# UBER CASE STUDY

Uber Supply-Demand Gap case study

By-Abhijeet Sengupta

# Business Understanding

- Uber Technologies Inc. is a peer-to-peer ridesharing, taxi cab, food delivery, and transportation network company headquartered in San Francisco, California, with operations in 633 cities worldwide.

- Uber is loosing out its revenue, If drivers cancel the request of riders or cars are unavailable while riders are trying to book the cab.

- Driver cancellation and non-availability of cars leading to loss of potential revenue.

# Business Objectives

- Identify the root cause of the following problem
  - cancellation initiated by cab driver
  - non-availability of cars
- Root cause(s) and possible hypotheses of the problem(s)
- Recommend ways to improve this situation of demand and supply gap.

# Methodology for Analysis

- ✓Gather and load data for the analysis
- ✓Understand the data by visual inspection
- ✓Clean and correct the data format( Timestamp , date time format )
- ✓Estimate Derived metrics
  - ✓Business Driven
  - ✓Data Driven
- ✓Perform Univariate and Segmented Analysis
- ✓Plot the results of analysis and find the root cause
- ✓Recommend the solution to improve the business.

# Data Understanding

- Data description

| Meta Data | Exit Data |
|---|---|
| Description | Uber cab booking request data |
| Source | UpGrad Assignement masked data |
| Format | csv |
| Number Of Rows | 6745 |
| Each Row | A ride detail of perticula customer |
| Sampling method | Pickup point city and Airport only |
| Prepared Data | NA |

| Column Name | Type | Description | Missing | Uniques |
|---|---|---|---|---|
| Request id | Numeric | A unique identifier of the request | 0 | 6745 |
| Pickup point | String | The point from which the request was made | 0 | 2 |
| Driver id | Numeric | The unique identification number of the driver | 2650 | 301 |
| Status | String | The final status of the trip | 0 | 3 |
| Request timestamp | Date Time | Time at which the customer made request | 0 | 5618 |
| Drop timestamp | Date Time | drop-off date and time if trip completed | 3914 | 2599 |

Format based on lecture: EDA data description

# Data Cleaning and Preparation

- Date/time provided are not in consistent format.

| Request.id | Pickup.point | Driver.id | Status | Request.timestamp | Drop.timestamp |
|---|---|---|---|---|---|
| 619 | Airport | 1 | Trip Completed | 11/7/2016 11:51 | 11/7/2016 13:00 |
| 867 | Airport | 1 | Trip Completed | 11/7/2016 17:57 | 11/7/2016 18:47 |
| 1807 | City | 1 | Trip Completed | 12/7/2016 9:17 | 12/7/2016 9:58 |
| 2532 | Airport | 1 | Trip Completed | 12/7/2016 21:08 | 12/7/2016 22:03 |
| 3112 | City | 1 | Trip Completed | 13-07-2016 08:33:16 | 13-07-2016 09:25:47 |
| 3879 | Airport | 1 | Trip Completed | 13-07-2016 21:57:28 | 13-07-2016 22:28:59 |
| 4270 | Airport | 1 | Trip Completed | 14-07-2016 06:15:32 | 14-07-2016 07:13:15 |
| 5510 | Airport | 1 | Trip Completed | 15-07-2016 05:11:52 | 15-07-2016 06:07:52 |
| 6248 | City | 1 | Trip Completed | 15-07-2016 17:57:27 | 15-07-2016 18:50:51 |
| 267 | City | 2 | Trip Completed | 11/7/2016 6:46 | 11/7/2016 7:25 |

Dates are separated by “/” and “-”. This Make it inconsistent for any analysis

.....

.....

.....

| Request.timestamp | Drop.timestamp |
|---|---|
| 2016-07-11 11:51:00 | 2016-07-11 13:00:00 |
| 2016-07-11 17:57:00 | 2016-07-11 18:47:00 |
| 2016-07-12 09:17:00 | 2016-07-12 09:58:00 |
| 2016-07-12 21:08:00 | 2016-07-12 22:03:00 |
| 2016-07-13 08:33:00 | 2016-07-13 09:25:00 |
| 2016-07-13 21:57:00 | 2016-07-13 22:28:00 |
| 2016-07-14 06:15:00 | 2016-07-14 07:13:00 |
| 2016-07-15 05:11:00 | 2016-07-15 06:07:00 |
| 2016-07-15 17:57:00 | 2016-07-15 18:50:00 |
| 2016-07-11 06:46:00 | 2016-07-11 07:25:00 |

