UBER SUPPLY - DEMAND GAP ASSIGNMENT

SUBMISSION

By –

Mr. Mohit Singh

APFE18803194

PGDDS June 2018

Business Objectives

- While travelling to and from the airport, Customers are facing the problems, either by the driver or non-availability of cars. These very issues are impacting the business of Uber and leading to loss of potential revenue.
- Analysis of the given data, to identify the root cause's of the problem (i.e. cancellation and non-availability of cars).
- The analyses and the recommendations using plots for the root cause(s) of the problem(s) .
- Find out the gap between supply and demand and show the same using plots.Find the time slots when the highest gap exists.
- Find the types of requests (city-airport or airport-city) for which the gap is the most severe in the identified time slots.
- Find possible hypotheses of the problem(s) .
- Recommend ways to improve the situation from City to Airport and vice versa.

Problem Solving Methodology

- Data Sourcing
- Data Cleaning
- Understanding data
- Derived metrics
- Visualize the frequency of requests that get cancelled or show 'no cars available' and identify the most problematic types of requests (city to airport / airport to city etc.) and the time slots (early mornings, late evenings etc.) using plots from data visualization.
- Find out the gap between supply and demand and show the same using plots.
- Find the time slots when the highest gap exists
- Find the types of requests (city-airport or airport-city) for which the gap is the most severe in the identified time slots
- The root cause(s) and possible hypotheses of the problem(s)
- Recommend some ways to resolve the supply-demand gap.

Data Cleaning

NA Values and Duplicates:

1. Na value analysis : 2650 NA’s are in Driver.id and 3914 NA's Drop.timestamp, adding to total 6564 NA's in uber_data.

2. No Duplicate values in uber_data.

Data Standardization issues:

1. Pickup.point and status are converted to factor data type, which can be helpful for data analysis and creating plots.

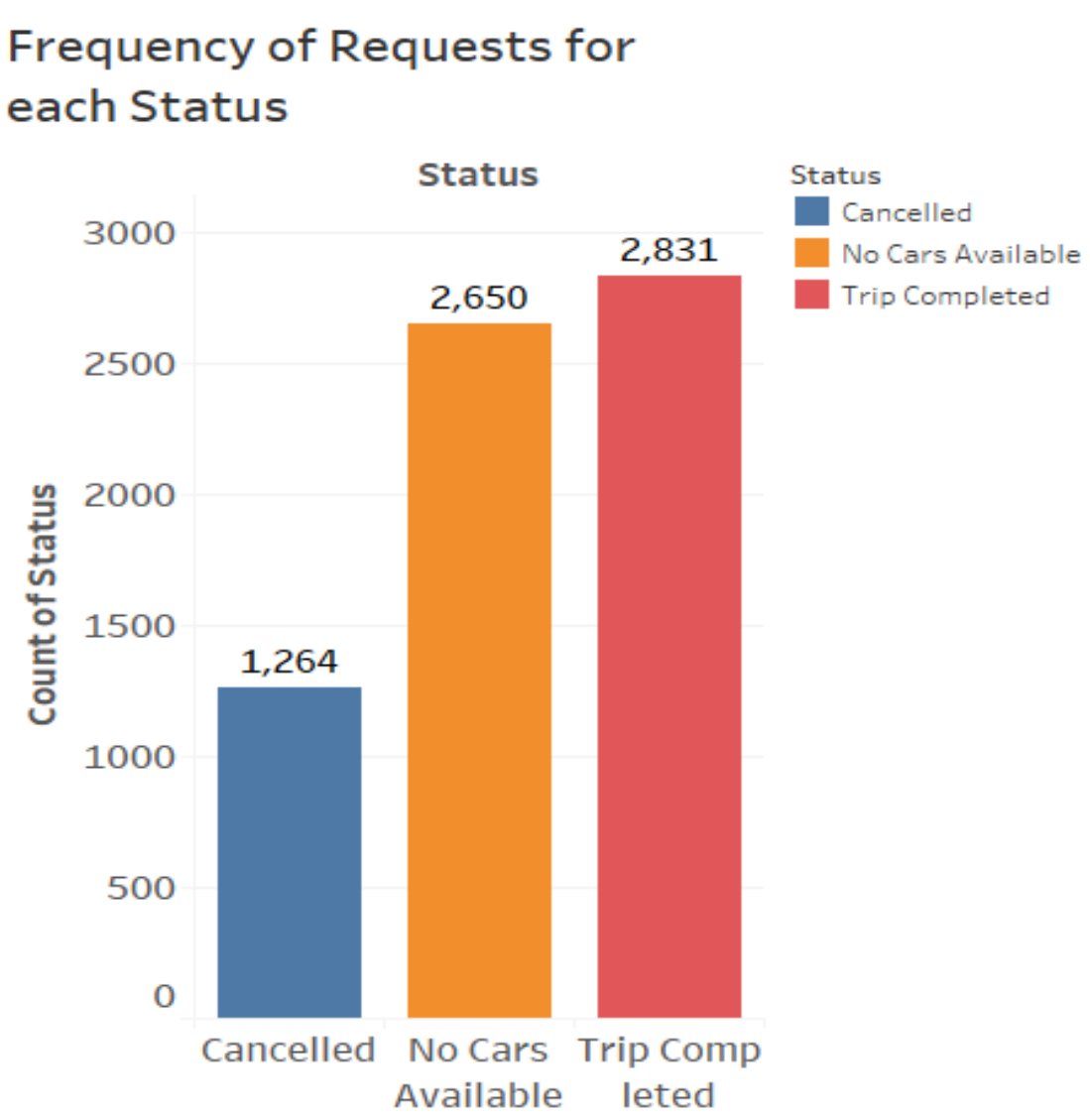
2. Dates are separated by “/” and “-” in date-time column and are replaced by “-” for uniformity of data, which can be helpful for data analysis.Then, These columns are converted to date format.

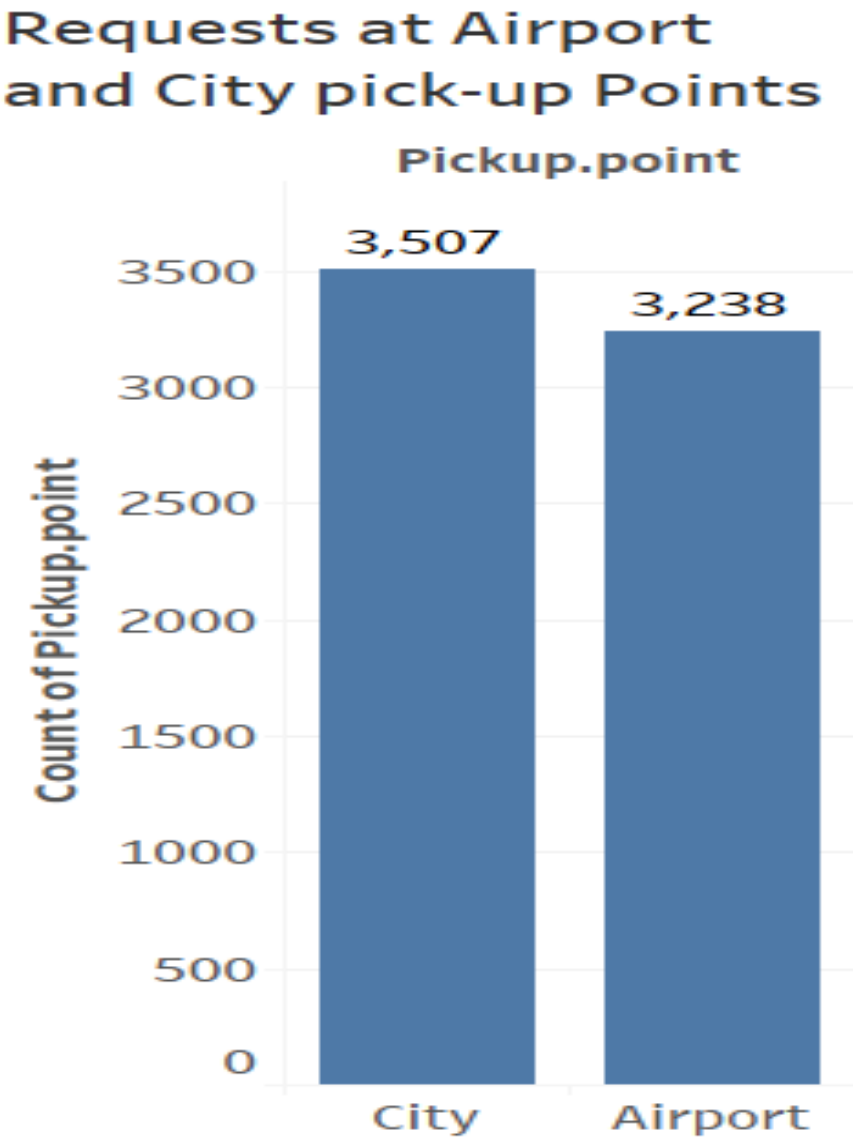
Data Understanding - Uber Request Data (Provided by the Client)

In the provided Uber data there are six attributes associated with each request made by a customer :

- Request id: A unique identifier of the request
- Time of request: The date and time at which 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
- Only the trips to and from the airport are being considered.
- 5 days Uber data is provided from 11-7-2016 to 15-7-2016.
- 6745 unique request's are made by customers in the given duration, TWO columns are of Integer type and FOUR columns are of Character type.
- Data provided is of 24 hours of each given days.
- NA's in Driver ID shows No cars are available at the location of request made.
- Driver ID is present for "cancelled " status, which shows that requests are cancelled by the drivers present at the location of request.

