Demand-Supply Gap in Uber
A Case Study

Presented By:

Ujwal Kiran Bhargava

Introduction, Background and Problem Statement

- Uber is a ridesharing and taxi cab services company headquartered in San Francisco, California.
- It has operations in 633 cities worldwide.
- Apart from cab rentals and transportation services, it is also involved in food delivery services.
- However, due to cancellation and non-availability of cars, Uber tends to face a loss of revenue every day.
- The main objective of the case study is to find and analyse the root cause of the problem (non-availability and cancellation of cars) and to recommend ways to rectify/improve the situation.
- However, the constraint in this case study is that only trips to and from the airport are considered.

Analysing and Understanding Data

There are six attributes associated with each customer request:

- Request ID: Unique identifier for each customer request
- Pickup point: The source at which the request was made
- Driver ID: Unique identification number for each driver
- Status: Indicates whether cabs not available, ride cancelled or trip successfully completed
- Request timestamp: Date and time at which customer request occurred
- Drop timestamp: Date and time at which the trip was completed

Based on the given problem statement, we should focus on Pickup point and Status based on the Request timestamp.

Metadata

The .csv file used in the case study (Uber Request Data.csv) has the following features:

- 6745 rows, 6 columns (Request id, Pickup point, Driver id, Status, Request timestamp, Drop timestamp)
- Details of trip to and from the airport
- Data collected on 5 days (11^{th} to 15^{th} July 2016) and across 24 hrs(00:00 to 23:59)

Data Issues and Cleaning

The main issues in the data in the .csv file are:

- NA values in columns
- Duplicate values of Request ID
- Inconsistent format of date

To make the date format consistent for the ease of analysis, we used strptime() function for both Request and Drop timestamps.

Also, we converted the required data into factor type for ease of analysis.

Assumptions

- Only trips to and from the airport are considered for analysis.
- Derived metrics are derived from the columns based on suitable criteria.
- Demand is defined as :

  Trips completed + No cabs available + Trips cancelled
- Supply is defined as:

  Trips Completed
- Timestamps are divided into 5 timeslots based on the hour of the day:
  - Early Morning (12:00 am – 4:59 am)
  - Peak Morning (5:00 am – 9:59 am)
  - Normal Daytime (10:00 am – 4:59 pm)
  - Peak Evening (5:00 pm – 8:59 pm)
  - Late Night (9:00 pm – 11:59 pm)

Steps Involved

The following steps are involved in this case study:

- Import csv file for reading in R, and load the required library packages.
- Understand, clean and convert data into the required format for ease of analysis.
- Derive new metrics based on existing data.
- Filter data based on Pickup point, Status, and time slot.
- Extrapolate the data, and perform univariate and segmented analysis.
- Plot the results using ggplot() and analyse them.
- Derive conclusions and final recommendations based on the result.