Consistent format and represented as date object

# Data Cleaning and Preparation….

- Introducing new derived metrics

| Request.id | Pickup.point | Driver.id | Status | Request.timestamp | Drop.timestamp | Request.Day | Request.Hrs | Drop.Day | Drop.Hrs |
|---|---|---|---|---|---|---|---|---|---|
| 619 | Airport | 1 | Trip Completed | 2016-07-11 11:51:00 | 2016-07-11 13:00:00 | 11 | 11 | 11 | 13 |
| 867 | Airport | 1 | Trip Completed | 2016-07-11 17:57:00 | 2016-07-11 18:47:00 | 11 | 17 | 11 | 18 |
| 1807 | City | 1 | Trip Completed | 2016-07-12 09:17:00 | 2016-07-12 09:58:00 | 12 | 09 | 12 | 09 |
| 2532 | Airport | 1 | Trip Completed | 2016-07-12 21:08:00 | 2016-07-12 22:03:00 | 12 | 21 | 12 | 22 |
| 3112 | City | 1 | Trip Completed | 2016-07-13 08:33:00 | 2016-07-13 09:25:00 | 13 | 08 | 13 | 09 |
| 3879 | Airport | 1 | Trip Completed | 2016-07-13 21:57:00 | 2016-07-13 22:28:00 | 13 | 21 | 13 | 22 |
| 4270 | Airport | 1 | Trip Completed | 2016-07-14 06:15:00 | 2016-07-14 07:13:00 | 14 | 06 | 14 | 07 |
| 5510 | Airport | 1 | Trip Completed | 2016-07-15 05:11:00 | 2016-07-15 06:07:00 | 15 | 05 | 15 | 06 |
| 6248 | City | 1 | Trip Completed | 2016-07-15 17:57:00 | 2016-07-15 18:50:00 | 15 | 17 | 15 | 18 |
| 267 | City | 2 | Trip Completed | 2016-07-11 06:46:00 | 2016-07-11 07:25:00 | 11 | 06 | 11 | 07 |

Four new columns, Req.Day, Req.Hrs, Drop.Day & Drop.Hrs is derived from Request.timestamp & Drop.timestamp columns.