Analysis of Request's at Pick-up Point's and Status





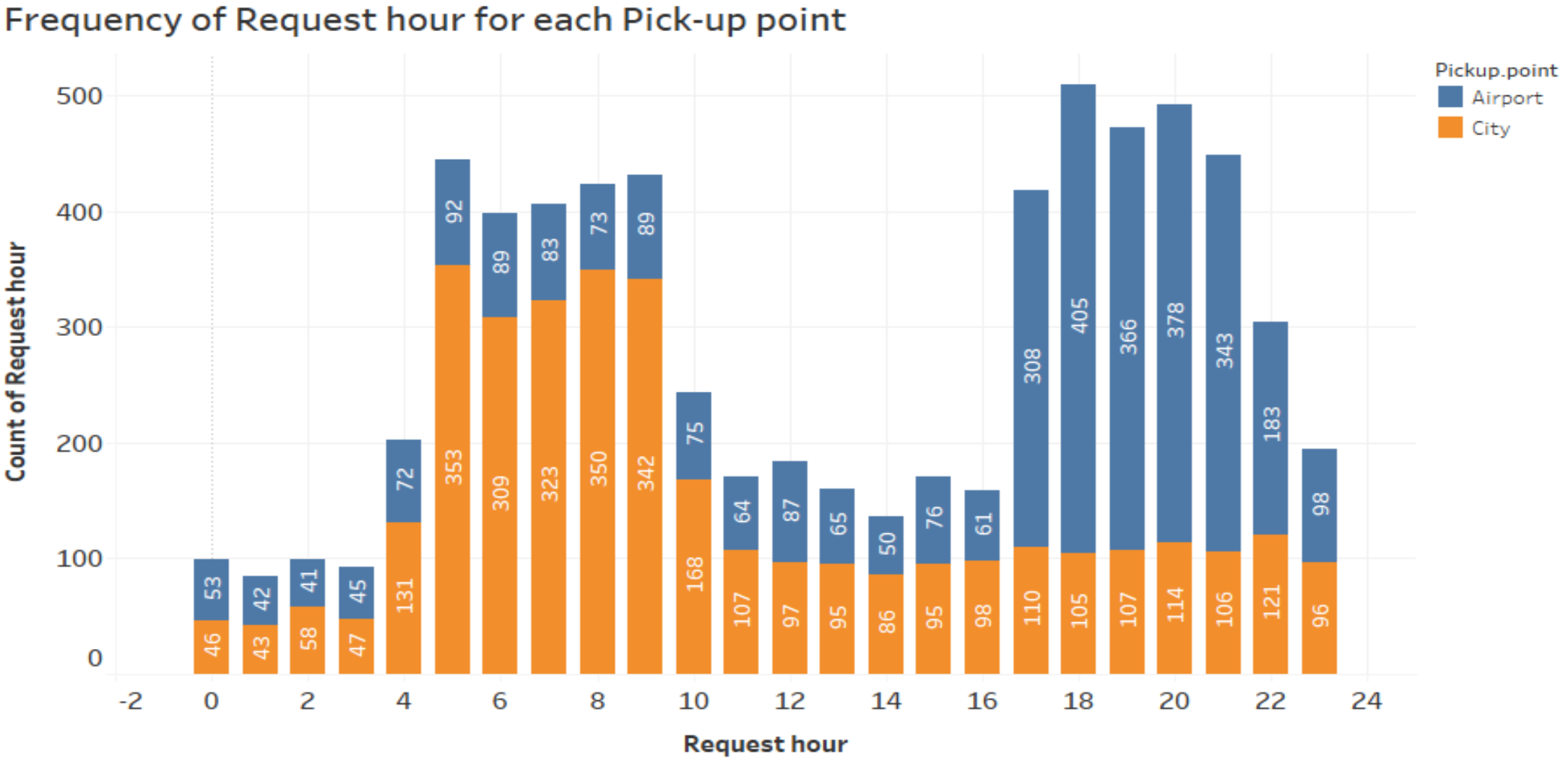
Derived Matrix

- Date & Hour are separated from the given Request and Drop date-time columns Request_dt, Request_hour and drop_dt, drop_hour are added to uber_data.

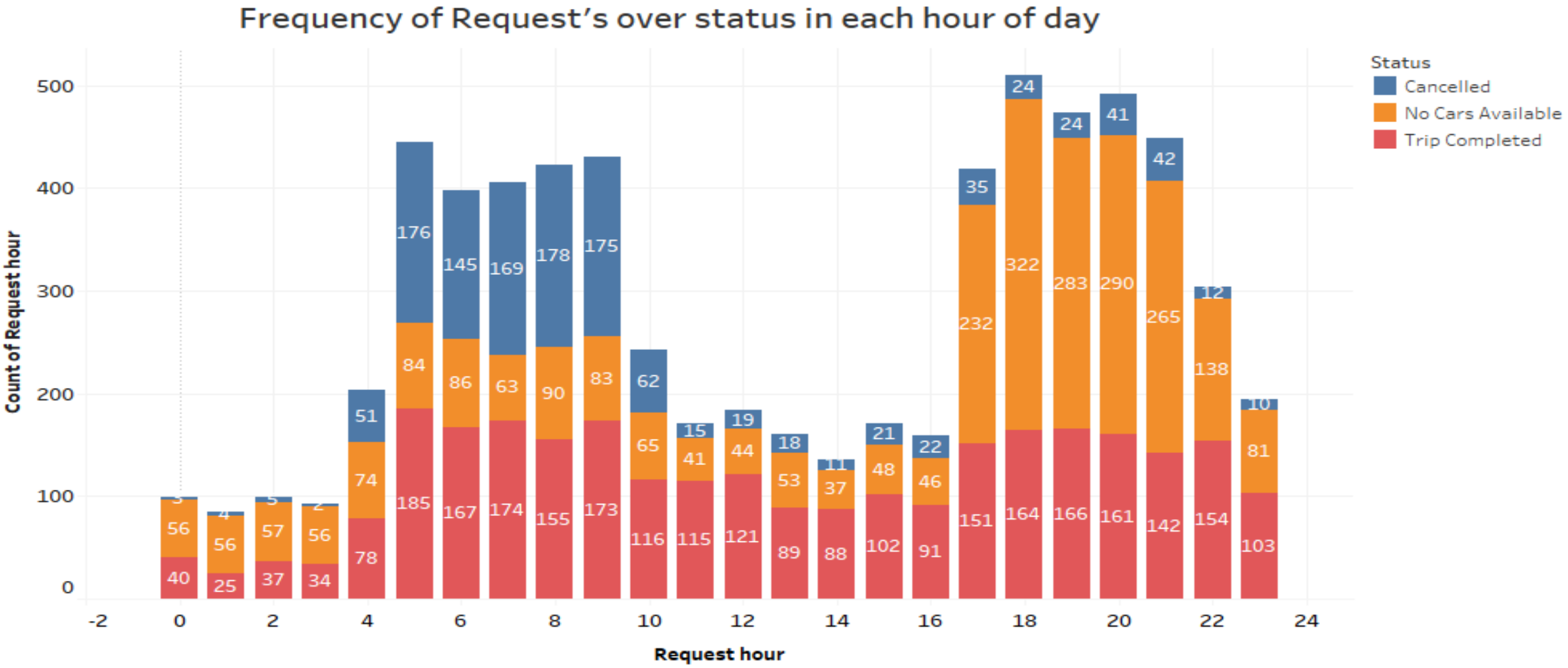
- Columns Supply and Duration (in minutes) is added to data frame :

1. Supply : “Supply available” is taken for cancelled & completed requests and “ No Supply” is taken for non availability of cars. Because the cars are available at the time of request
2. Duration (in minutes) : Time taken between City and airport.