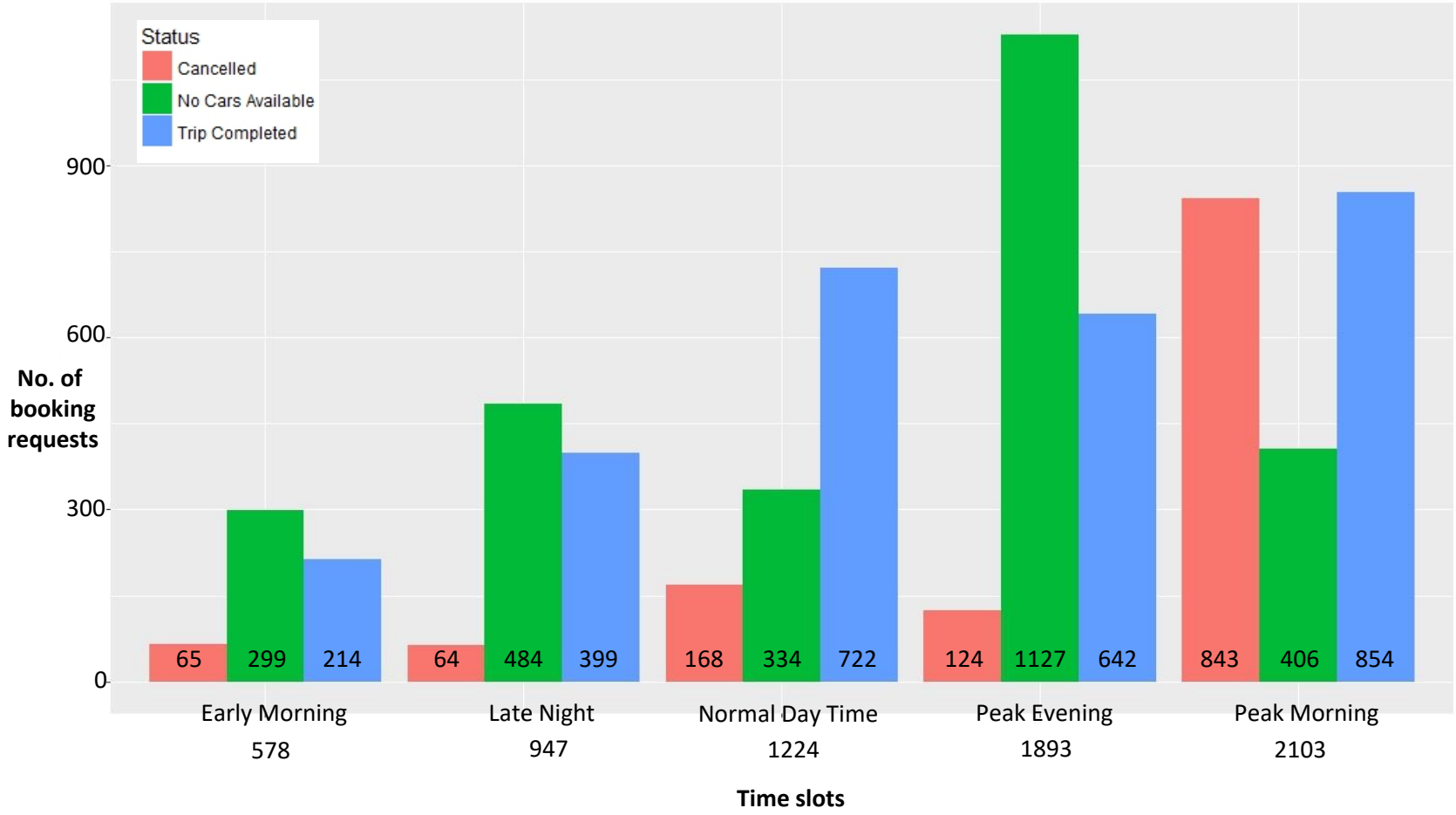
Derived Metrics Used in Graphs and Analysis

The following are the major derived metrics that are used in the case study for plotting the graphs:

- requestHour – derived from Request.timestamp, used to indicate the hour in which the booking was made by the customer
- Timeslots – derived from requestHour, used to indicate the part of day in which the request was made
- Peak – derived from timeslots and Status, used to indicate cancelled bookings from in the peak morning timeslot
- Evening – derived from timeslots and Status, used to indicate non-available car bookings from the airport to the city in the peak evening timeslot
- Airport – derived from Pickup.Point, used to indicate the trips made from the airport to the city
- City – derived from Pickup.Point, used to indicate the trips made from the city to the airport

Analysis 1

Identifying the frequency of cancellations and car non-availability (Segmented Univariate Analysis)