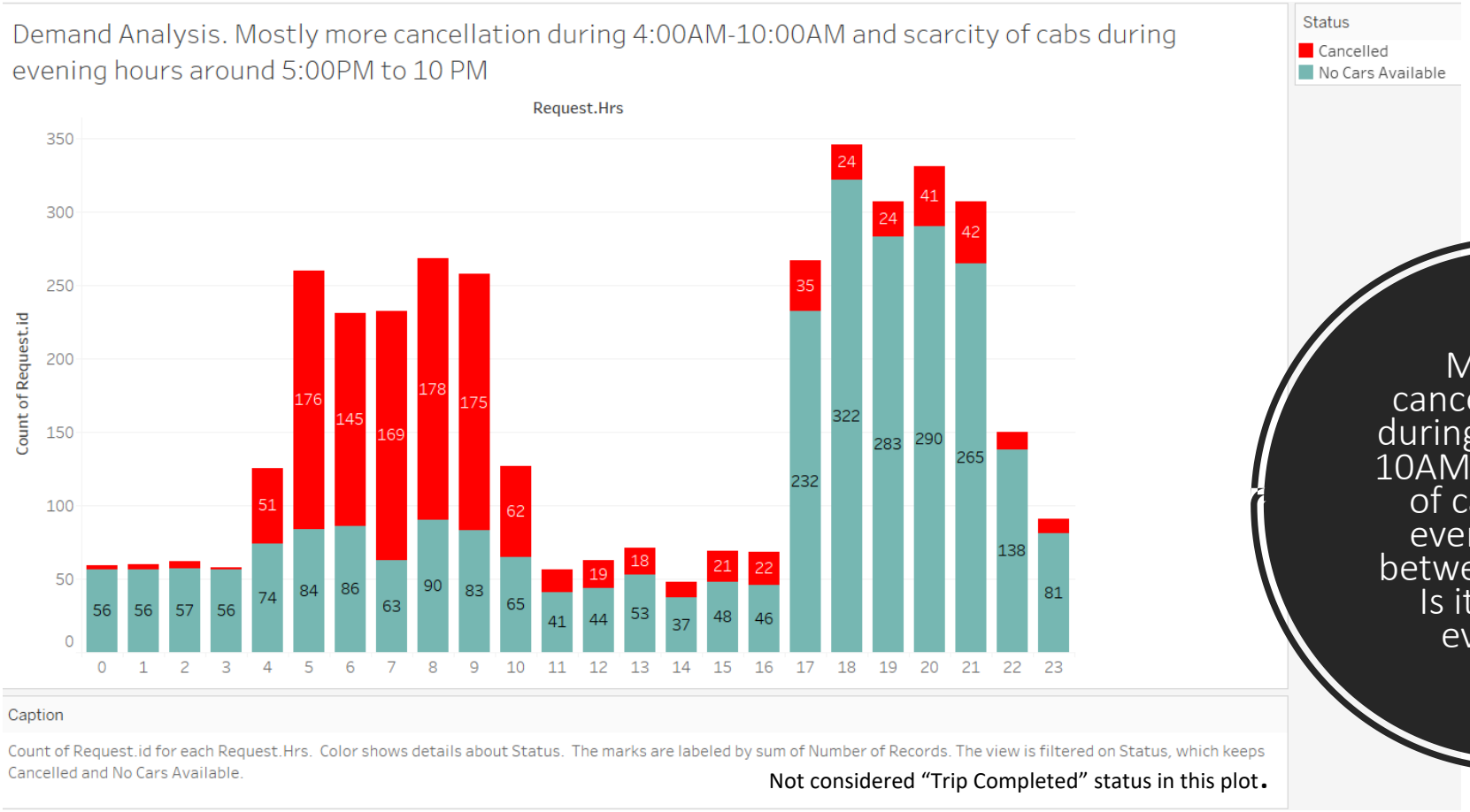
Post Data cleaning data stat: `6745 obs. of 10 variables`

# Request Analysis – Cancellation and No Cars Available is the major issue.

- We found there is high number of "NO CARS AVAILABLE" issue. Approximately the same as that of  number of trip completed.
- Cancelled request ratio is also very high.
- This needs to be Investigated for root cause analysis.

| | Status | CountOfRequest | perc |
|---|---|---|---|
| | <chr> | <int> | <dbl> |
| 1 | Cancelled | 1264 | 18.7 |
| 2 | No Cars Available | 2650 | 39.3 |
| 3 | Trip Completed | 2831 | 42.0 |