Identification of Most Pressing Problems for Uber

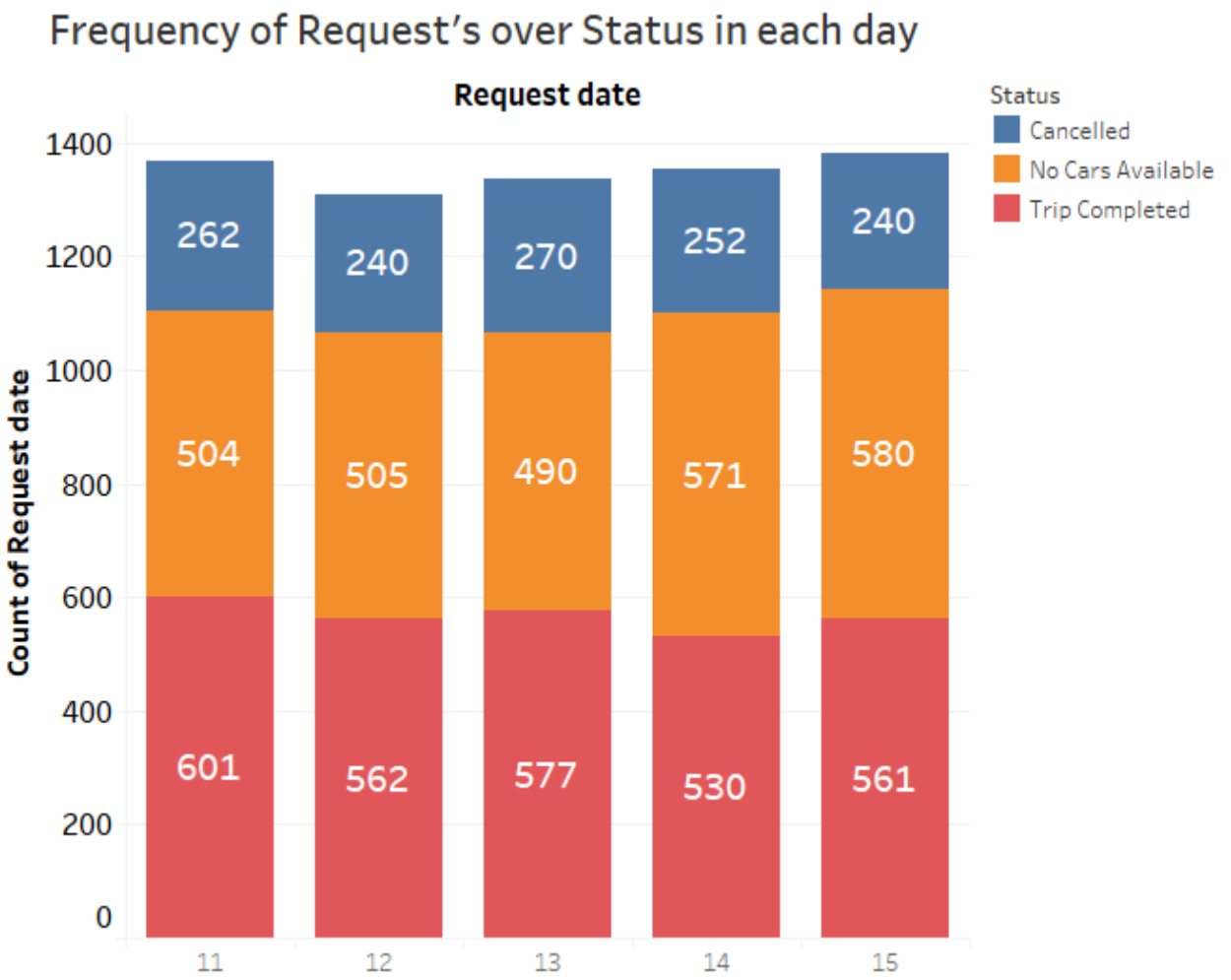


It can be seen that demand is more in City in the Morning hours and demand is more at Airport in from Evening till late night hours.

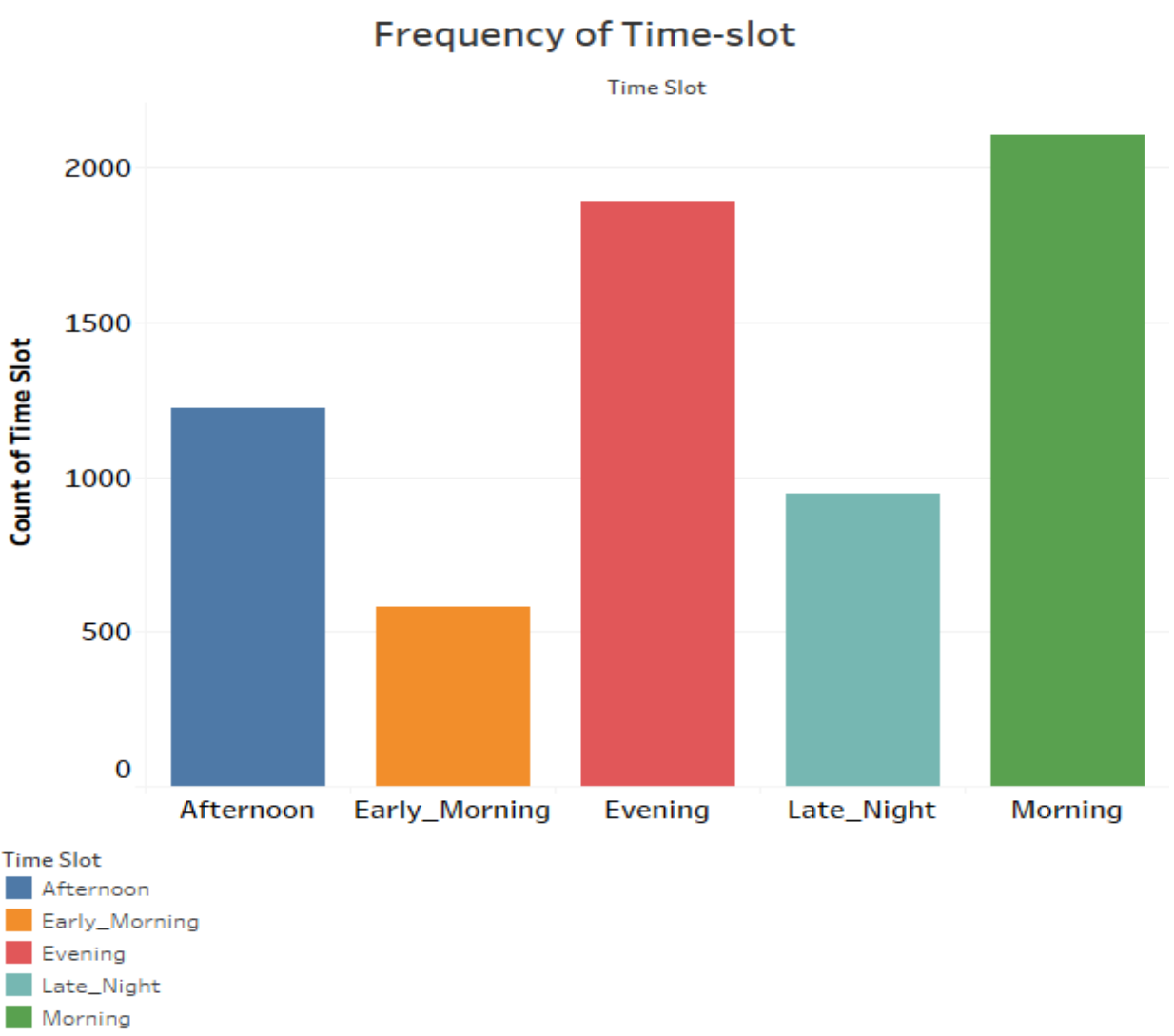


There is more distribution of "No Cars Available" as compared to "Cancelled" cars, More number of "Cancelled" cars are seen only in the early morning hours as compared to other hours and problem of "No Cars Available" is more in Morning hours & from Evening till late Night hours.

