UBER SUPPLY DEMAND GAP ANALYSIS

SUBMISSION

Santanu Dey (DDA1730043)

Abstract of Uber Supply Demand Gap Analysis

Introduction

- Customers of Uber often face the problem of cancellation by the driver or non-availability of cabs.
- These problems also affect the Uber's business, as Uber loses out on its revenue.

Goals

Present to the client:

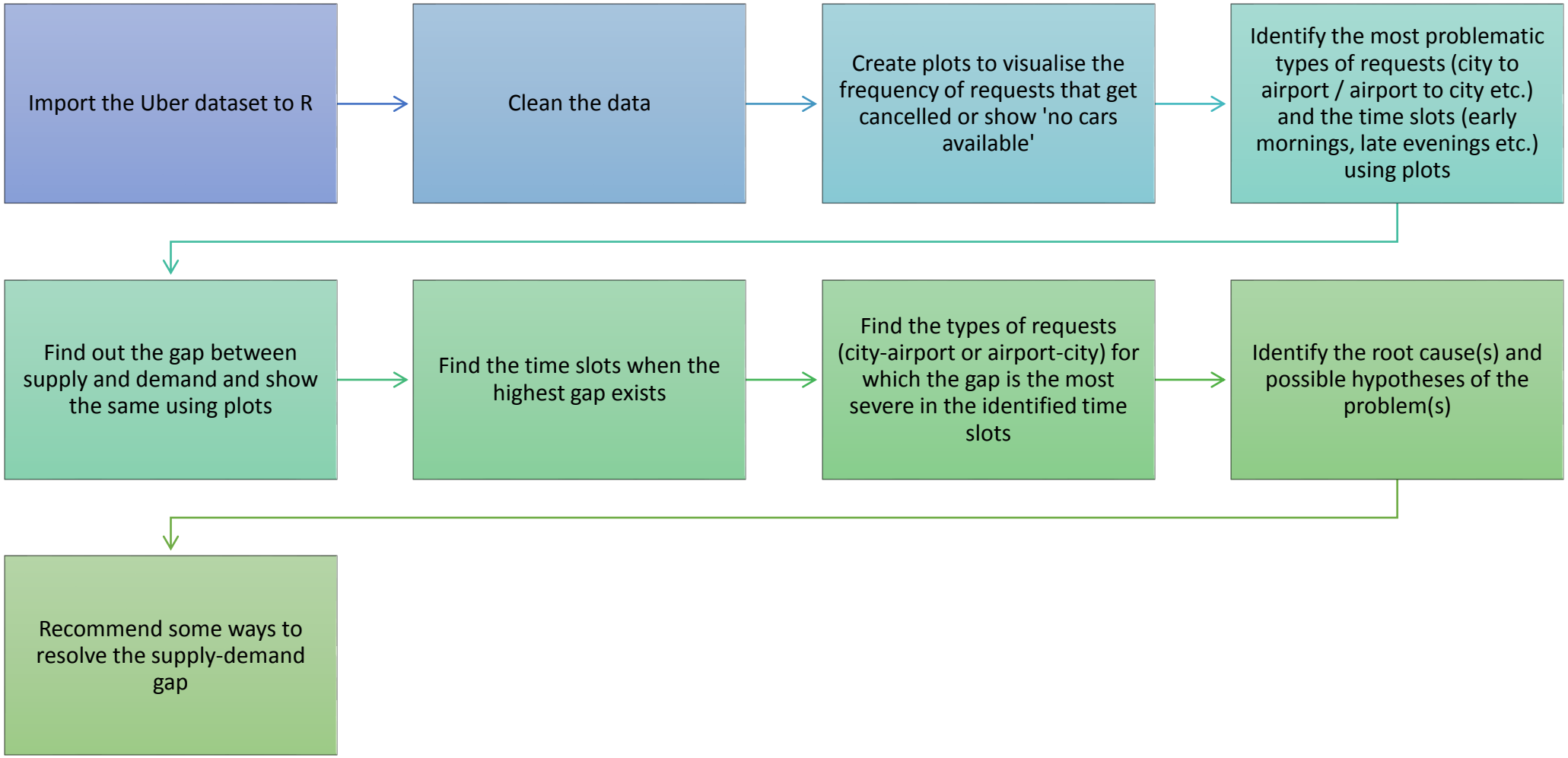
- the root cause(s) and possible hypotheses of the problem(s)
- recommend ways to improve the situation.

Constraints

The data available for analysis has the following constraints:

- Only the trips to and from the airport are being considered.
- Only the request timestamp and the drop timestamp (for completed rides) are available. Hence, no concrete analysis on the idle time, if any, can be performed.