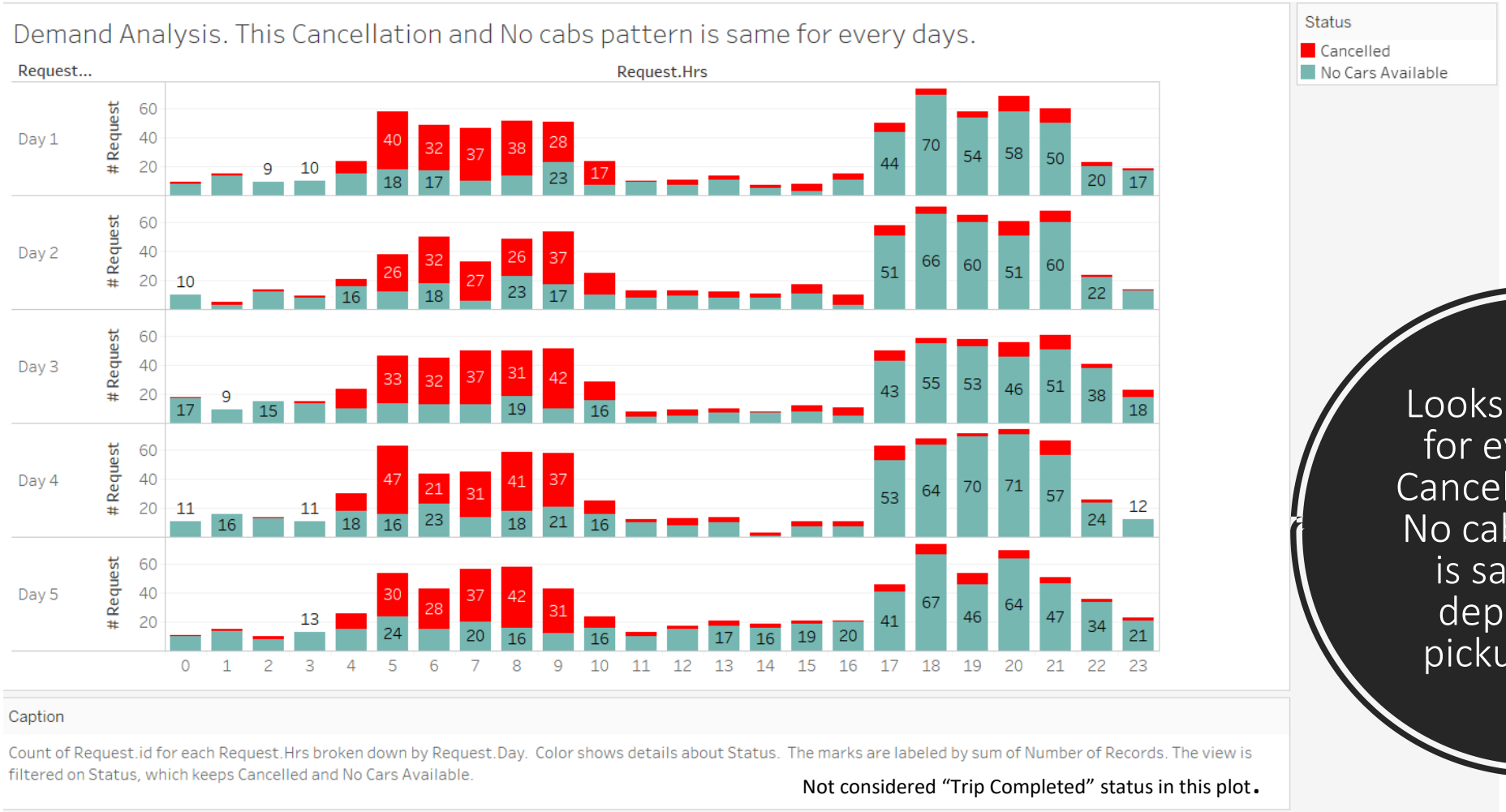
# Request Analysis- Cancellation and Cabs Availability




Not considered “Trip Completed” status in this plot.

Mostly all cancelations are during morning 4-10AM and scarcity of cabs during evening hours between 5-10PM. Is it same for everyday?