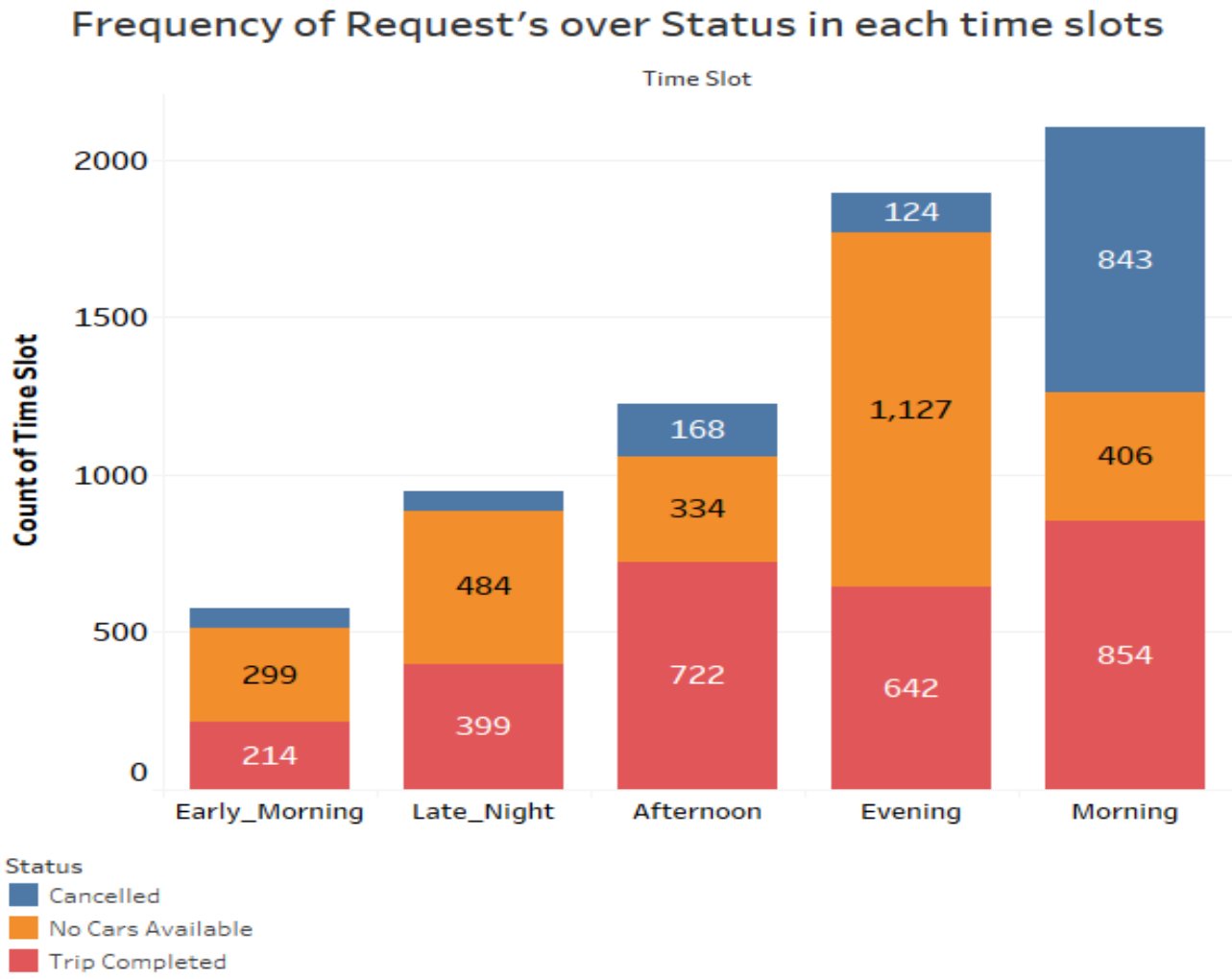
Major time blocks can be noticed based on variations in requests made by customer in 24 hours and the distribution date-wise showed less significant variation on plot, so hours are used for final analysis.



Throughout the days their respective distribution remained almost the same Problem of "No Cars Available" is more reflected in all 5 days as compared to "Cancelled" cars, But this days-wise plot is not giving any strong evidence of any raised problem to be taken out of further analysis.

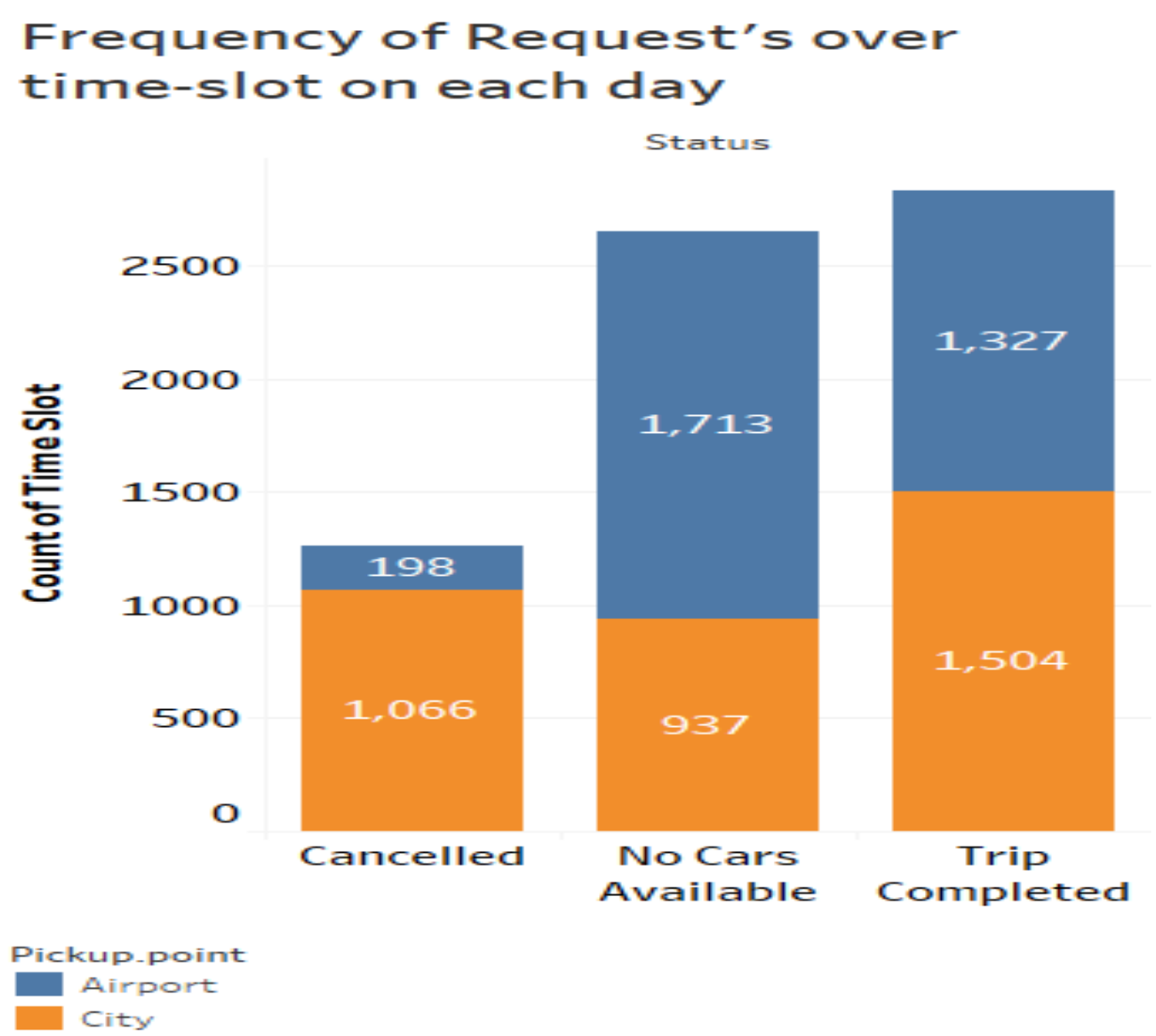


From respective counts of time slots, It is evident that more requests are made during Morning hours, than during Evening hours and lesser for other time slots.



