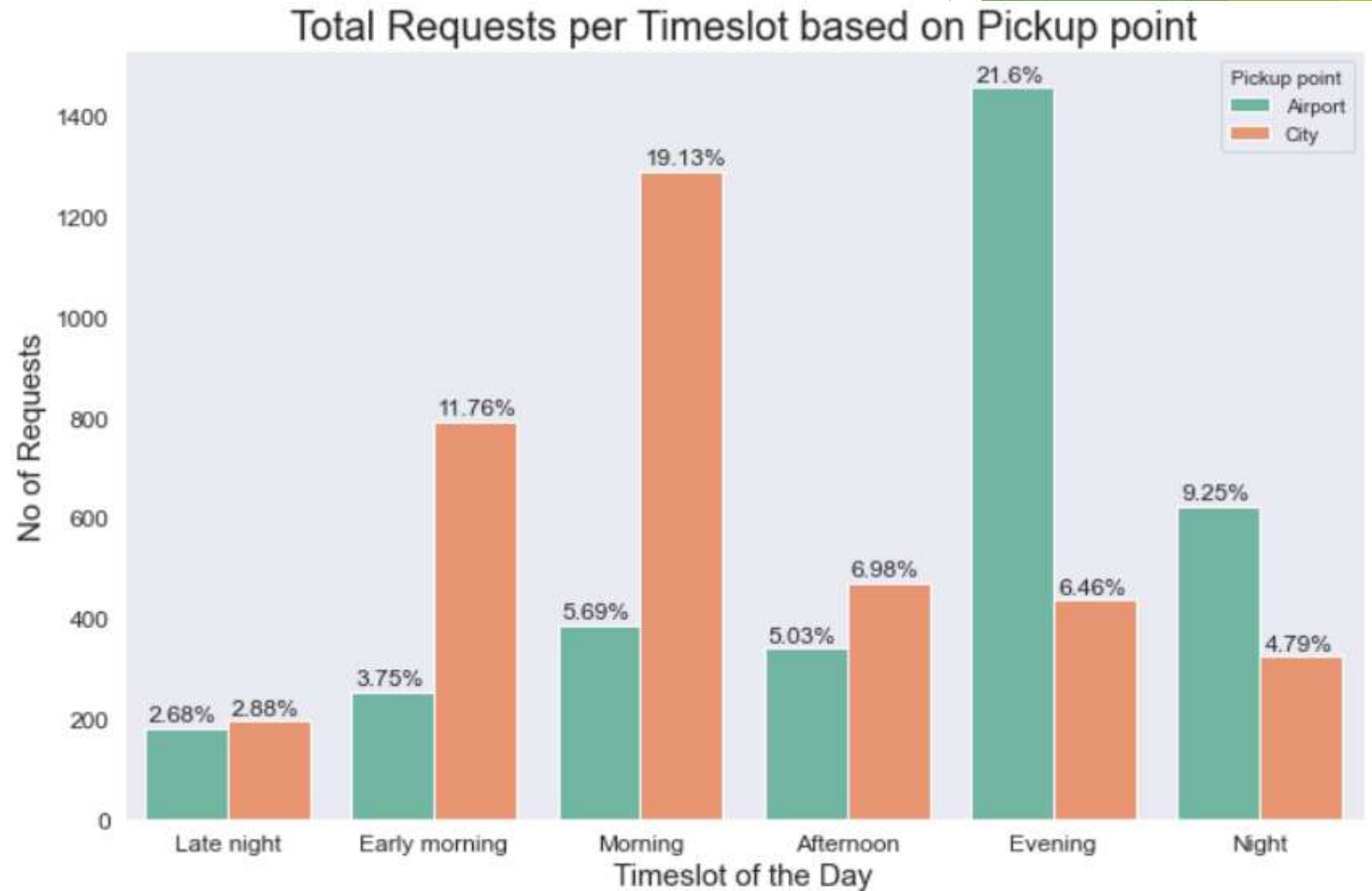
Total requests per Timeslot based on Status:

- The graph indicates that there is a huge shortage of cabs in the Evening timeslot while the Cancellation rate is very high during the Early morning and Morning timeslots.

- Combining this knowledge with the previous two graphs, we can observe the below:

- From City to Airport:
  * The requests are high during Early morning and Morning slots, but there are a huge number of cancellations. We also see some shortage of Cars

- From Airport to City:
  * The requests are high during Evening slots, but there is a huge unavailability of cars.