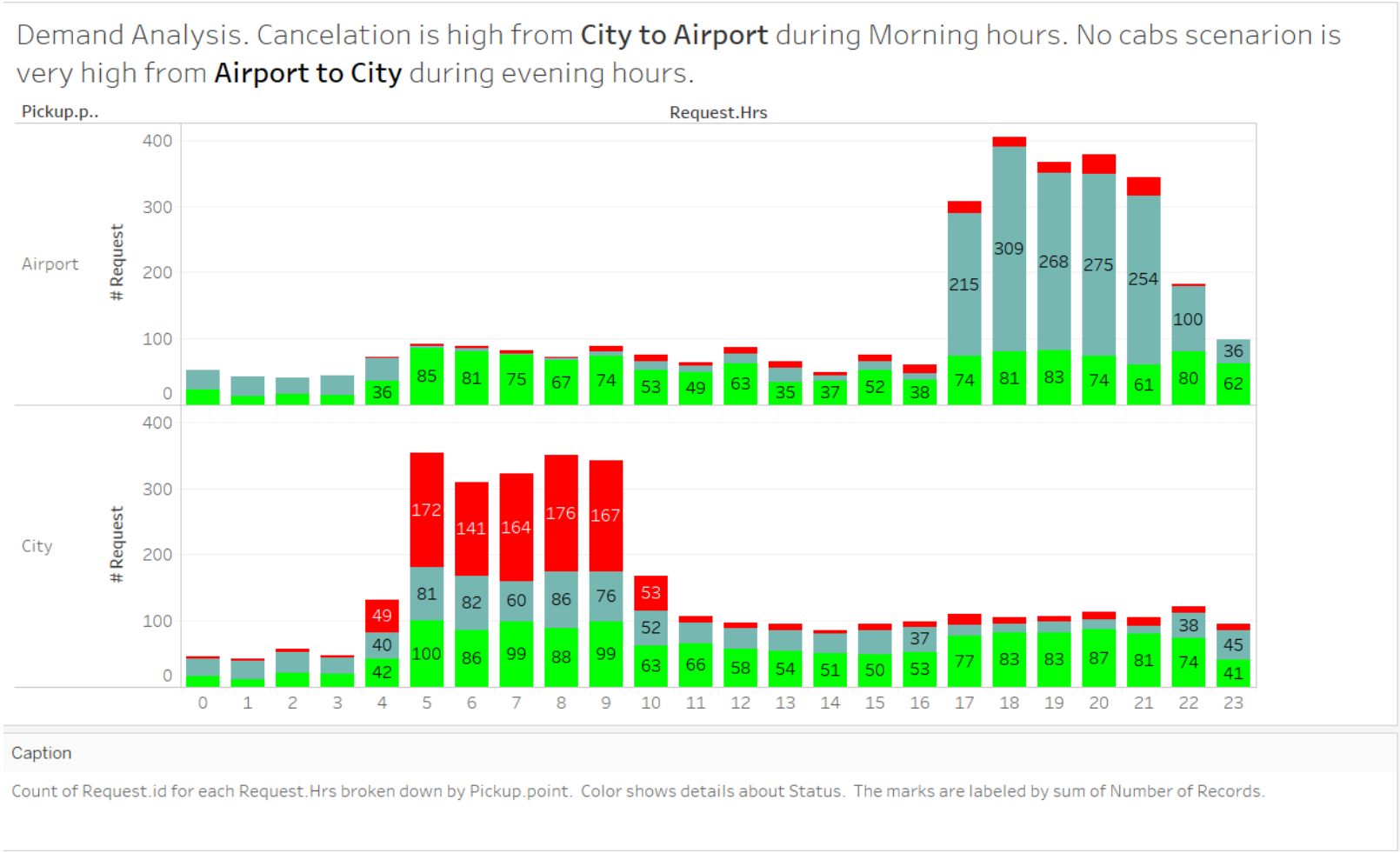
# Request Analysis- Cancellation and Cabs Availability…




Not considered “Trip Completed” status in this plot.

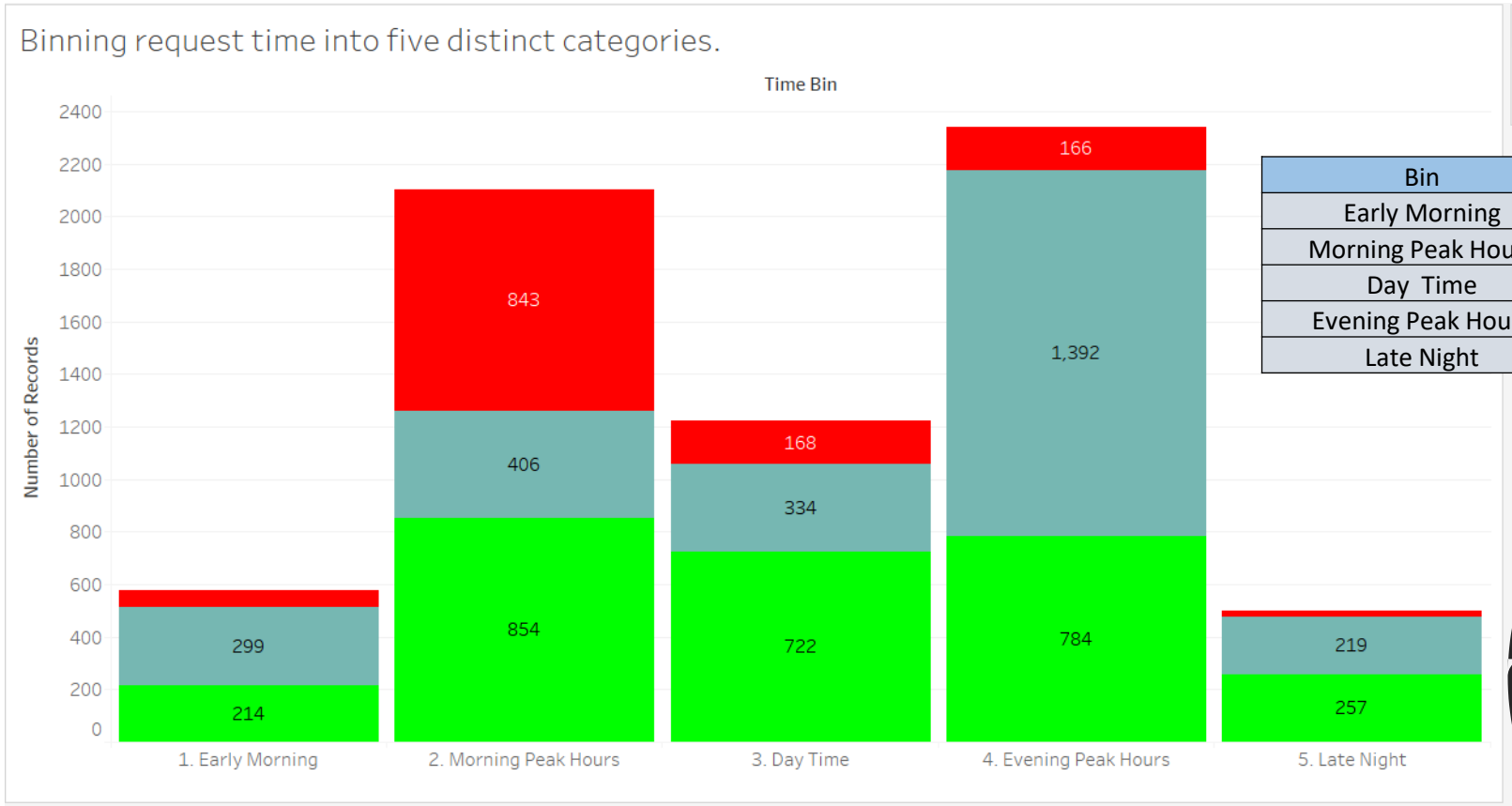
Looks like same for every day. Cancellation and No cabs pattern is same. Is it depends on pickup point?