Problem Solving Methodology



Data Understanding

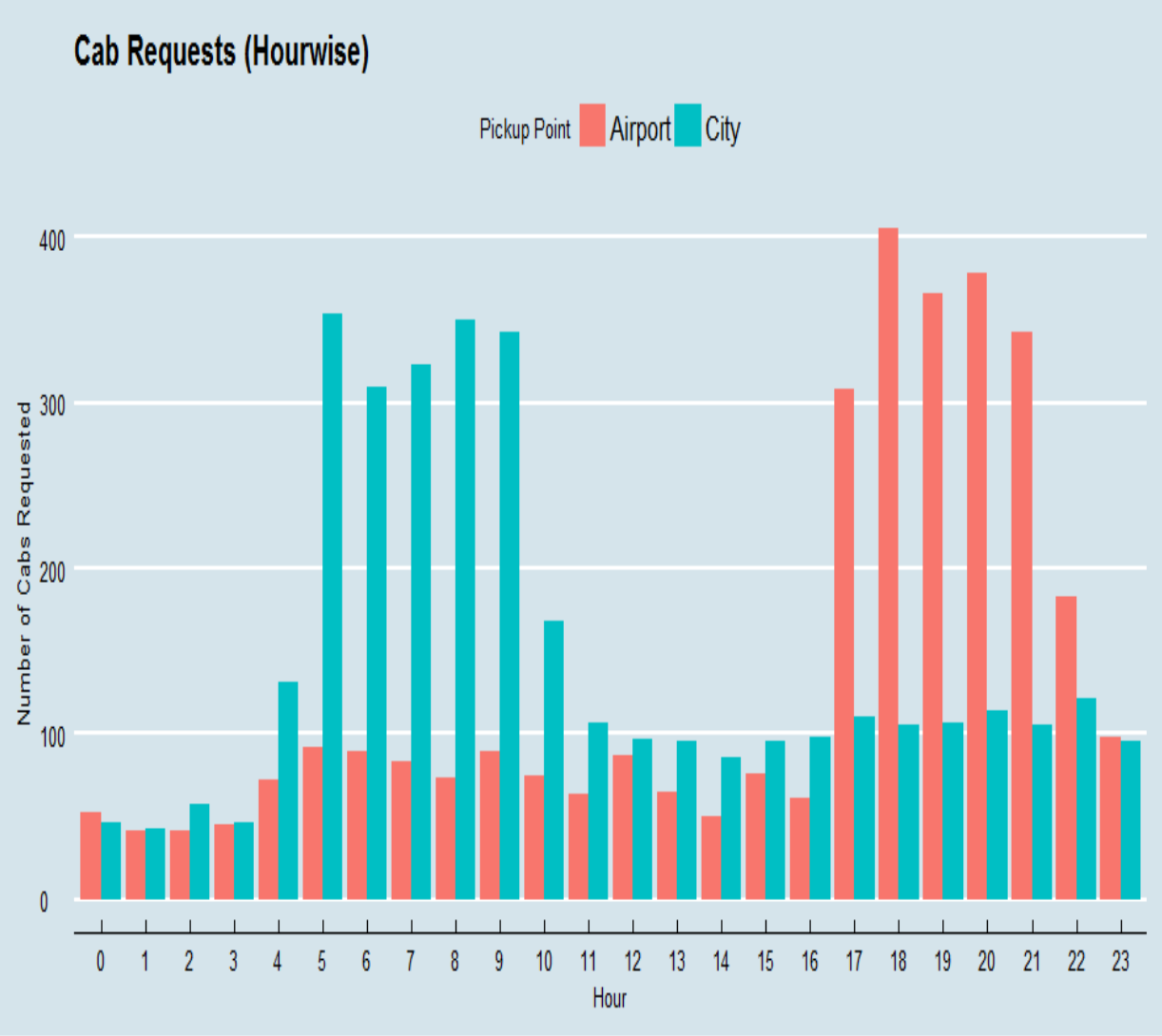
- Load the dataset *Uber Request Data.csv* into *uberData* data frame and view the data frame contents.
- There are 6745 ride requests (observations) with the following six attributes associated with each request made by a customer:
  - *Request id*: A unique identifier of the request
  - *Time of request*: 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

- *Status* has 3 unique values: *No Cars Available*, *Trip Completed*, and *Cancelled*.
- *Pick-up point* has 2 unique values: *City*, and *Airport*.
- Find the number of entries for each of the 3 unique *Status* values
- Find how many NA values are present in each column of the data frame:
  - Driver.id column has 2650 NA entries, which corresponds to the 2650 entries for "No Cars Available" Status entries.
  - Drop.timestamp column has 3914 NA entries, which corresponds to the sum of 1264 "Cancelled" and 2650 "No Cars Available" Status entries.

