UBER CASE STUDY

SUBMISSION

By:
Nikhil Gupta

Business Objective

The aim of analysis is to:

- Identify the root cause of the problem (i.e. cancellation and non-availability of cars)

- Recommend ways to improve the situation.

Notes:

- The analysis is done for weekday requests only as the provided data contains requests made by customers on a weekday only.

- The analysis is done only for city-airport and airport-city requests.

Let's understand the data

There are six attributes 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, that can be either completed, cancelled by the driver or no cars available

Analysis Methodology

1. Importing the data into Rstudio

2. Understand the data by visual inspection

3. Fixing Rows and Columns for missing values, standardising values, removing invalid values

4. Separating relevant information from columns for further analysis (ex: Day of the month from Request Timestamp)

5. Extrapolate the derived metrics

6. Filter data on the basis of Pickup.point

7. Perform Univariate and Segmented Univariate analysis on the data

8. Perform Bivariate analysis on the data

9. Plot the results and extract the root cause of the problem
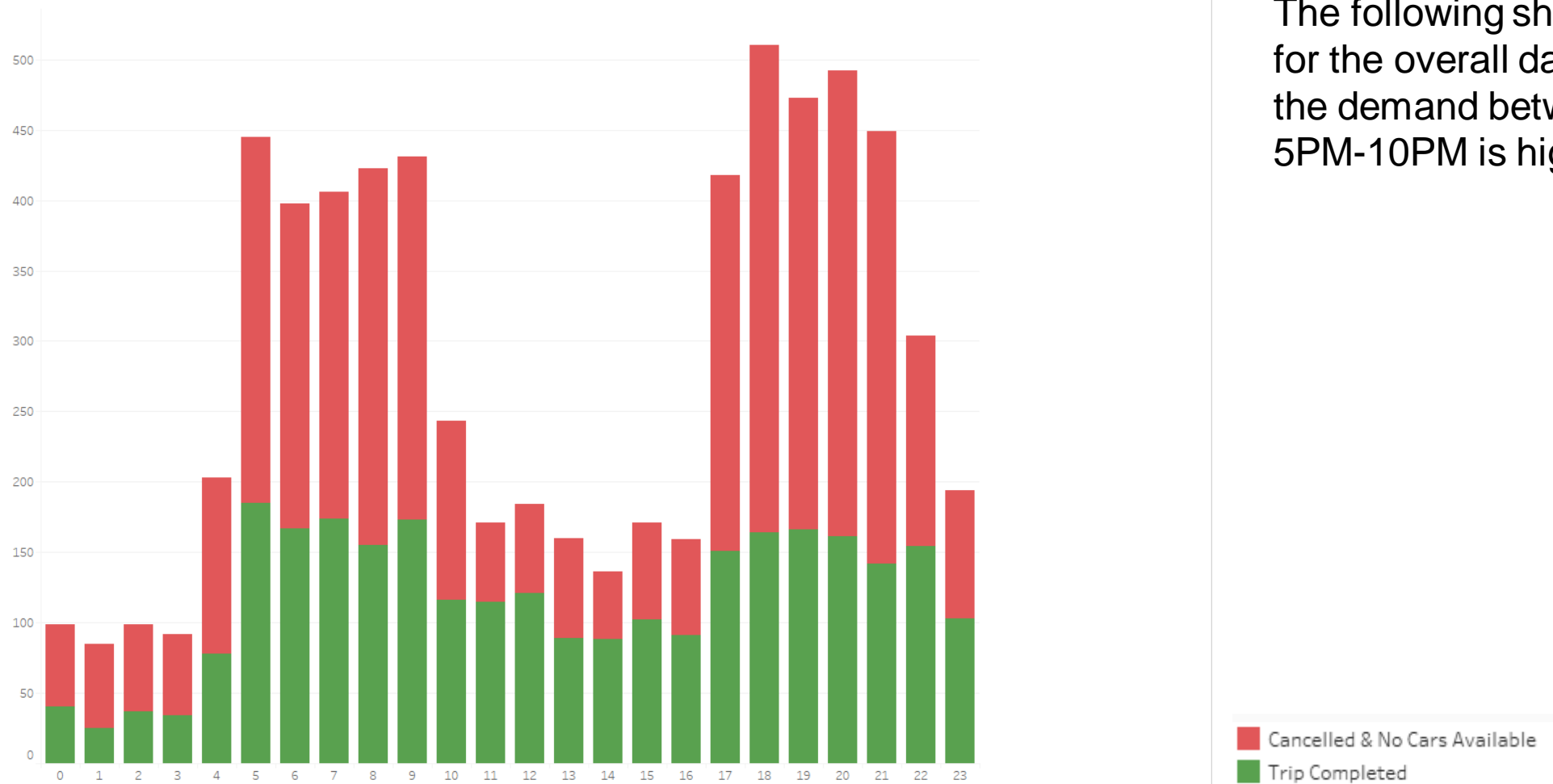
Plot-1 Hour of the day Vs Requests
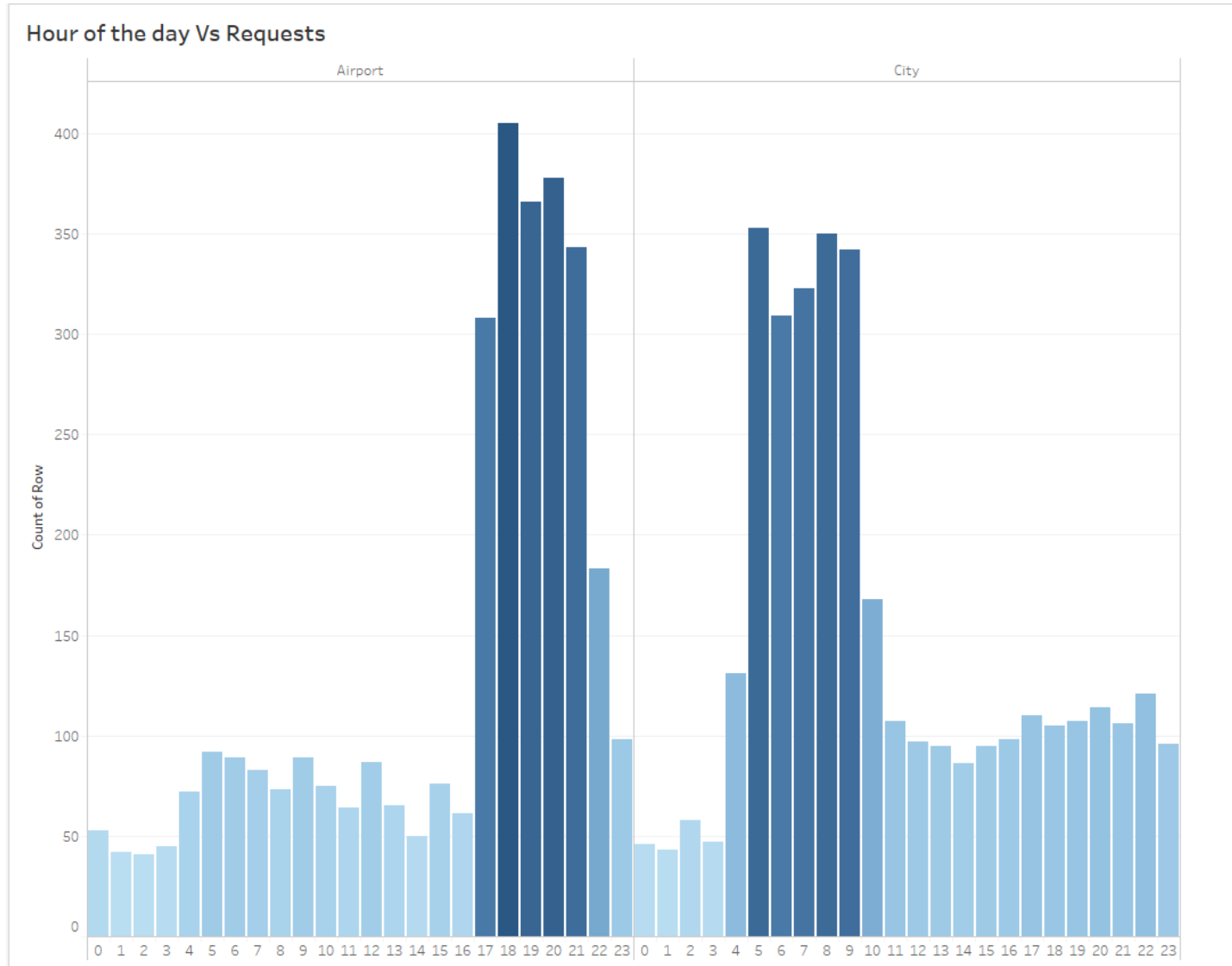
There is a high demand of cabs at city between 04:00 AM to 10:00AM for airport.