# Request Analysis- Cancellation and Cabs Availability…




Cab drivers are cancelling more City to Airport trips during Morning hours. Huge shortage of cabs in the evening from Airport to City

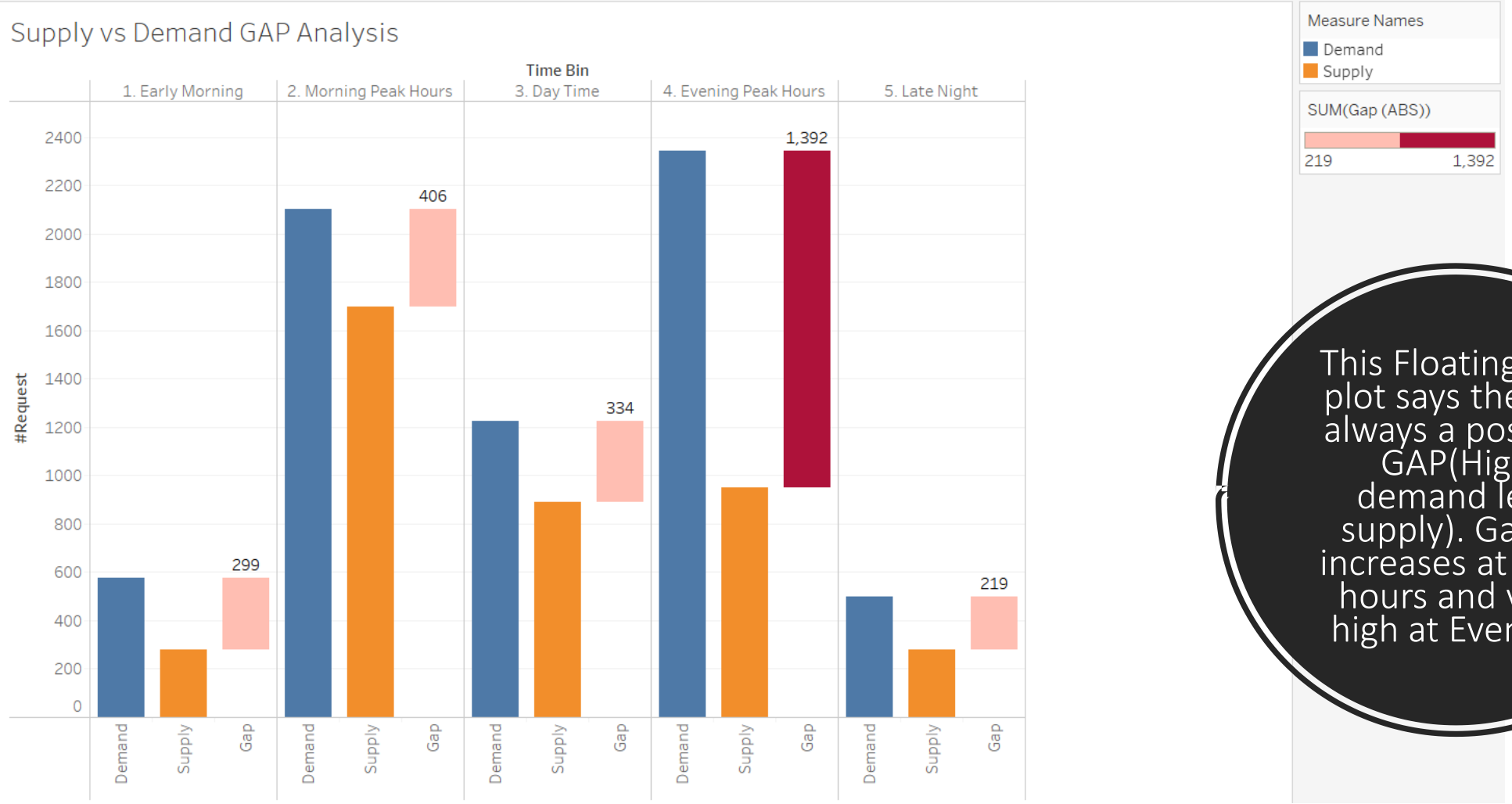
# Request Analysis- Cancellation and Cabs Availability…