Data Cleaning

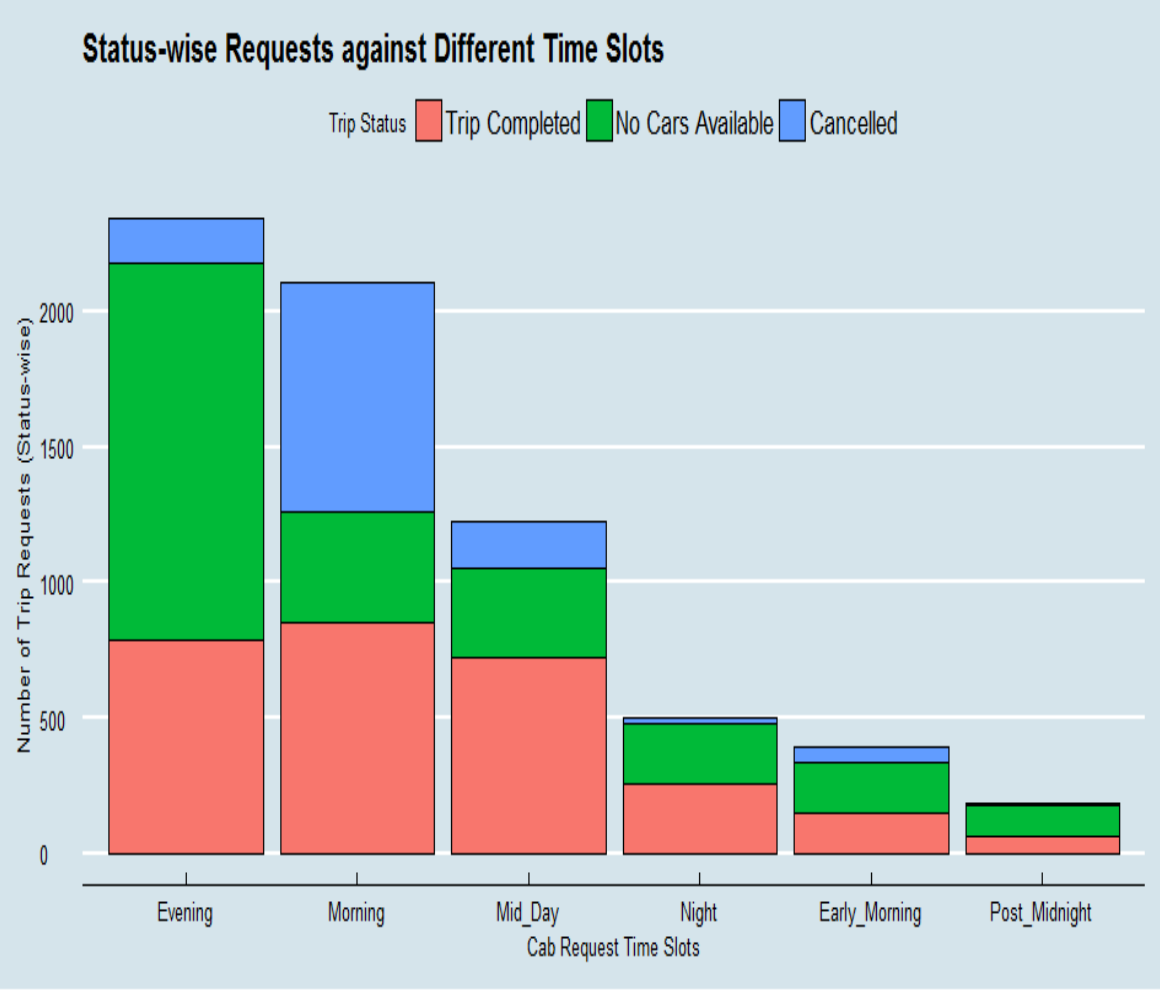
- We observe that there are multiple formats present in the dataset for the timestamp columns.
- Convert each timestamp column to the same format. We also ensure that the existing NA values are not impacted during this conversion.
- Separate each timestamp column into date and time columns, and drop the timestamp columns.
- Find the unique request dates: *07-13-2016, 07-14-2016, 07-15-2016, 07-12-2016, 07-11-2016*
- Extract the hour (in 24-hour format) from the time columns for request and drop.

Data Cleaning (contd.)



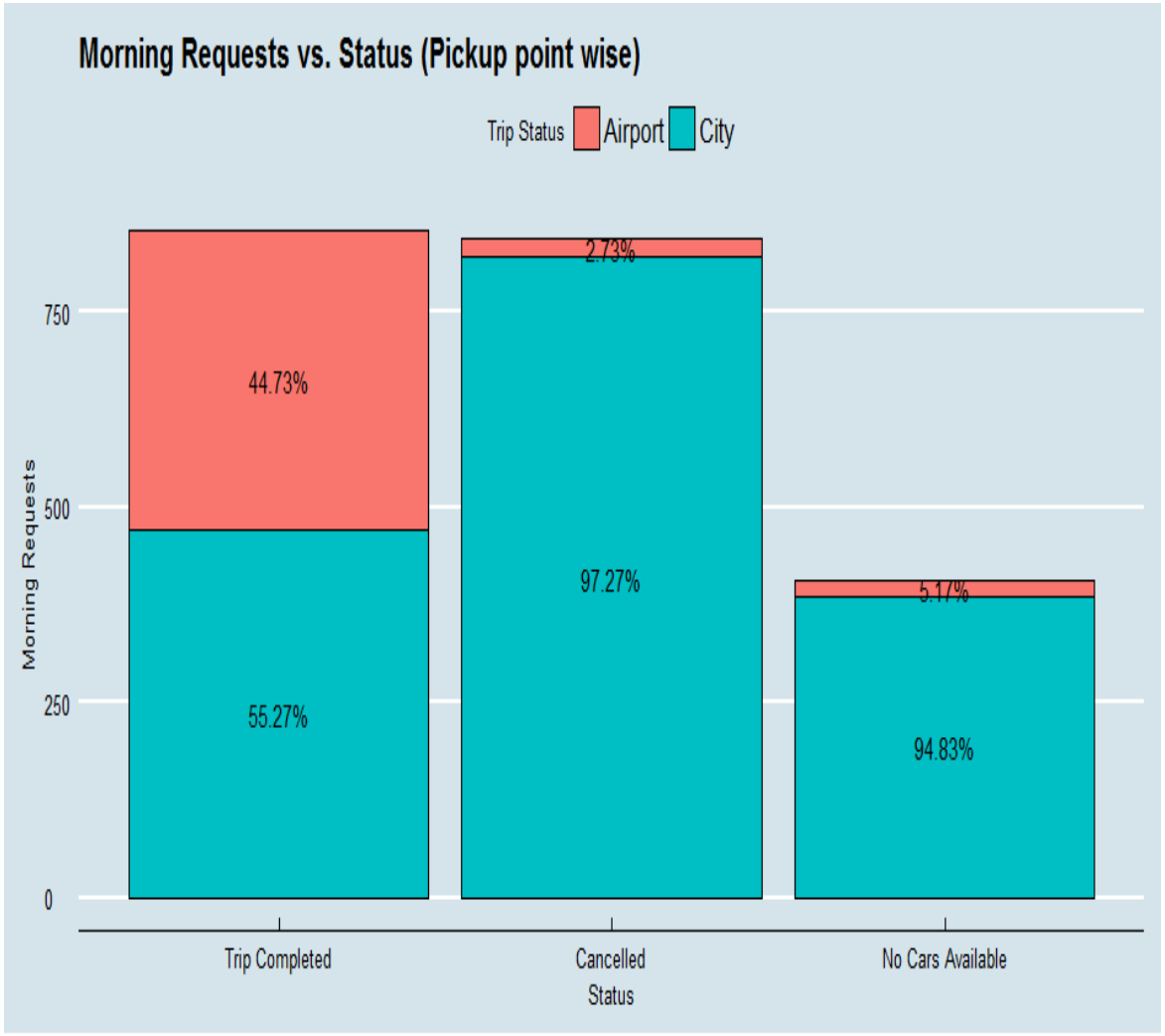
- Plot the number of cab requests in a particular hour for all five days.
- We observe that:
  - Frequency of requests from City to Airport is highest during the morning hours 5 to 9.
  - Frequency of requests from Airport to City is highest during the evening hours 17 to 21.
  - Based on these observations, it is suitable to partition the data into different time slots, in order to have a focussed approach for the analysis.