Total Requests per Status at each Timeslot
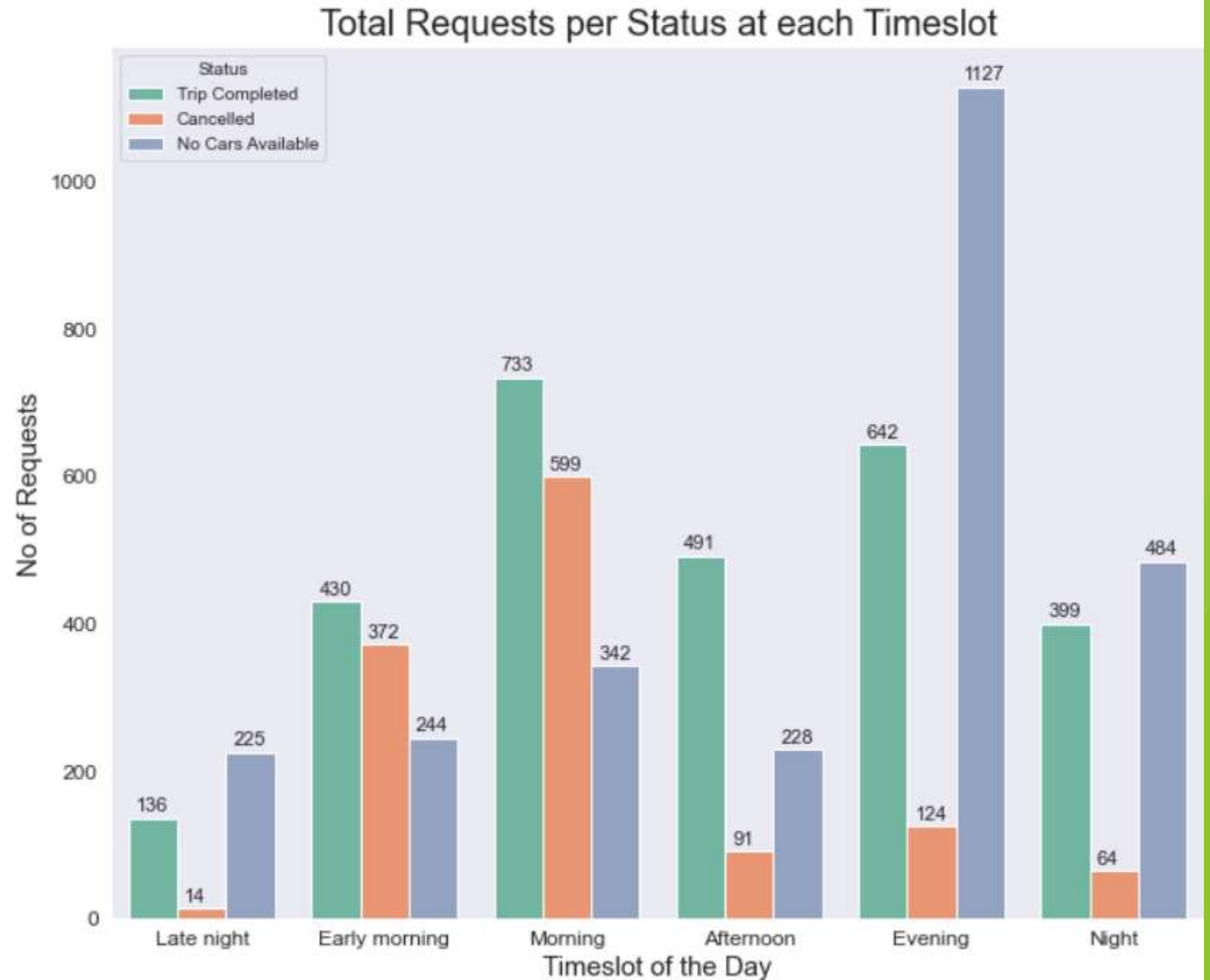
No of Rides fulfilled by Driver per day (Boxplot):

- This shows that there is almost a medium spread of rides among the riders each day.

- However, there are drivers(for instance on Wednesday), who are fulfilling about 8-12 rides, while few are only fulfilling 1-4 rides.

- Uber could also look into this and come up with some ways to spread the number of rides almost evenly among more drivers throughout the day.