| Bin | Time |
|---|---|
| Early Morning | 12 AM- 5 AM |
| Morning Peak Hours | 5 AM - 10 AM |
| Day Time | 10 AM - 5 PM |
| Evening Peak Hours | 5 PM - 10 PM |
| Late Night | 10 PM - 12 AM |

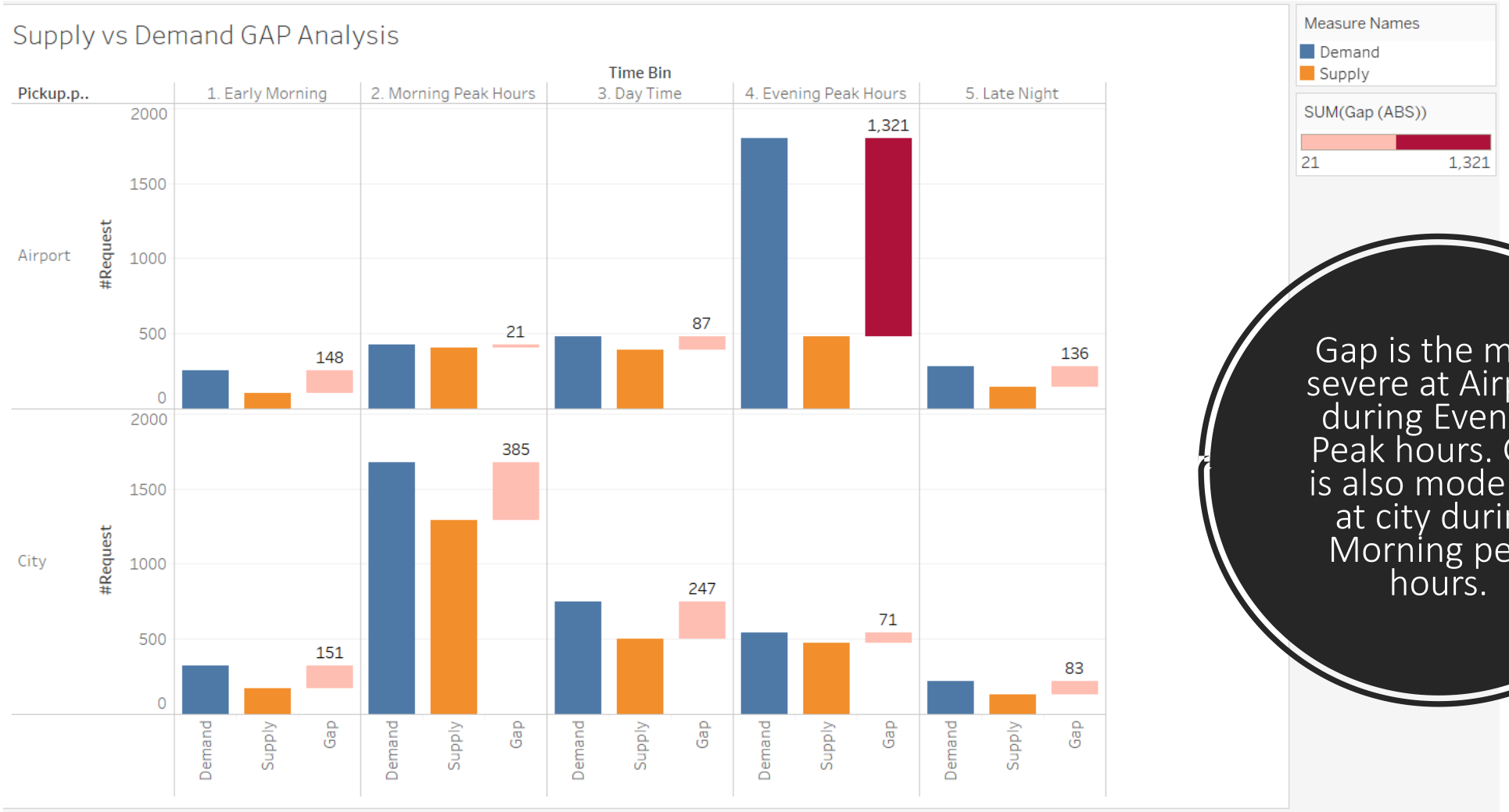
Graph Shows two major issues clearly,
1) Cancellation during morning peak hours
2) Unavailability of cabs during evening peak hours.