- Assign one of the following time slots to each request hour:
  - Post_Midnight (0, 1)
  - Early_Morning (2, 3, 4)
  - Morning (5, 6, 7, 8, 9)
  - Mid_Day (10, 11, 12, 13, 14, 15, 16)
  - Evening (17, 18, 19, 20, 21)
  - Night (22, 23)



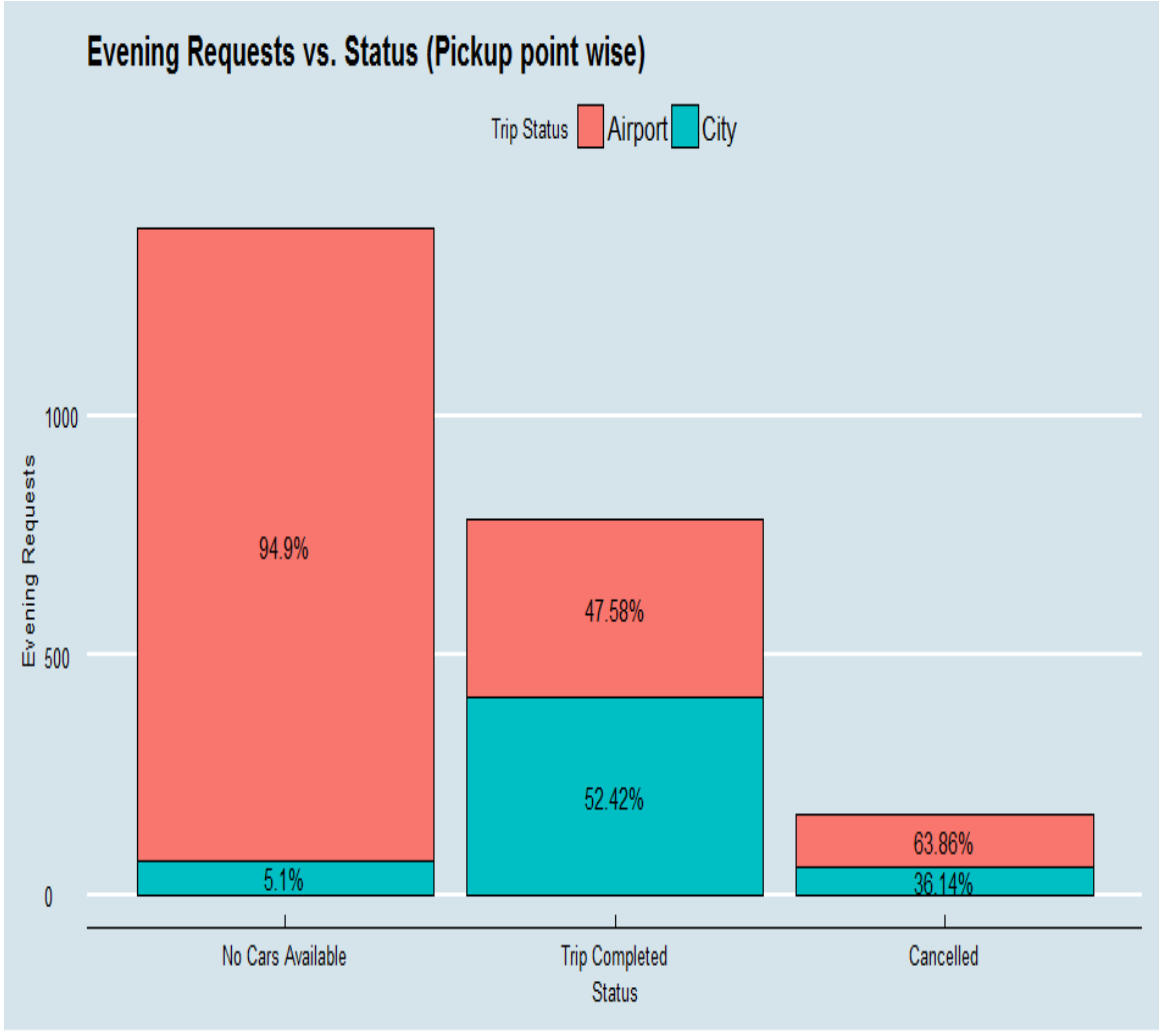
Data Analysis

- Plot a bar chart for request time slots by number of requests
- We observe that the TOP two problems are as follows:
  - A lot of cab requests get CANCELLED in the MORNING Time Slot
  - A lot of cab requests get the NO CARS AVAILABLE response in the EVENING Time Slot.



Data Analysis – Top Problem

- Plot a bar chart for Status by Number of Requests for Morning time slot
- We observe that in the Morning Time Slot:
  - 97.27% of all CANCELLED Requests are from the City
  - 94.83% of all NO CARS AVAILABLE are from the City



Data Analysis – Second Top Problem

- Plot a bar chart for Status by Number of Requests for Evening time slot