A very high demand at the airport between 05:00 PM to 10:00 PM and the trips completed by drivers are very less.
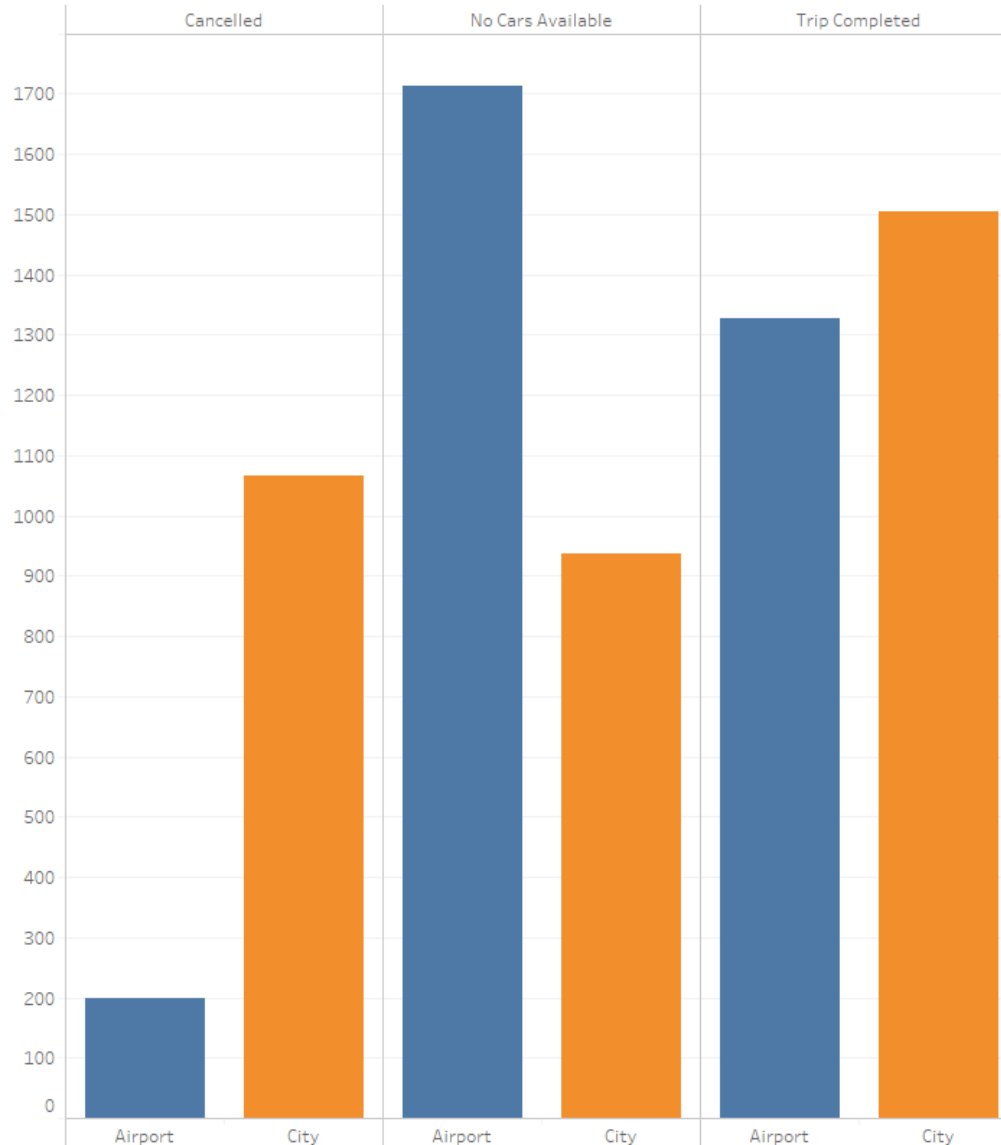
Plot-2 Demand Vs Supply