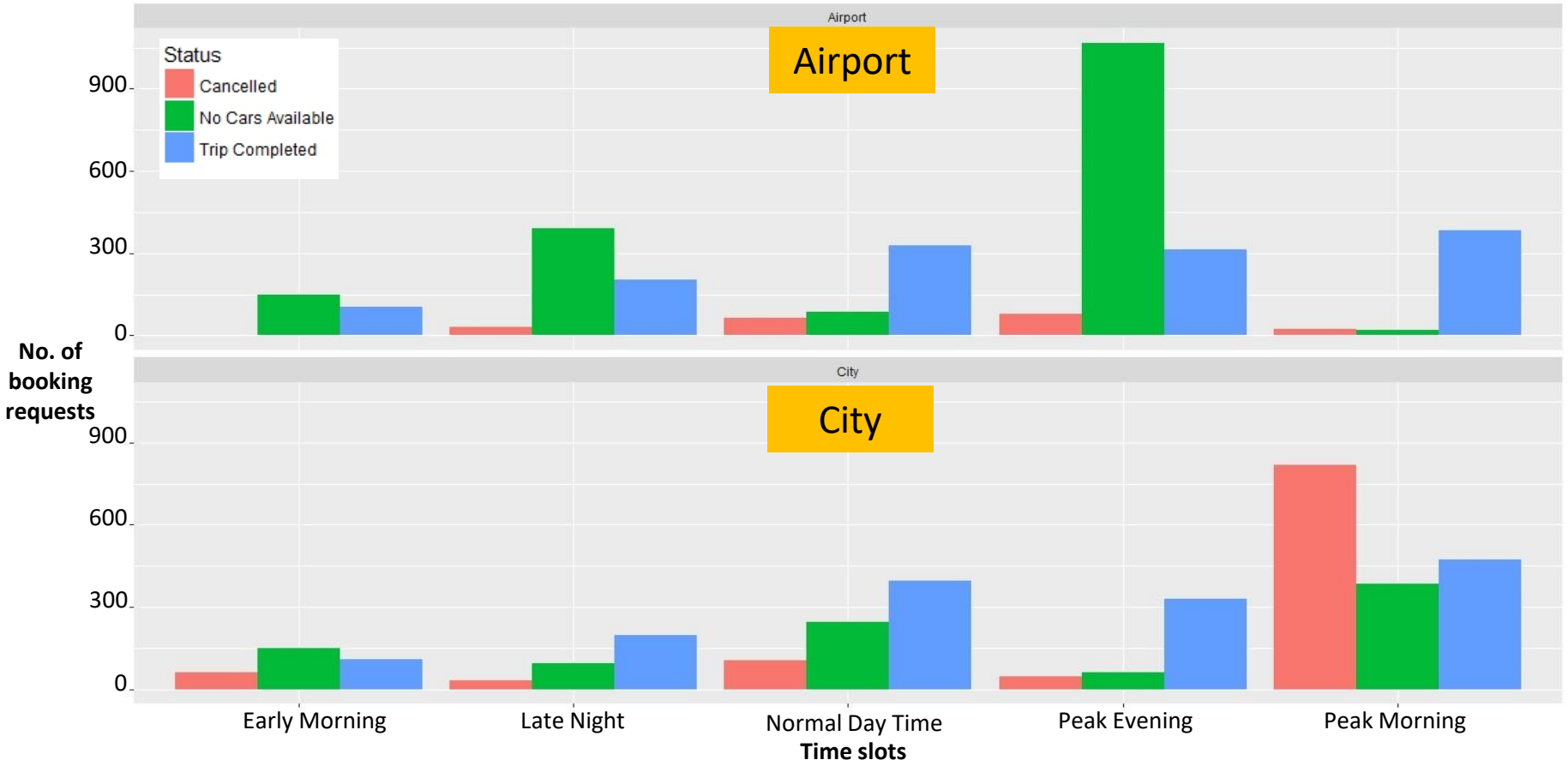


Maximum number of cancellations happens in the peak morning time slot (5:00 am – 9:59 am).

Maximum number of car non-availability happens in the peak evening timeslot (5:00 pm – 8:59 pm).

Analysis 1

Identifying the frequency of cancellations and car non-availability (Univariate Analysis w.r.t. Pickup point)