- We observe that in the Evening Time Slot:
  - 63.86% of all CANCELLED Requests are from the Airport
  - 94.9% of all NO CARS AVAILABLE are from the Airport

Data Analysis – Demand Supply Gap Analysis

- Prepare the *demand* data frame
- Prepare the demand supply *gap* data frame

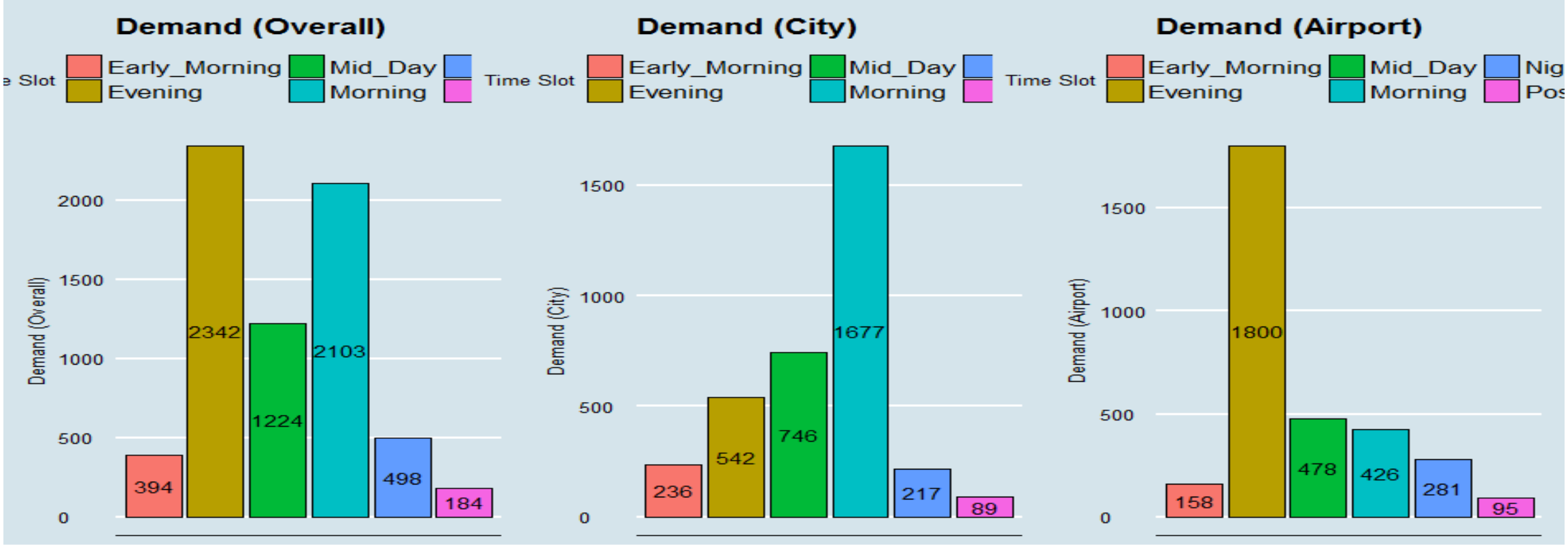
- **Considerations**:
  - *Demand*: All customer requests contribute to the Demand
  - *Supply*: Only completed Trips contribute to the Supply
  - *Gap*: All Cancelled and No Cabs Available statuses contribute to the Gap, which is essentially the difference between the Demand and the Supply

- Plot bar graphs for Demand, Supply and the Gap for each time slot for the following three categories:
  - Overall
  - City to Airport
  - Airport to City

- **Assumption**:

As it is previously identified that Morning and Evening time slots are the most effected ones, we will focus our observations only on these two time slots.