Demand and Supply at different Hours

The following shows the demand vs supply for the overall data. It clearly shows that tha the demand between 4AM-10PM and 5PM-10PM is high and the supply is low
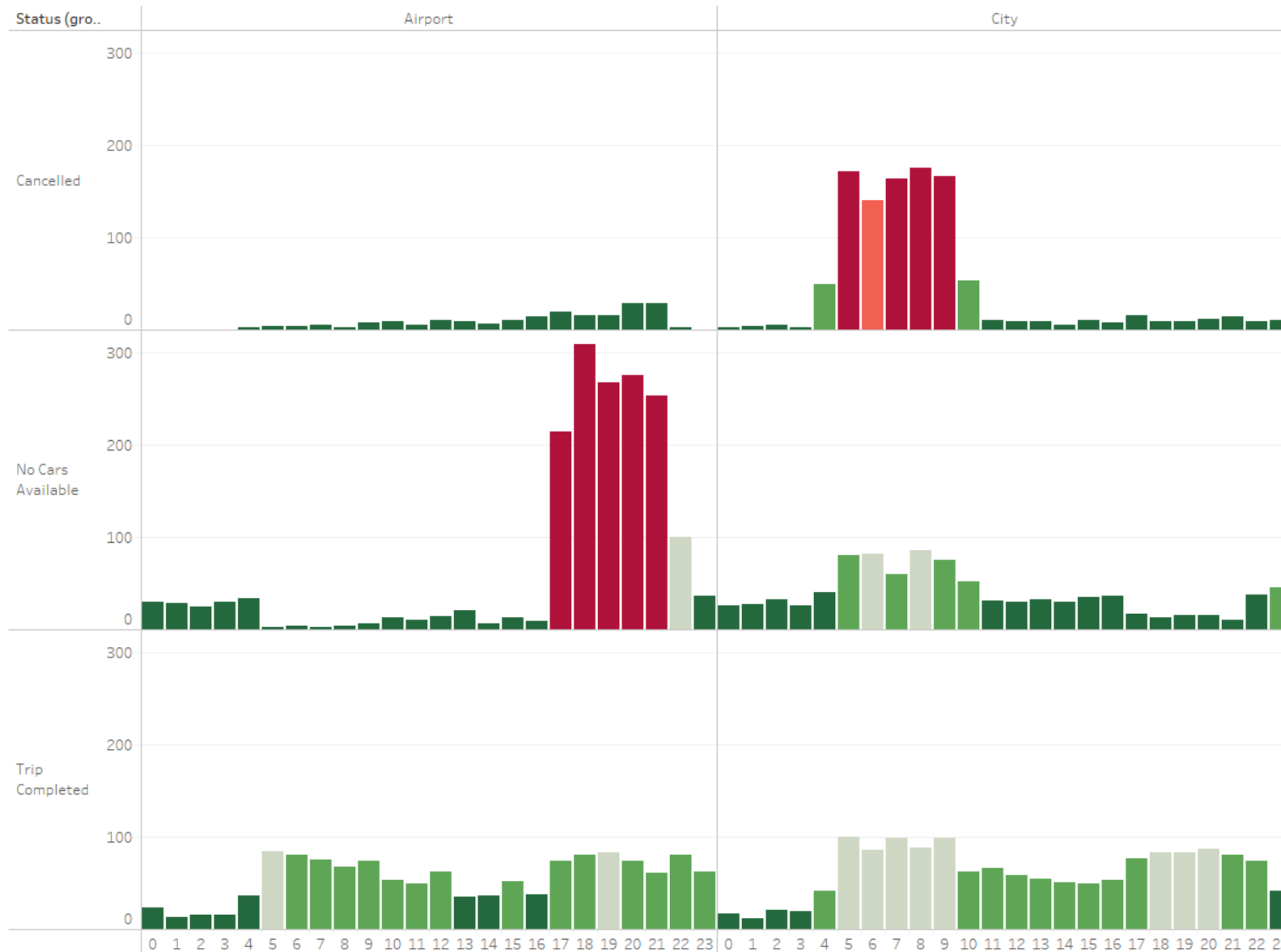
Plot-3 Pickup point Vs Status