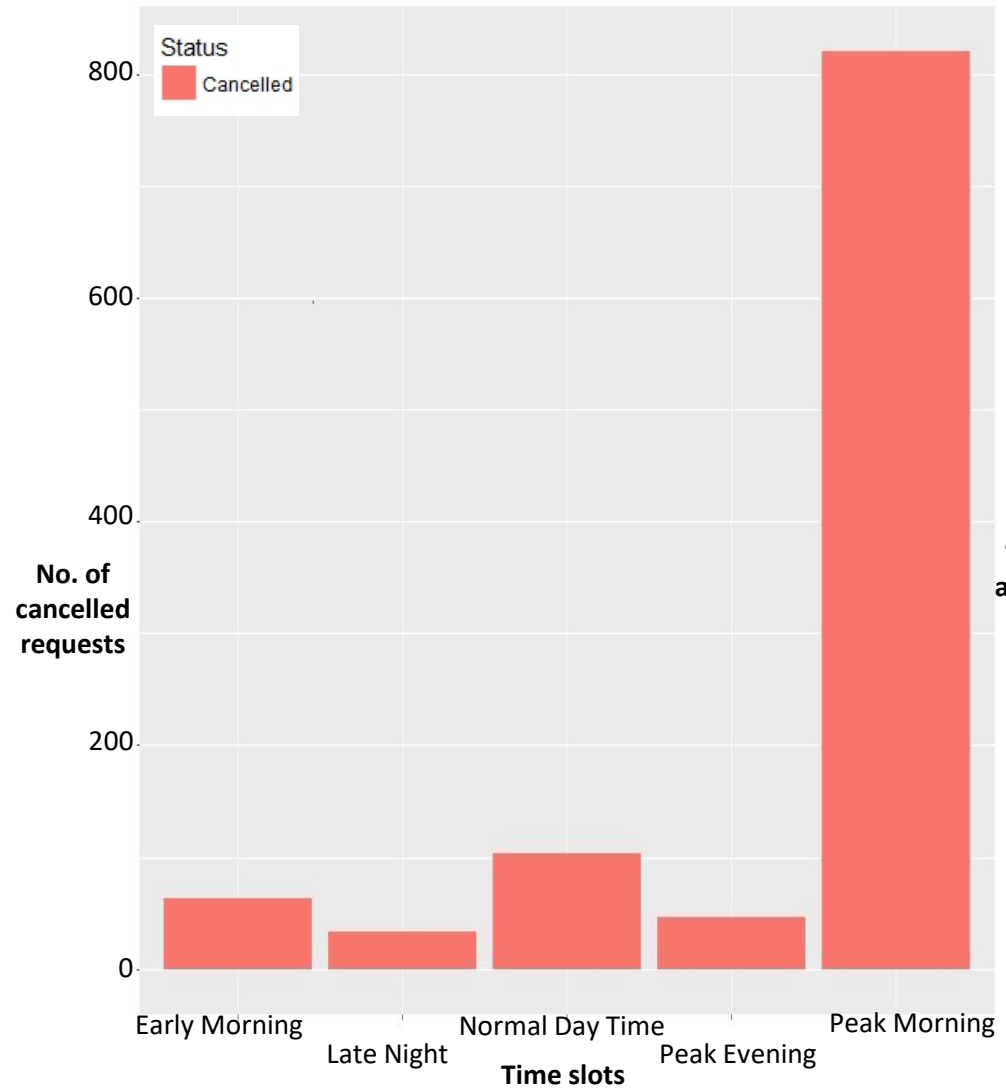
Maximum number of car non-availability happens in the peak evening timeslot (5:00 pm – 8:59 pm) from the airport.

Maximum number of cancellations happens in the peak morning time slot (5:00 am – 9:59 am) from the city.