Data Analysis – Demand Analysis



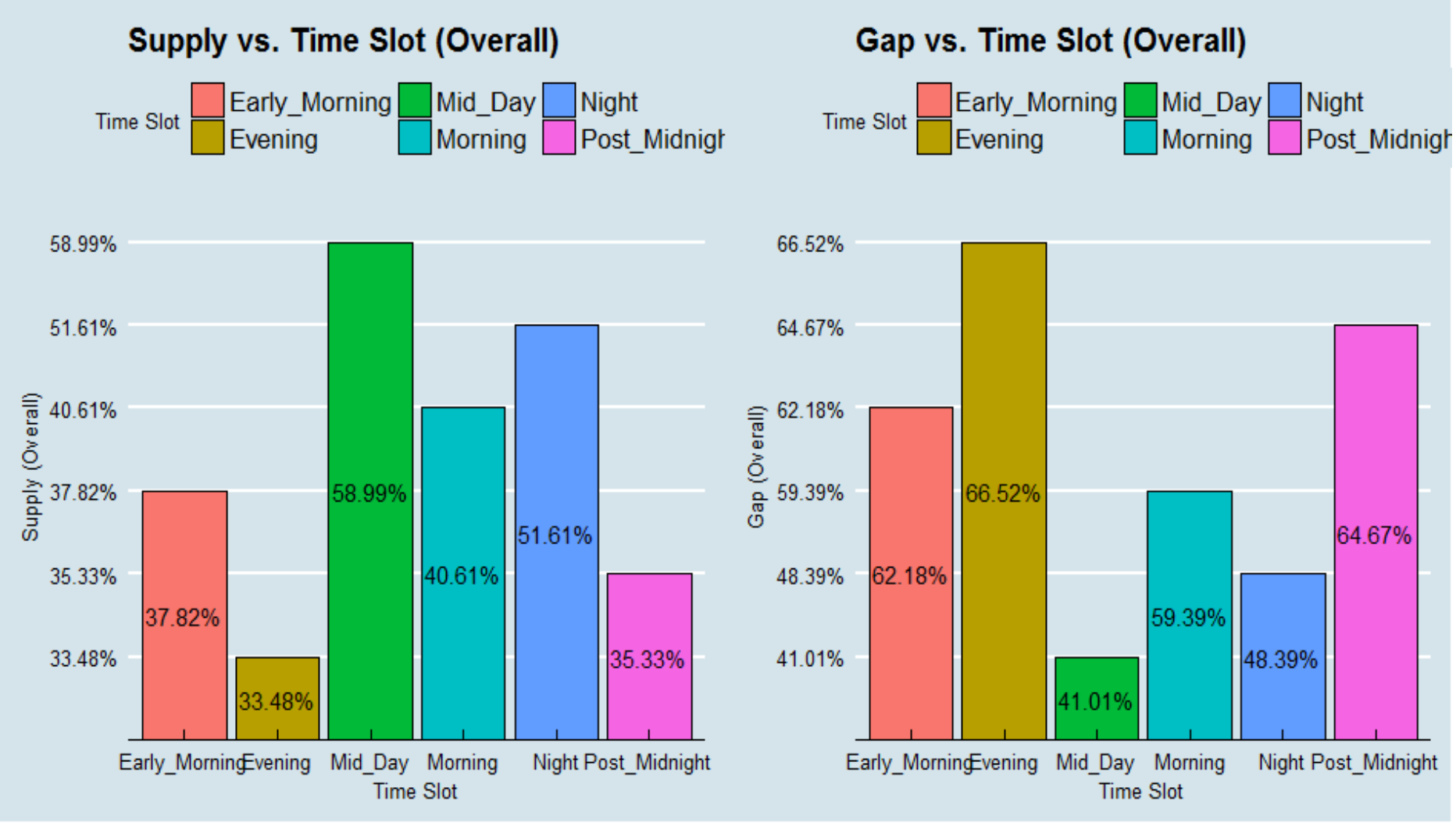
Plot Observations:

- Overall highest demand is in the Morning and Evening time slots.
- Highest demand from City to Airport is in the Morning time slot.
- Highest demand from Airport to City is in the Evening time slot.

Data Analysis – Demand Supply Gap Analysis (Overall)

Plot Observations:

- Only 40.61% Supply overall in the morning.
- Only 33.48% Supply overall in the evening