- We could put a daily target for the number of rides to be completed by each driver based on the average number of rides per day.

Distribution of Drivers taking given number of rides each day
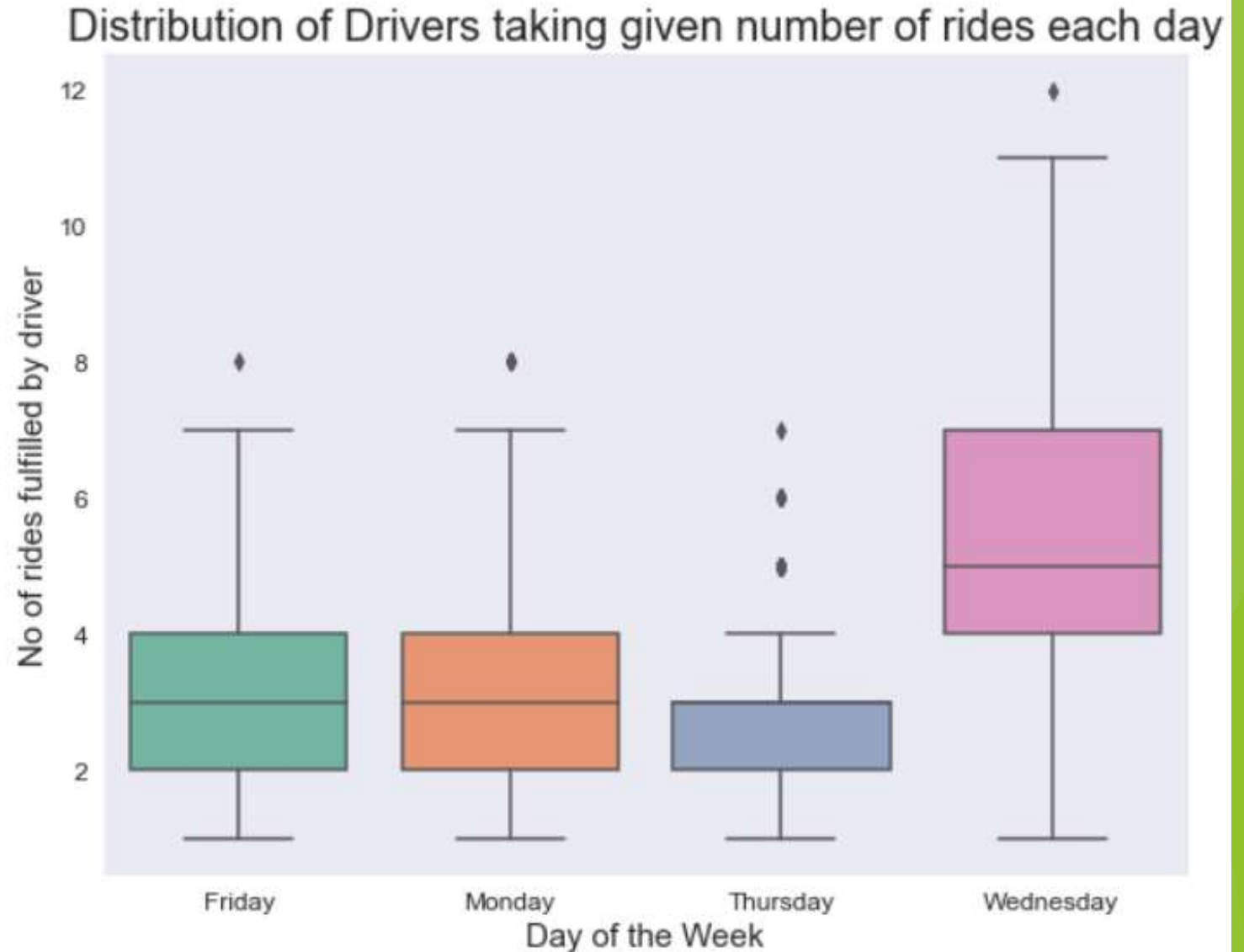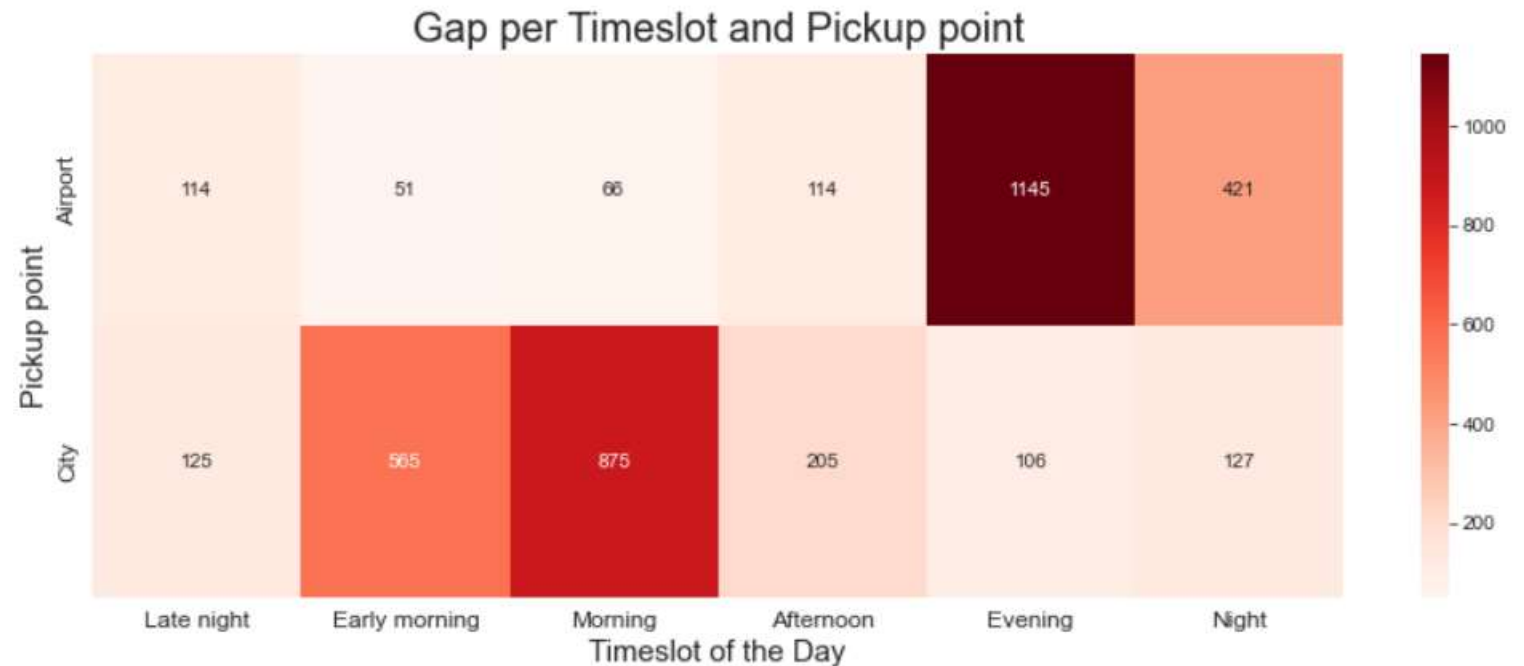