Trip status for City & Airport at different Hours

Plot-5 Supply vs Demand for City and Airport
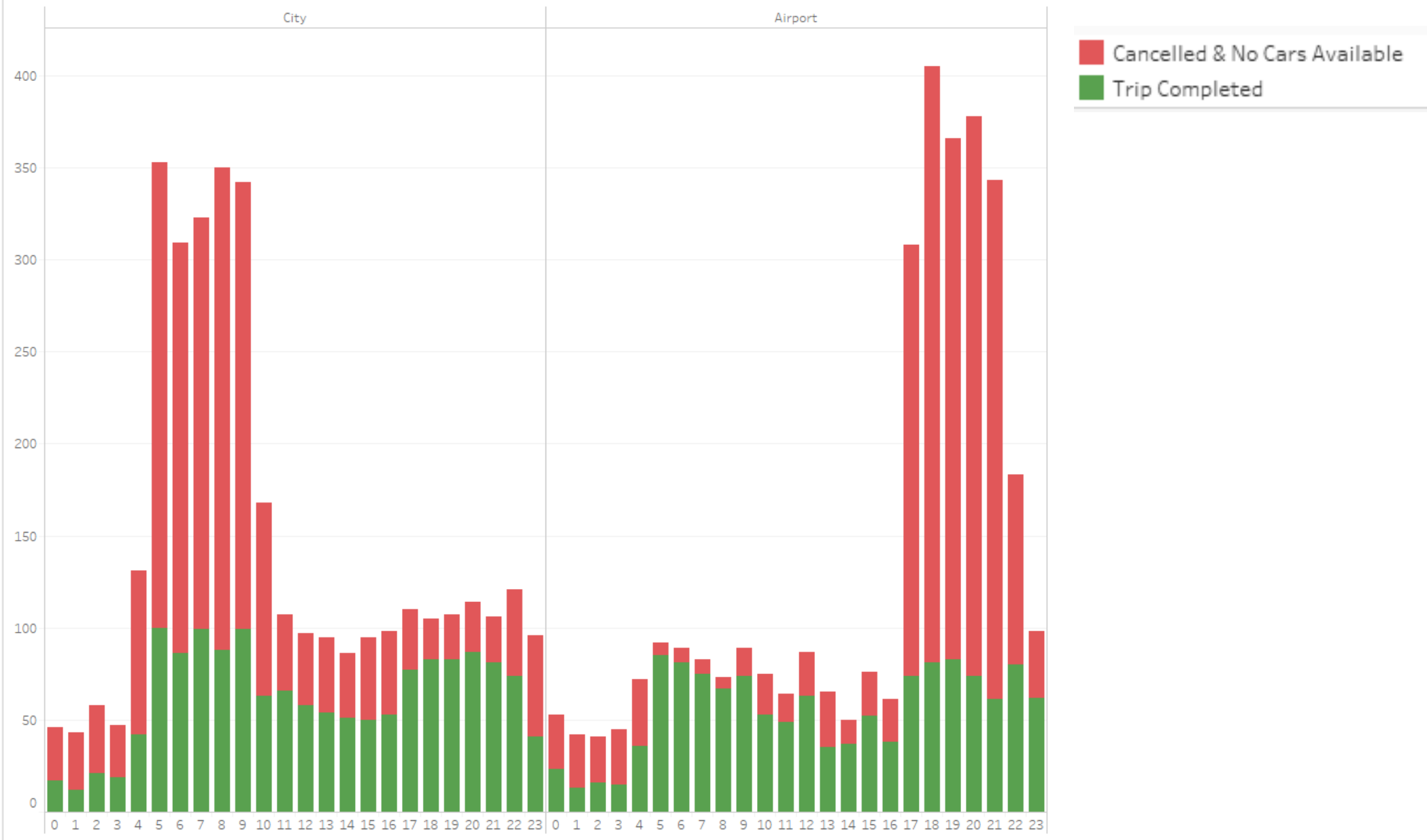
Cancellations at City

1. As visualized in plot-4, there is a sharp increase in cancellations from 04:00AM to 10:00AM

2. The trips completed during peak hours are highest for the day

3. The demand at airport for cabs during peak hours at city is moderate

4. An average trip from city to airport takes 52 minutes.

5. If a driver accepts a trip, there are strong chances that he might have to wait for a longer time to find another trip

6. This is the reason drivers are cancelling trips for a better business within the city.

No cars at Airport

1. As visualized in plot-4, there is a sharp increase in 'No cabs' from 05:00PM to 10:00PM

2. The demand at the city for cabs to the airport during these hours is very low. That's why no cabs are available.

Recommendations

The root cause of the problem is the demand being unidirectional during peak hours. Some of the possible solutions are:

Cancellations in city:

1. The rides taken during peak hours in city can be incentivized to encourage less cancellations

2. Based on available data, new cabs can be added to Uber services and that would ease the problem

3. Drivers can be educated about these issues.

No cabs available at airport:

1. The 'no cabs' problem can be solved by incentivizing the trips taken from airport to city.

2. An algorithm which doesn't assign new 'within city' rides to the cabs near airport during peak hours but balances the city demand as well can be coded to meet the airport demand.

3. Collaborations with the offices near airport to maintain a regular flow of cabs in the area throughout the day.