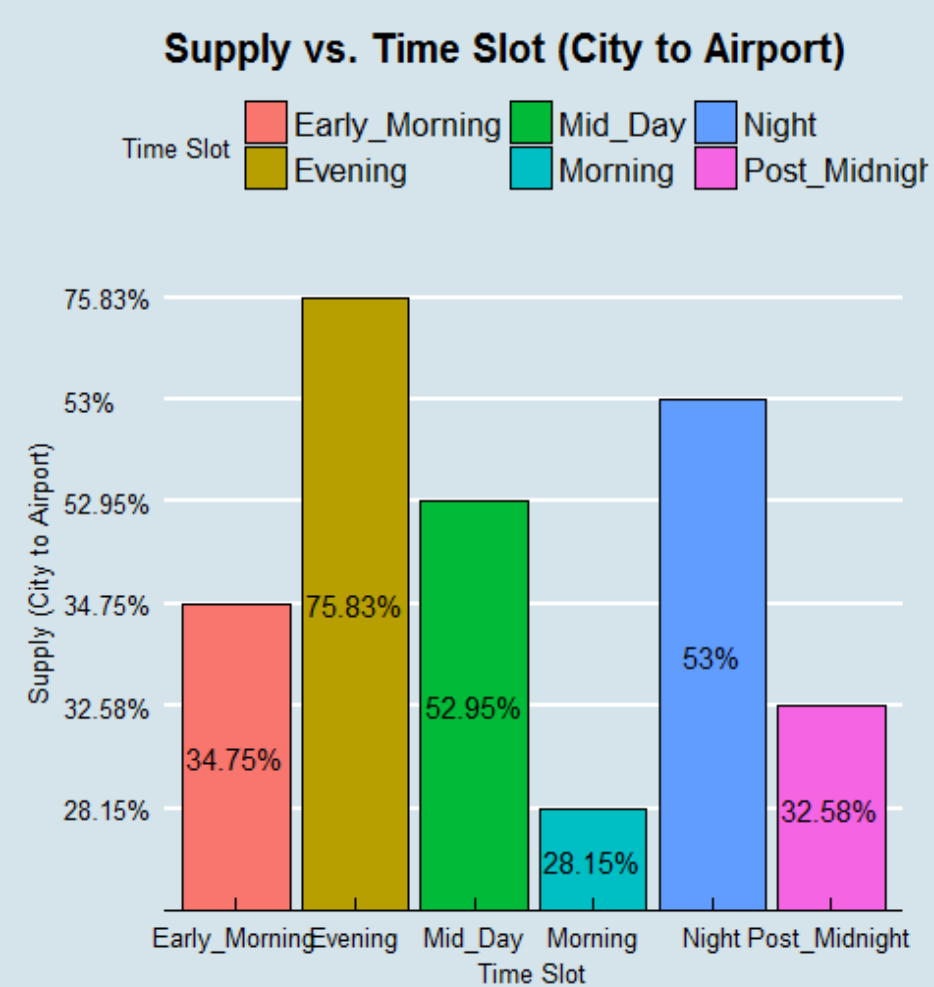


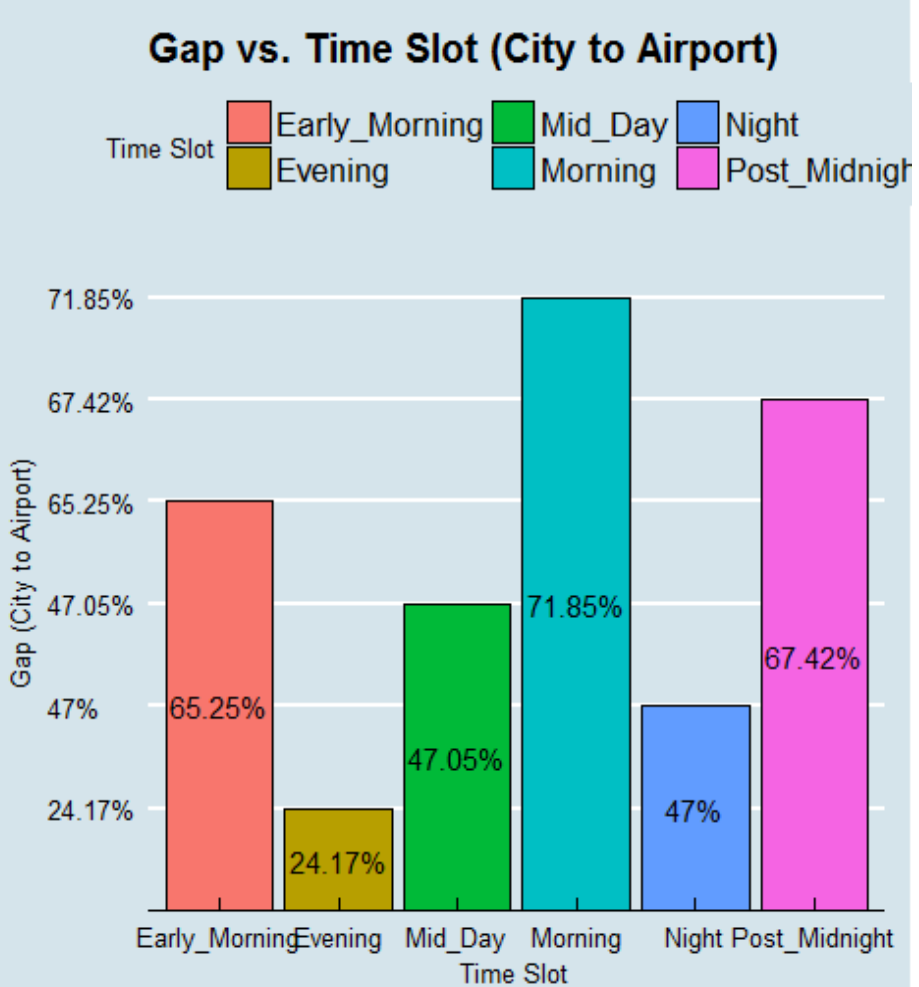
Data Analysis – Demand Supply Gap Analysis (City to Airport)