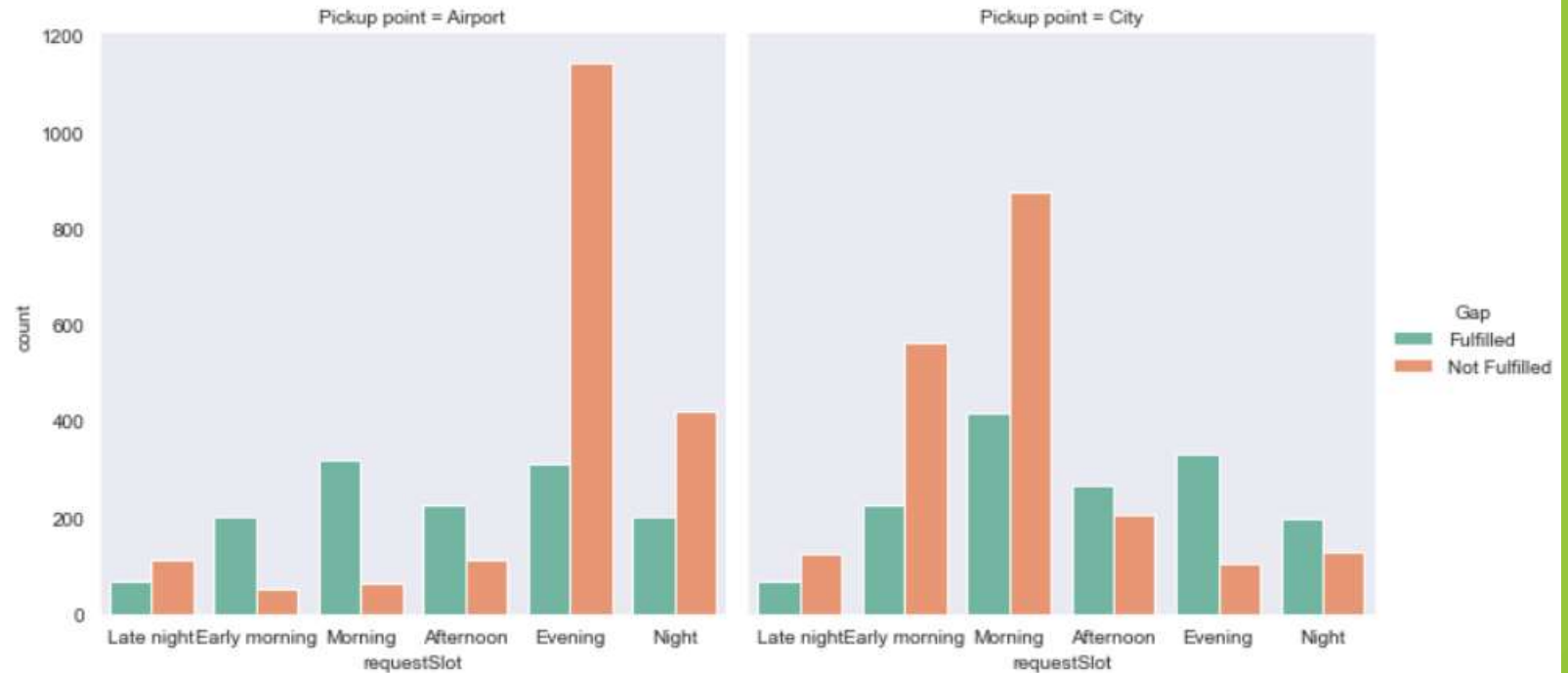
Supply-Demand Gap Analysis
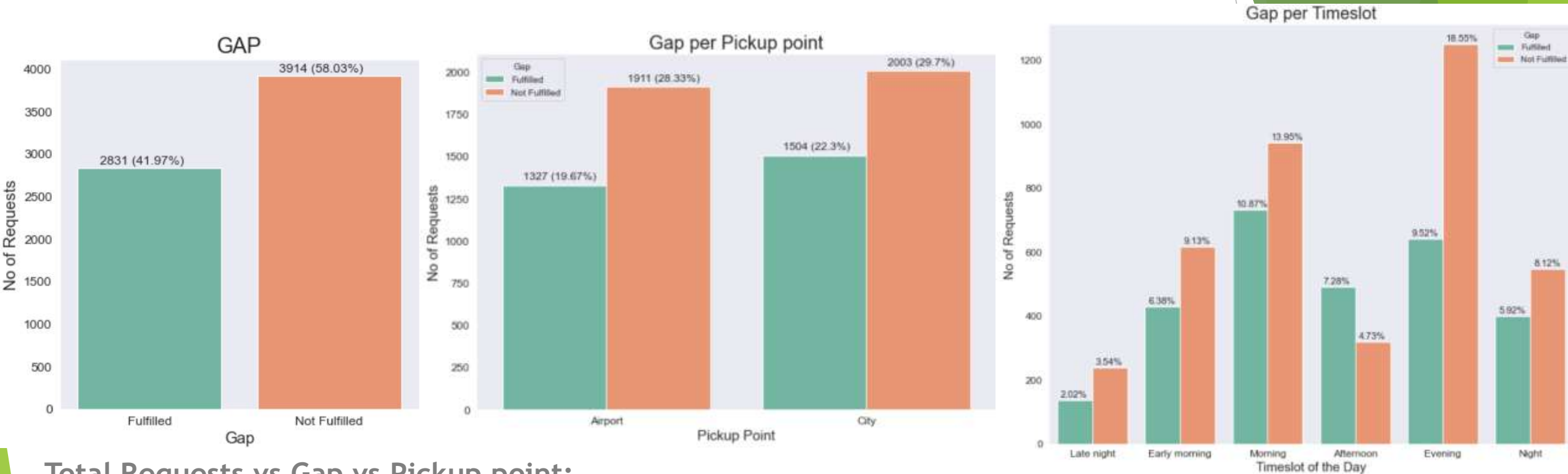
Total Requests vs Gap vs Pickup point:

- If we consider Gap as the difference between supply and demand, then the above graph shows that there is about a **58.03% unfulfilled Gap**.

- The unfulfilled Gap at the **Airport is about 28.33%**, while in the **City they are about 29.7%**.

- The unfulfilled Gap is **highest during the Early morning(9.13%), Morning(13.95%)** and **Evening slots(18.55%)**, while it is least during the **Late night slot(3.45%)**.

Total Requests vs Timeslots vs Gap for each Pickup point:

- These graphs give a clearer picture of the problem areas at each pickup point during different times of the day and it is all aligned with our analysis so far.

- There is a huge unfulfilled Gap at the Airport during Evening and Night slots, while huge unfulfilled Gap in the City during Early morning and Morning slots, which needs to be addressed.

Observations from the Analysis

1. Only **41.97%** of requests are fulfilled successfully. Through the Supply, Demand and Gap parameters, we are able to identify that there is a huge gap between the Supply and Demand and hence some measures need to be taken to improve Uber's Revenue

2. There is a **high demand** for rides from the **City to the Airport** during the **Early morning(15.51%)** and **Morning(24.82%)**, and from **Airport to City** during the **Evening(28.07%)**. However, there is very less demand at these pickup points for the remaining slots respectively.
   - This could be the reason why the Cab drivers are cancelling the rides from the City to Airport because there is no longer much demand till the evening at the Airport and they will have to wait till then or come back to the city idly.
   - This could also be the reason why there are a huge number of unfulfilled demands from the Airport to the City in the evening because there are not many drivers around the airport

3. There is a greater failure due to **"No cars available"(39.29%)** than **"Cancelled"(18.74%)** in general. This could mean, that we should look into the shortage of Cabs and possibly increase the number of drivers to fulfill this Gap.

4. We also see that the rides are not evenly spread for all the drivers, and there is scope to improve that spread as well to reduce the workload on a few drivers.

Analysis Hypothesis

1. <u>PEAK MORNING ISSUE (CITY TO AIRPORT)</u>:

   - As per the analysis, there is a huge demand from City to the Airport in the morning hours(5am to 10am) and a large number of requests are being cancelled, leading to an increase in the Supply & Demand Gap. There is a shortage of drivers as well leading to further increase in the Gap.

   - The high demand could be due to a lot of morning flights departing from the City.

   - The supply could be less because the drivers might not be willing to travel that far to the airport from the city during the morning possibly because there is less demand from the Airport back to the City during the afternoon. The drivers might not prefer waiting that long or coming back idle to the city. They may prefer taking smaller rides within the City itself.

2. <u>PEAK EVENING ISSUE (AIRPORT TO CITY)</u>:

   - As per the analysis, there is a huge demand from Airport to the City in the evening(5pm to 10pm) and but there is a huge shortage of drivers, leading to an increase in the Supply & Demand Gap.

   - The high demand could be due to a lot of evening flights arriving in the City.

   - The supply could be less since not many drivers are willing to come to the airport in the morning, leading to fewer drivers in the area by the evening.

Recommendations

- Add **dedicated Uber stands at the airport** making cabs available all the time, thus reducing the unfulfilled requests

- **Add a few more Uber drivers** since a major reason for unfulfilled trips is **"No cars available"(39.29%)**

- **Incentivize drivers** for taking rides to the Airport **during peak hours**. This would encourage them to take more rides from City to the Airport, thus reducing the number of cancellations. (Positive reinforcement)

- **Penalize drivers who frequently cancel the rides**, thus reducing the number of cancellations. (Negative reinforcement)

- **Promote pre-booking among the customers** by giving some discounts so that the drivers are well aware of the schedule in advance and can plan their routine accordingly.

- Educate drivers and share the analysis with them, **talk about peak hours and opportunities**, so that they know when and where to be during the day to fulfill more requests.

- Enhance the algo to find drivers by **making it prioritize the drivers who have taken lesser rides** during the day so far, thereby ensuring even distribution of rides among the drivers

- Have a **target number of rides to be achieved** by the drivers based on the average number of rides requested each day. We could reduce the incentives for riders who fault on this frequently.