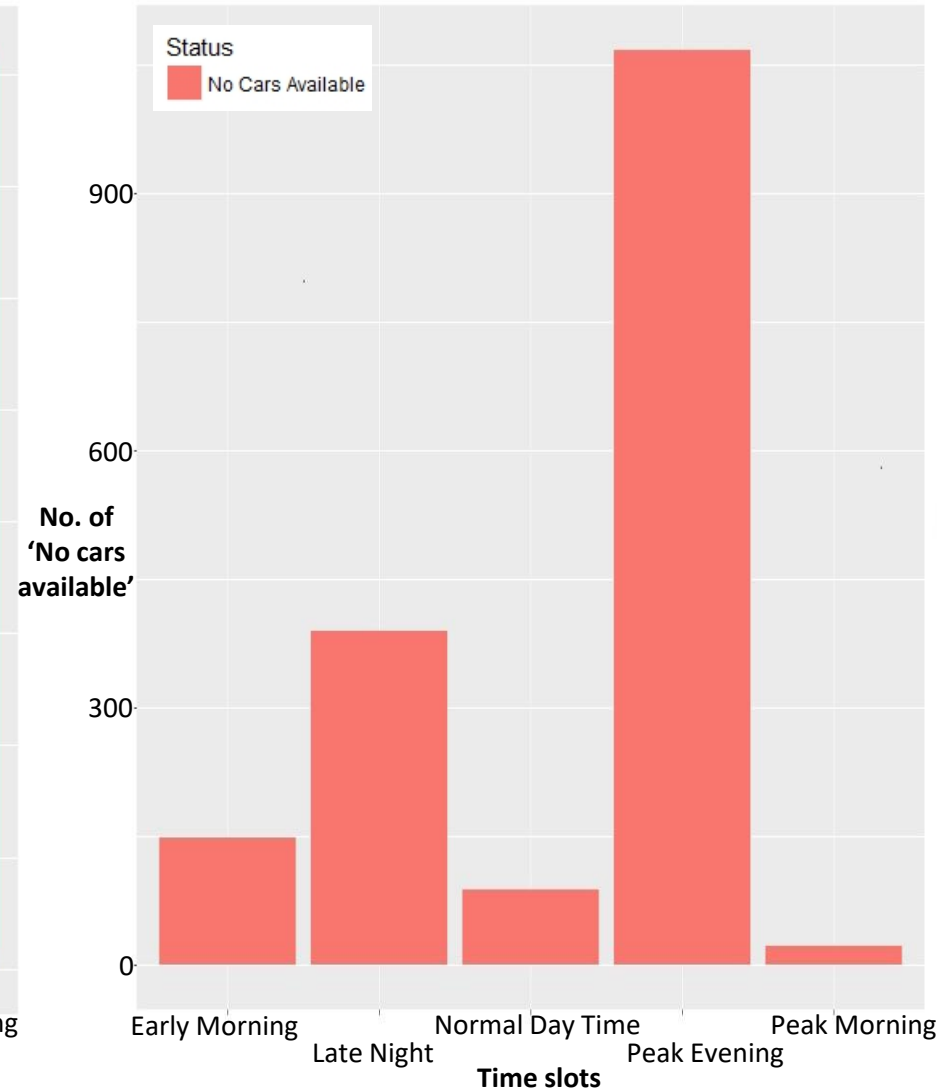
Analysis 1

Identifying maximum cancellations and car non-availability

City



Airport

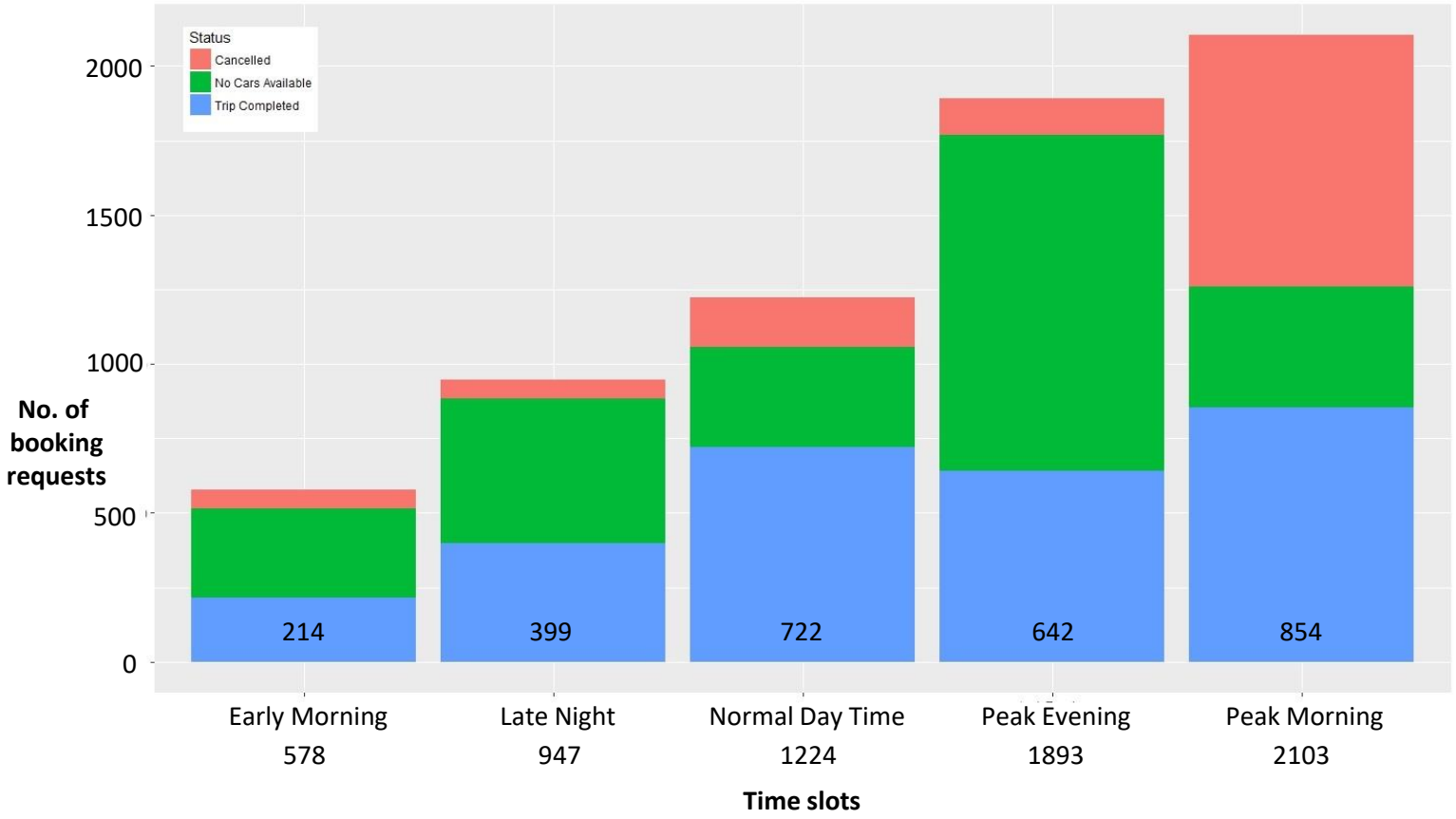


Maximum cancellations (40.09%) occur in the peak morning timeslot (5:00 am – 9:59 am) from the city.

Maximum unavailability of cars (59.54%) occurs in the peak evening timeslot (5:00 pm – 8:59 pm) from the airport.

Analysis 2

Understanding the Demand-Supply Gap (Segmented Univariate Analysis)