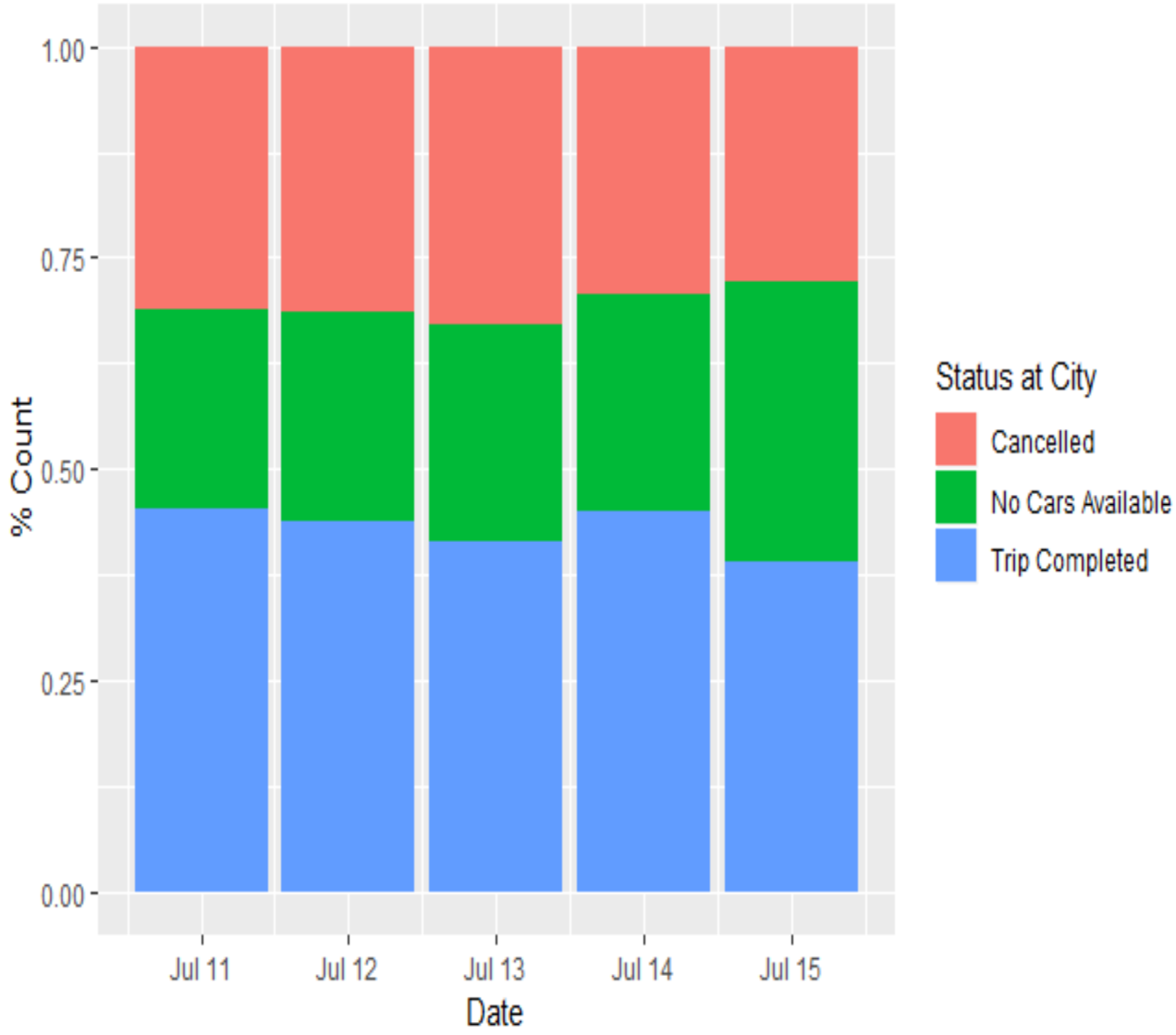
Among all time-slots "Cancelled" issue is more in morning hours.

Among all time slots "No cars available" issue is more in evening hours.

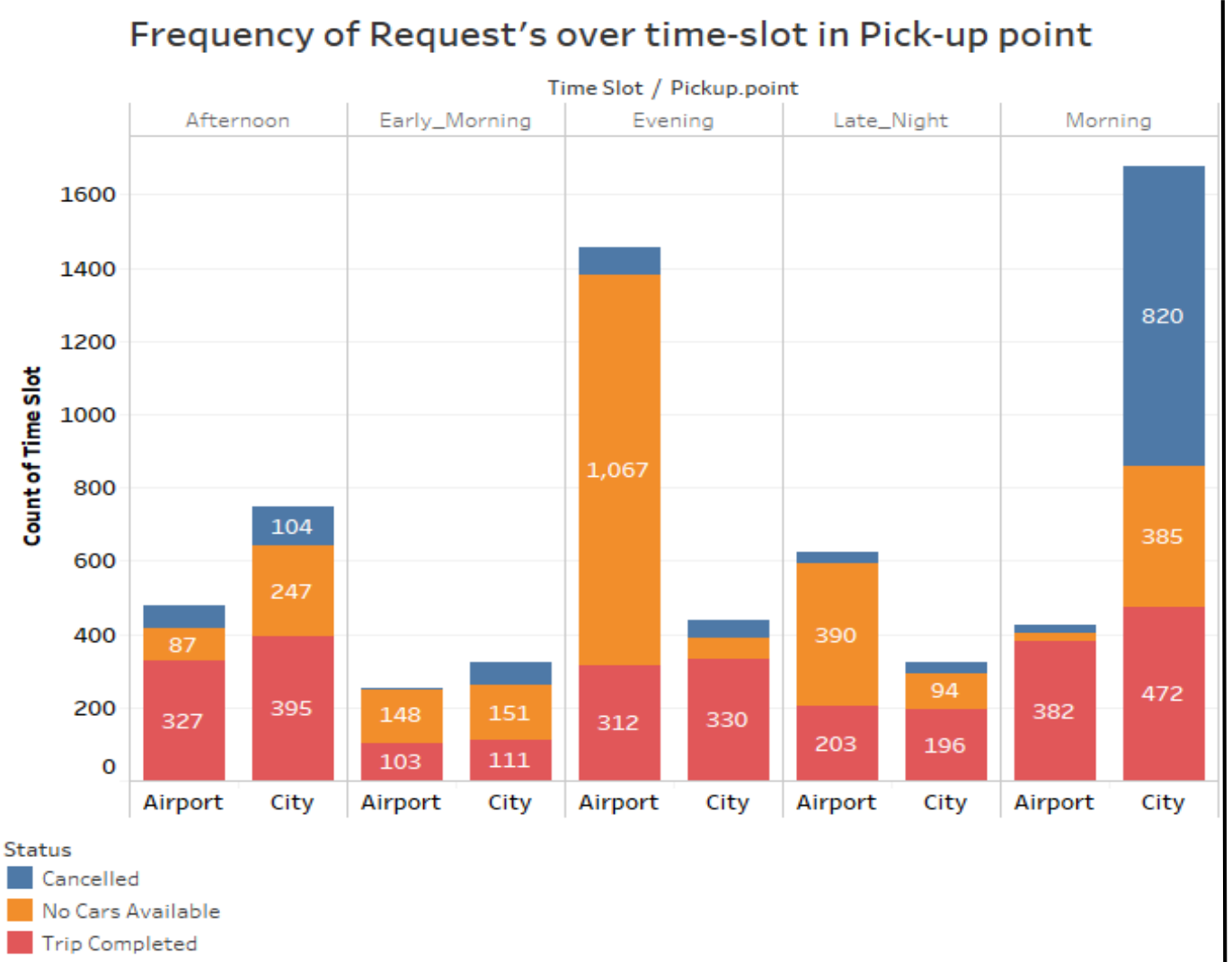


Plot shows Both "Cancelled" and "No Car Available" problems of cars are significantly high at City.

Plot shows that "No Car Available" problems of cars are very high at Airport.