Plot Observations:

For requests from City to Airport:

- Only 28.15% Supply in the morning.
- A very high 75.83% Supply in the evening



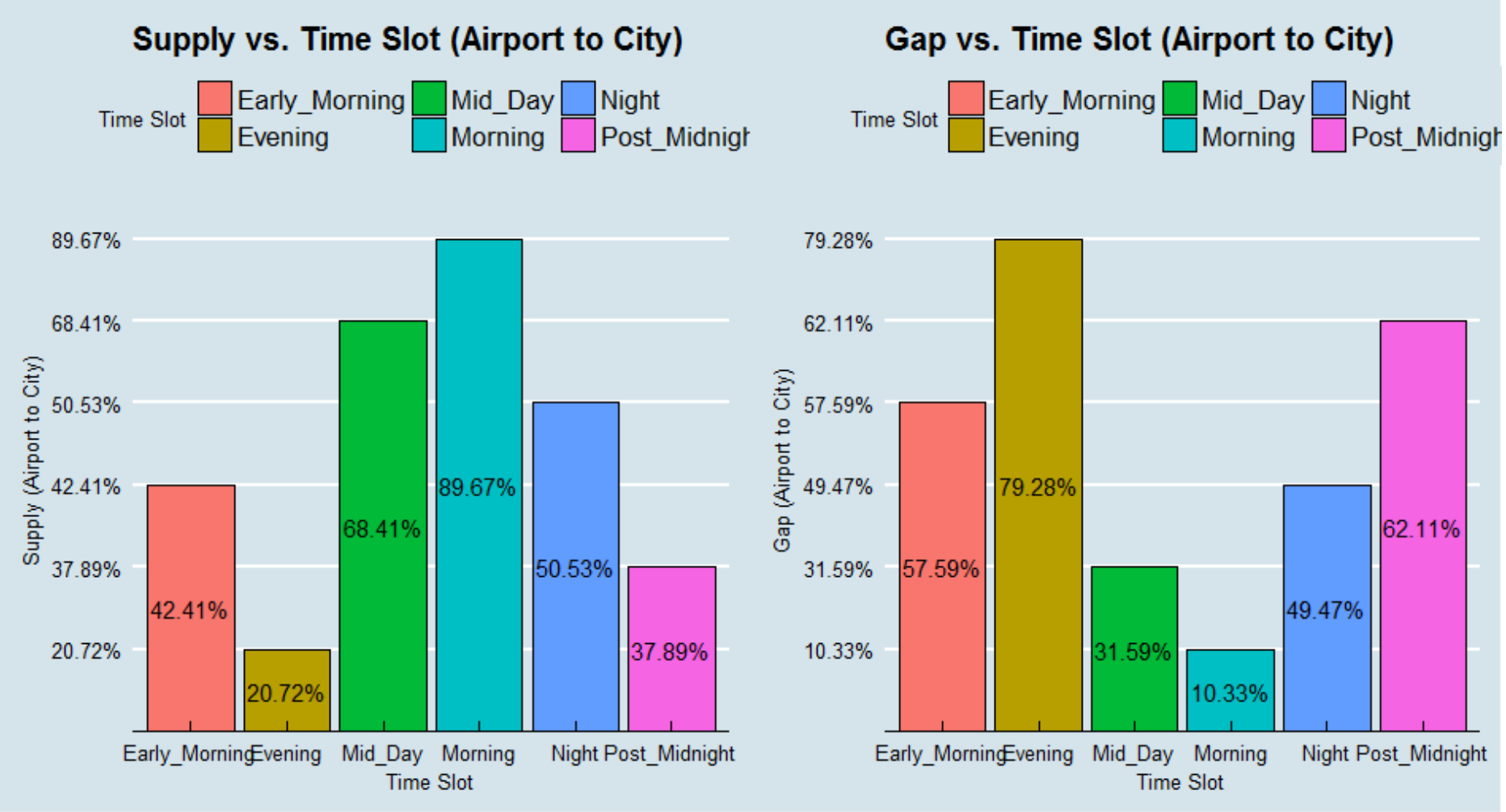


Data Analysis – Demand Supply Gap Analysis (Airport to City)

Plot Observations:

For requests from Airport to City:

- A very high 89.67% Supply in the morning.
- Only 20.72% Supply in the evening



Data Analysis – Demand Supply Gap Analysis (Possible Reasons)

- **Reason 1**: Due to heavy office traffic during morning time slot, and high demand for office-home route, which are usually shorter distances, drivers tend to cancel requests from City to Airport.

- **Reason 2**: Most drivers pick up rides for office to home in the evening time slot, causing a lot of No Cars Available.

- **Reason 3**: From the airport, drivers tend to cancel rides after checking the distance, duration and traffic after accepting a particular ride.

Data Analysis – Demand Supply Gap Analysis (Ways to Resolve)

- **Resolution 1**: Incentivize the City to Airport route in the morning and evening time slots. This will motivate drivers to cancel requests less often.

- **Resolution 2**: Assign dedicated cabs for Airport to City route in the morning and evening time slots, as these are the time slots which have the highest demand.

- **Resolution 3**: Prioritize pre-booking by customers for travel to and from the Airport. This would allow dedicated services to be allocated much better between the City and Airport.

Thank You