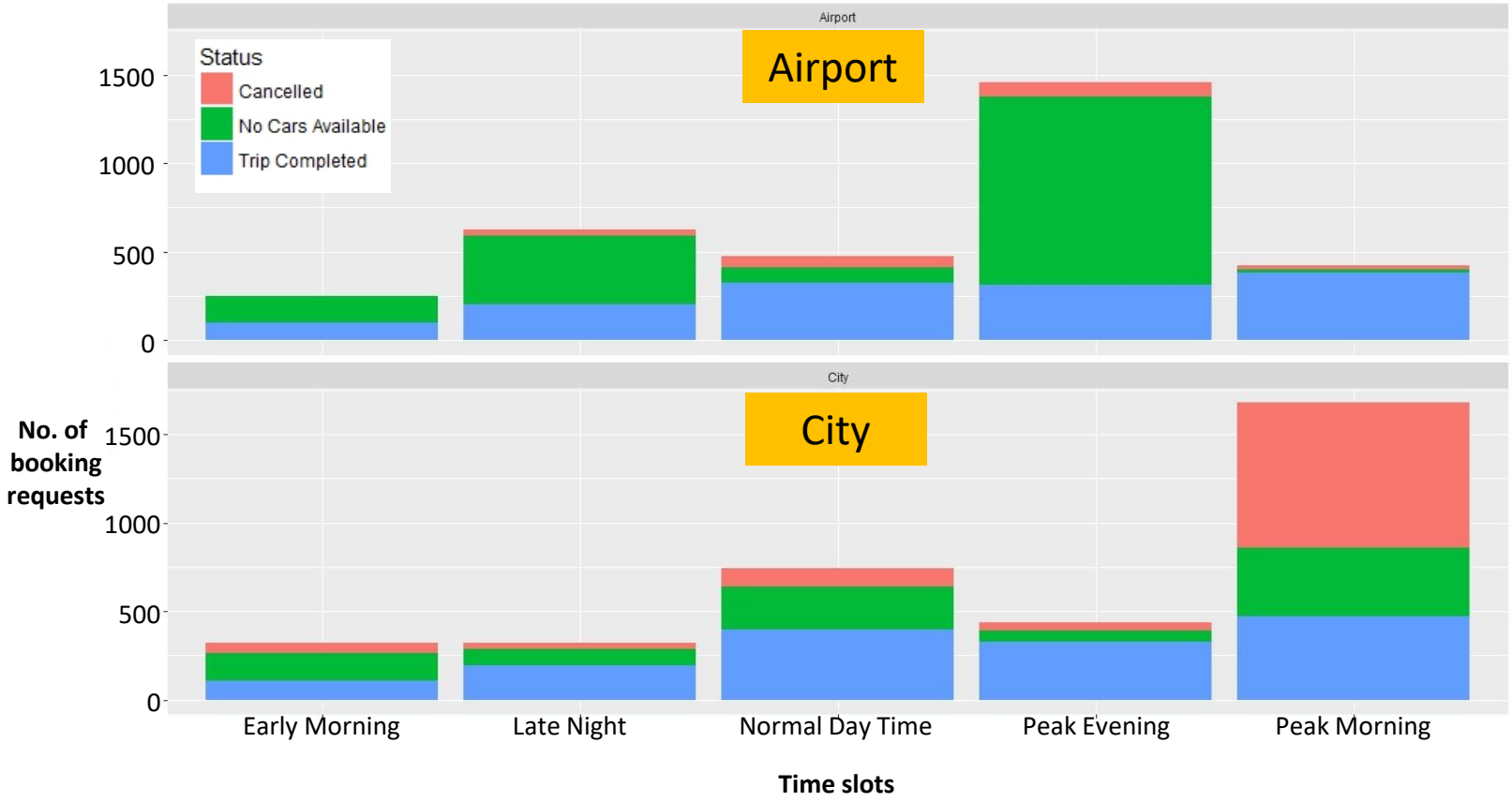


Maximum demand-supply gap for cars occurs in the peak morning (5:00 am – 9:59 am) and peak evening (5:00 pm – 8:59 pm) time slots.

Demand: Total number of trips (Trips Cancelled + No Cars available + Trips successfully completed)
Supply: Trips successfully completed

Analysis 2

Understanding the Demand-Supply Gap (Univariate Analysis w.r.t. Pickup point)



Maximum demand-supply gap for cars occurs in the peak morning (5:00 am – 9:59 am) time slot from the city.

Maximum demand-supply gap for cars occurs in the peak evening (5:00 pm – 8:59 pm) time slot from the airport.

Demand: Total number of trips (Trips Cancelled + No Cars available + Trips successfully completed)

Supply: Trips successfully completed

Analysis 2
Understanding the Demand-Supply Gap

From the graphs on the previous two slides, it is clear that:

- There is a lot of gap between demand for and supply of cabs from the airport in the peak evening timeslot (5:00 pm – 8:59 pm) and from the city in the peak morning (5:00 am – 9:59 am) timeslot.
- The demand-supply gap from the airport to the city is much severe in the peak evening timeslot (5:00 pm – 8:59 pm) as only 21.41% of the total trips in that particular timeslot are successfully completed.
- The demand-supply gap from the city to the airport is much severe in the peak morning timeslot (5:00 am – 9:59 am) as only 28.15% of the total trips in that particular timeslot are successfully completed
- This is largely responsible for high cancellation and unavailability of cars during these times.
- This indicates poor utilization of cabs during these timeslots.
- The reason for this gap is that a lot of people want to catch flights or return home during these peak hours, thus leading to increased demand for cabs, which subsequently outweighs supply, leading to unavailability of cars and fewer trips being successfully completed.
- The wait time for drivers leads to cancellations, as they move to the city for better business opportunities during peak time slots.

Recommendations

The following actions can be taken to bridge the demand-supply gap:

- Increasing the number of cabs and hiring more drivers to ensure more availability of cabs during the peak hour timeslots.
- Strict monitoring of cabs as the average time taken to commute from the city to the airport (and vice versa) is around 45 minutes.
- Incentivising drivers to complete trips in the peak hour timeslots and compensating them suitably for the wait time in these particular timeslots.
- Encouraging car pooling in the peak hour timeslots so that lesser number of cars serve more passengers.