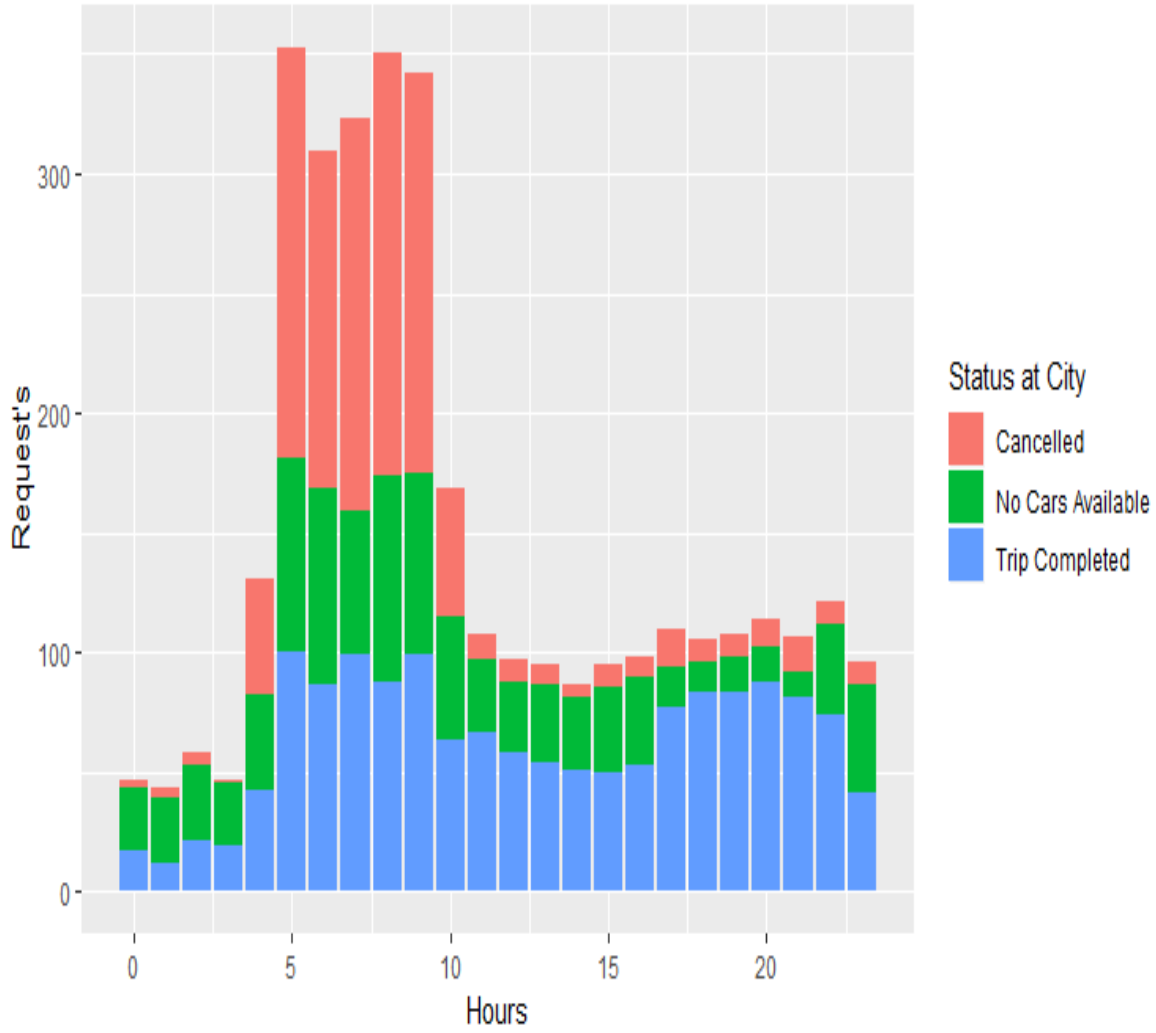


There is no significant variations in distribution among days, which cannot show evidence of problem when checked on each day



In the Morning hours "Cancelled" and "No Car Available" cars problems are very high as compared to other time slots

Hence, We can take Morning hours for further analysis.

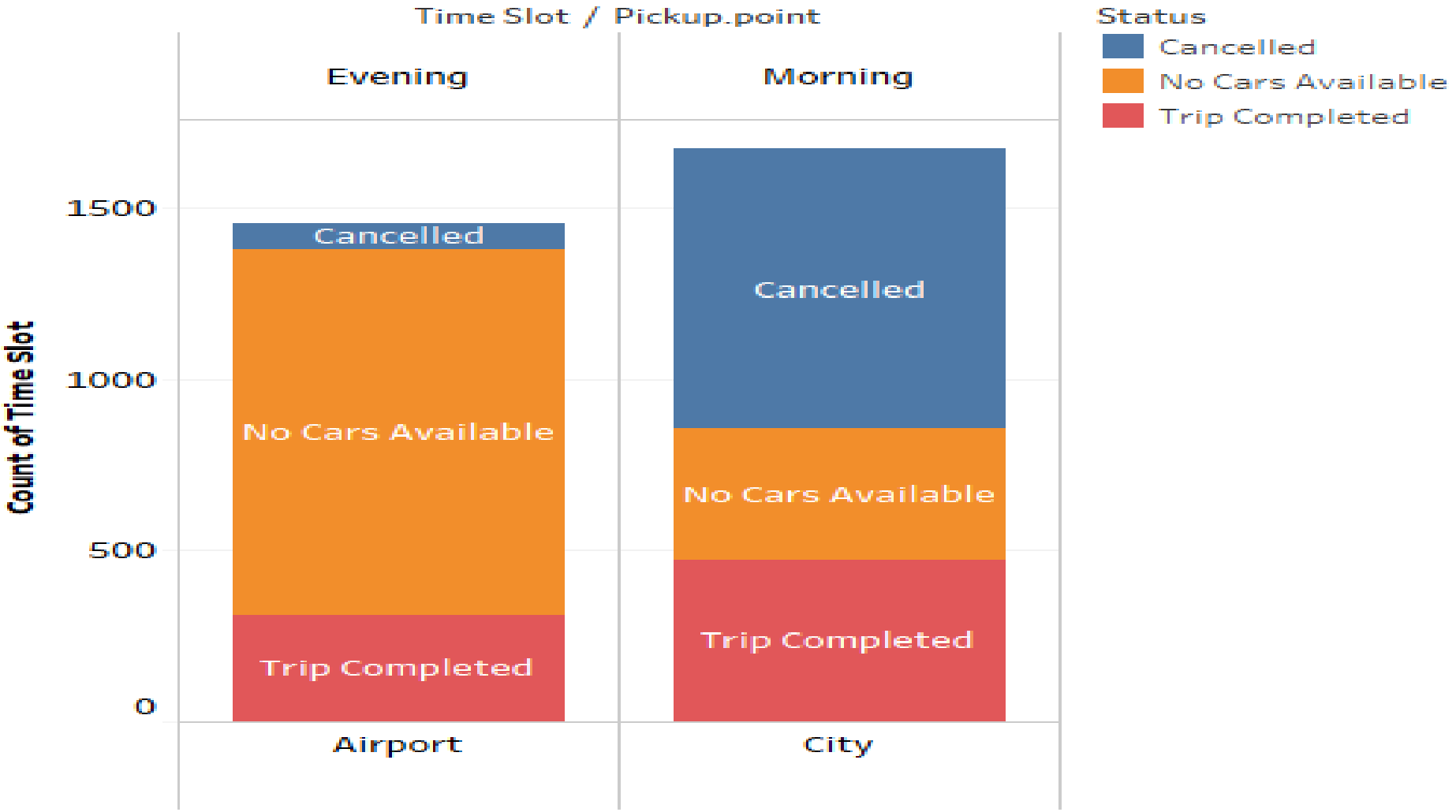


In the Morning hours from 5am to 9am,

"Cancelled" and "No Car Available" problems are

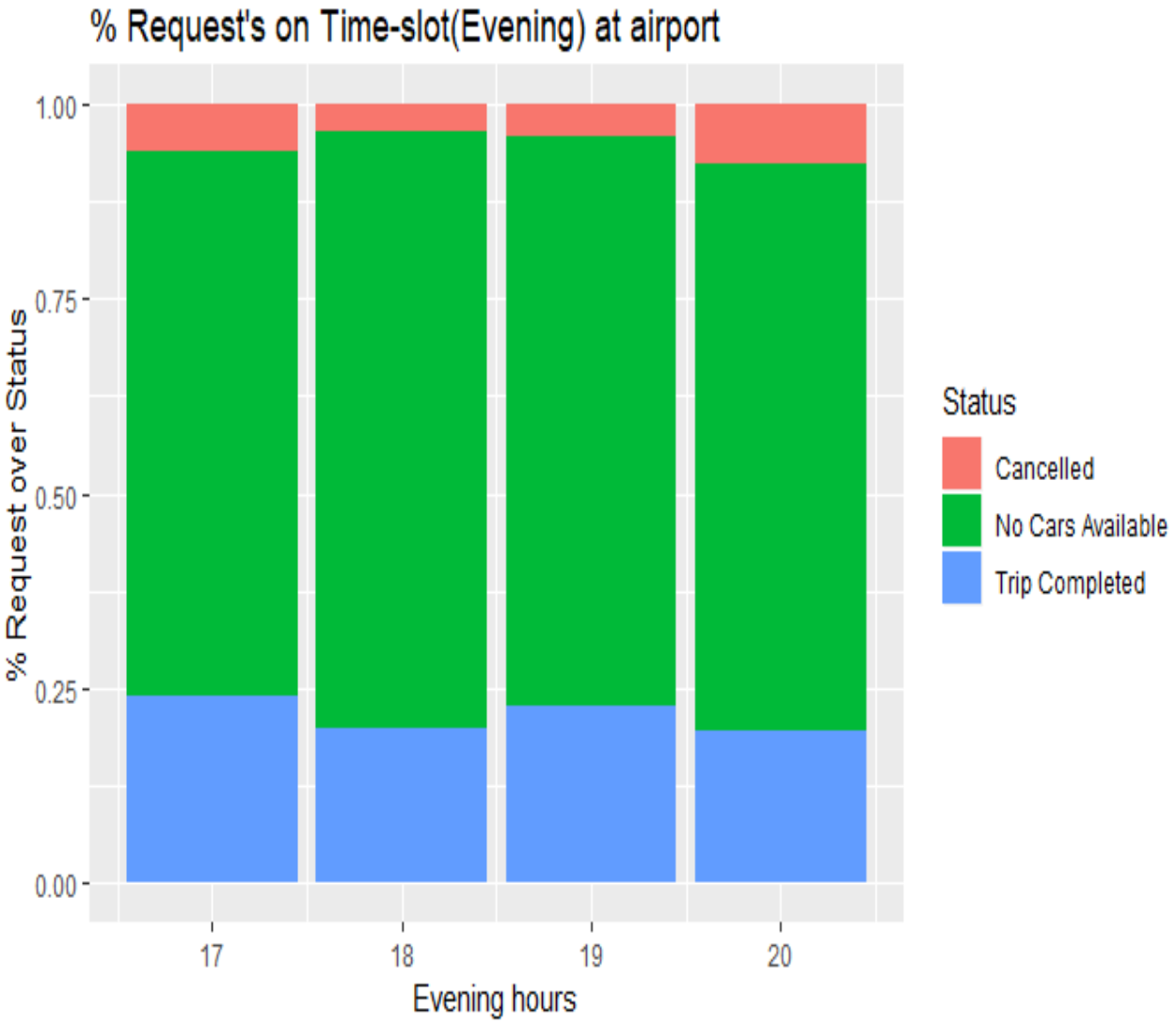
very high as compared to other time of the day.

Frequency of Request's over Status in each time slots

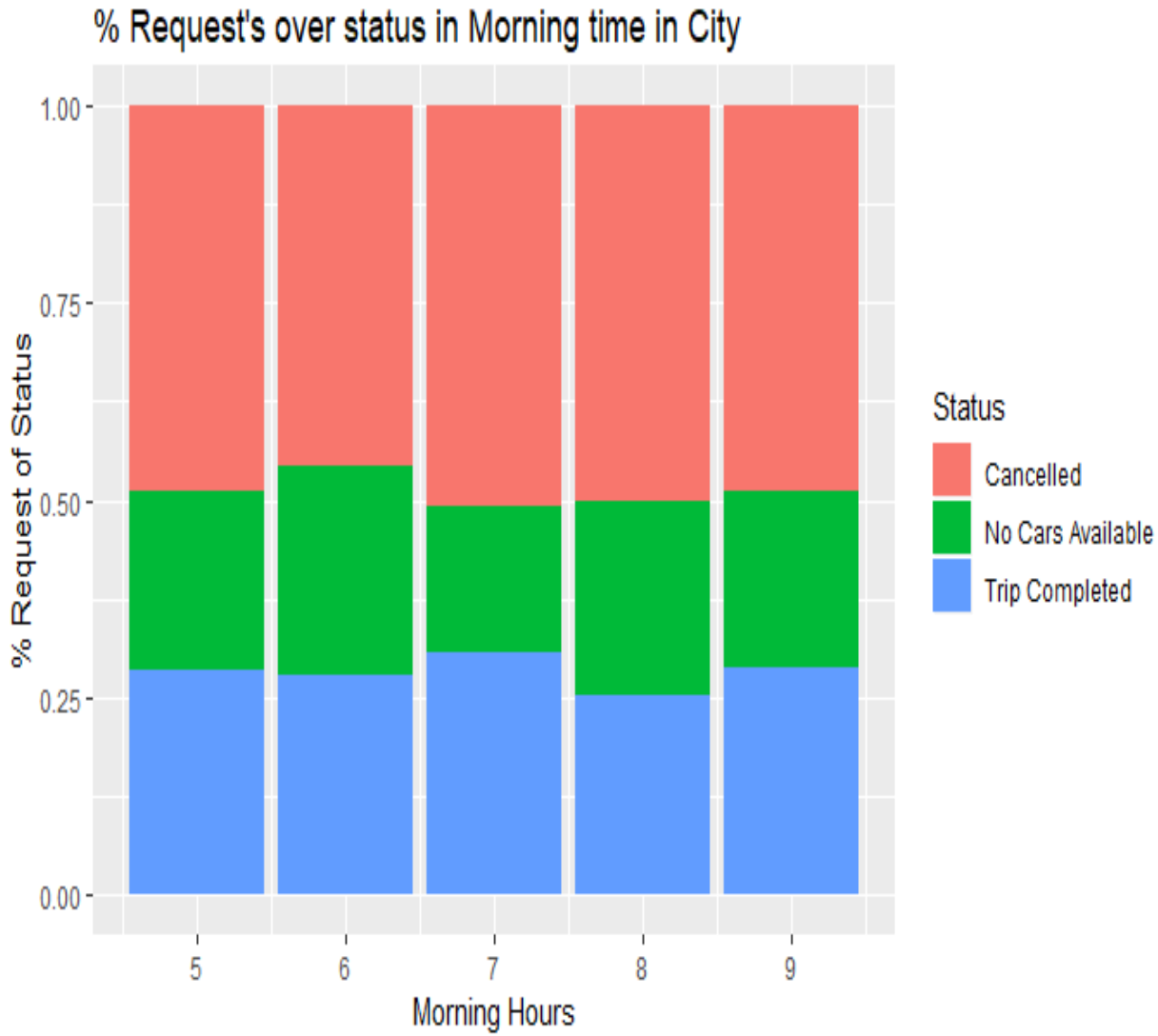


In Morning time-slot "Cancelled" cars problems are very high at City .

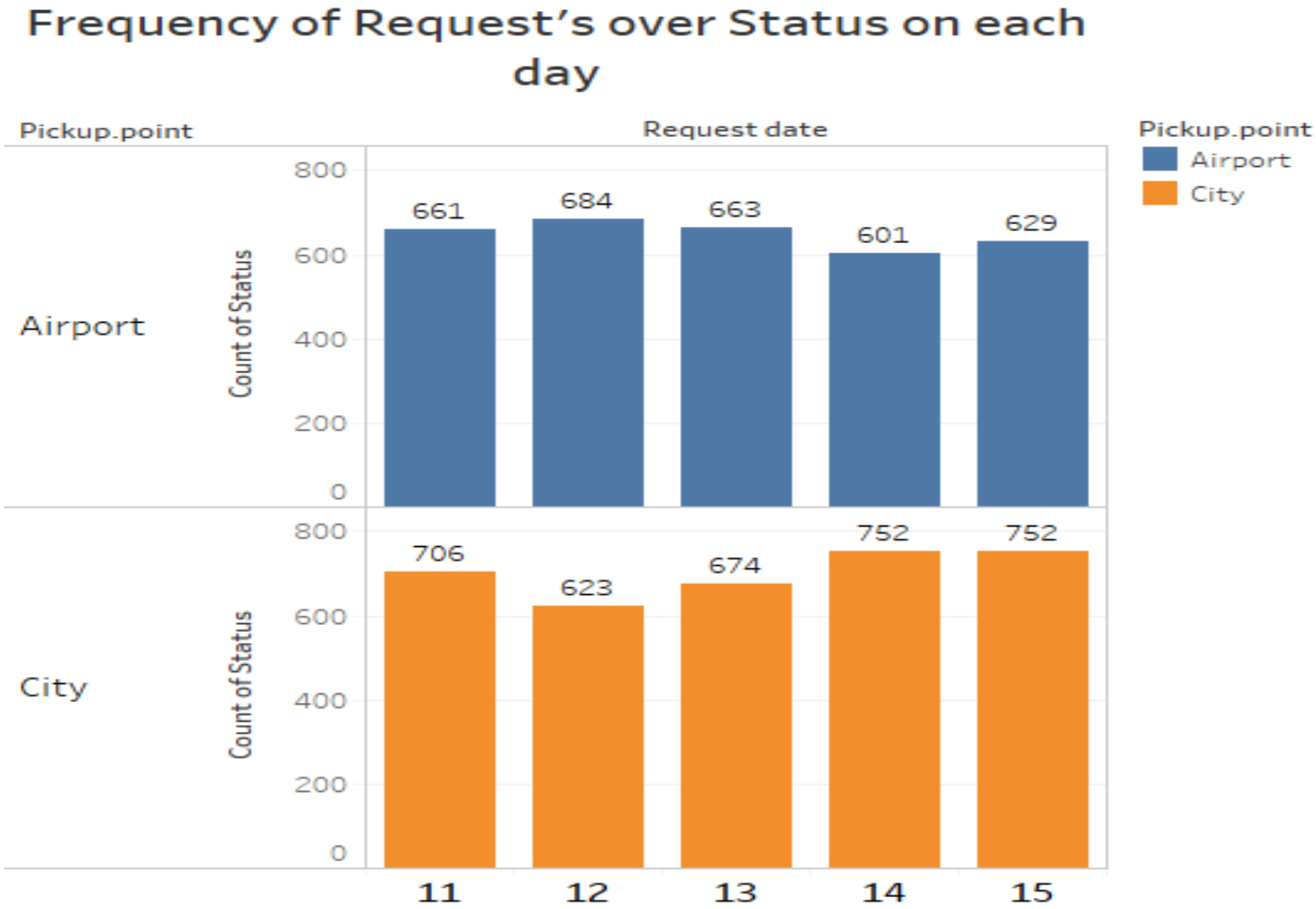
In Evening time-slot " No Cars Available" problems are very high at Airport.



Throughout Evening time-slot "No Car Available" problems are very high and "Cancelled" cars issue is very low, Drivers are cancelling very less number of requests.



The bar graph shows the peak of the problem is found around 5pm, which is of only slight variation than other morning hours, because of less variation among hours, we cant take it for further analysis on hour basis.



For both City and Airport, the bar graph shows only slight variation in day-wise, because of less variation among days, we can't take it for further analysis on day basis.

TWO MOST PRESSING PROBLEMS FOR UBER

- Based on the above analysis of issues, there are 2 most pressing problems for Uber as mentioned below :

Problem 1:

- Most number of requests are made in morning hours for City to Airport and most issue's are occuring in Morning hours itself,
- out of 2103 total morning Requests only 854 were completed, which is around 40% of the total Morning requests,
- 1249 requests were not completed by the Drivers, which are around 60 % of the total morning requests.
- Cancelled cars are high as 843, which are around 40 % of the total morning requests.

Problem 2:

- Most number of requests are made in evening hours for Airport to City and most issue's are occuring in Evening hours itself,
- out of 1893 total evening Requests only 642 were completed, which is only around 34% of the total Evening requests
- 1251 requests were not completed by the Drivers, which are around 66 % of the total morning requests, dominated by "No Cars Available" issue.

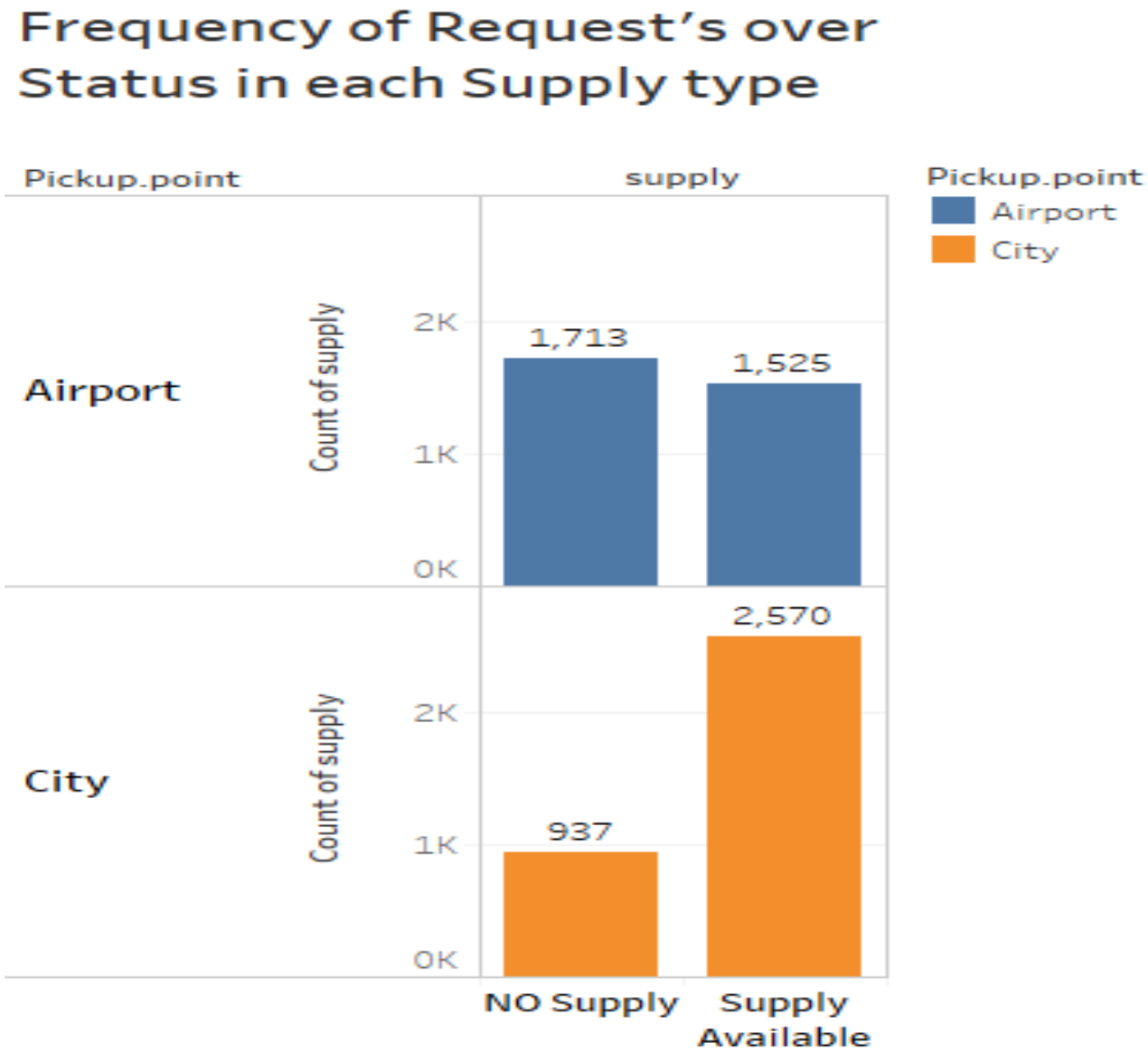
Supply and Demand Gap Identification

Analysis showed there are evidence of huge Supply and Demand gap between what is required & is available at the city and airport.

Assumptions :

- For Demand at CITY & AIRPORT = Trip Completed, Cancelled, No Cabs Available are considered.
- For Supply at CITY & AIRPORT = Trip Completed, Cancelled are considered.
- Reasons : Cancelled are also considered in supply because requests were made and also cars are available there, it's driver who cancelled, Because of which uber will face revenue loss.

Supply and Demand Gap Identification



- Hence, it is clear that in morning Out of 1677 requests made at City, Inspite of having supply of 1292, which is around 77 percent of request's at city in morning hours, Only 472 trips are actually getting completed, which is around 28 percent of request's at city in morning hours

- In evening out of 1457 requests made at airport, only supply of 390 cars was available, which is around 27 percent of request's at airport in evening hours, Only 312 trips are actually getting completed, which is around 21 percent of request's at airport in evening hours.

Supply and Demand Gap Hypothesis

Average distance covered between city and Airport During Morning and Evening hours is same as throughout time slots, so time taken to reach the destination would not be the reason for cancellation.

1. City to Airport (Morning Hours) : Issue of Cancellation of cars is found more, possibly because of more requests from other locations that are nearby or those locations where after reaching there will be less wait time as compared to Airport after morning hours.

2. Airport to City (Evening Hours) : Issue of Non-availability of cars is found more, possibly because of less number of drivers reaching airport before evening hours. Reason for less flow of cars before evening hours will be long wait time till time of flights in evening hours. So, Drivers will be more routed towards less wait time locations. There is no point in reaching to a location that is around 1 hour distance and wait for long hours, or may be incase return without customer.

Supply and Demand Gap Recommendations

- In order to prevent the drivers from cancellation of requests, as in case of city morning hours status, compensation of time in terms of some % of their monthly salary can be provided. Which will reduce cancellations and also help customers on reaching to their destinations on time.
- For Such situations where uber customer's face more problems what analyzed, some special cabs can be fixed to meet the purpose and their pay to dealt accordingly, as they will be spending long hours idle at destination locations.
- Such Problematic hours of the day can be fixed for some special offers for drivers to motivate them to serve the purpose and meet more customer satisfaction.
- Drivers can be paid for their fuel, in case they completing more requests in such Problematic hours and returning back from destination without any customer.
- We can pool the rides of customers, so that less number of cars present at the locations in such needy hours can prove helpful.