# Supply vs Demand GAP Analysis




This Floating BAR plot says there is always a positive GAP(High demand less supply). Gap is increases at peak hours and very high at Evening.

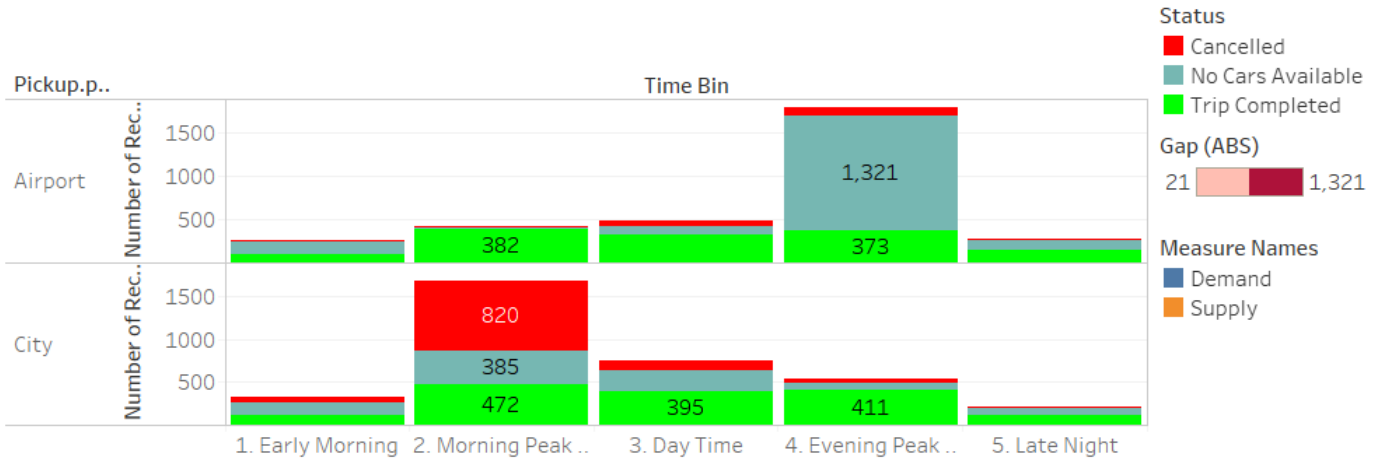
# Supply vs Demand GAP Analysis…



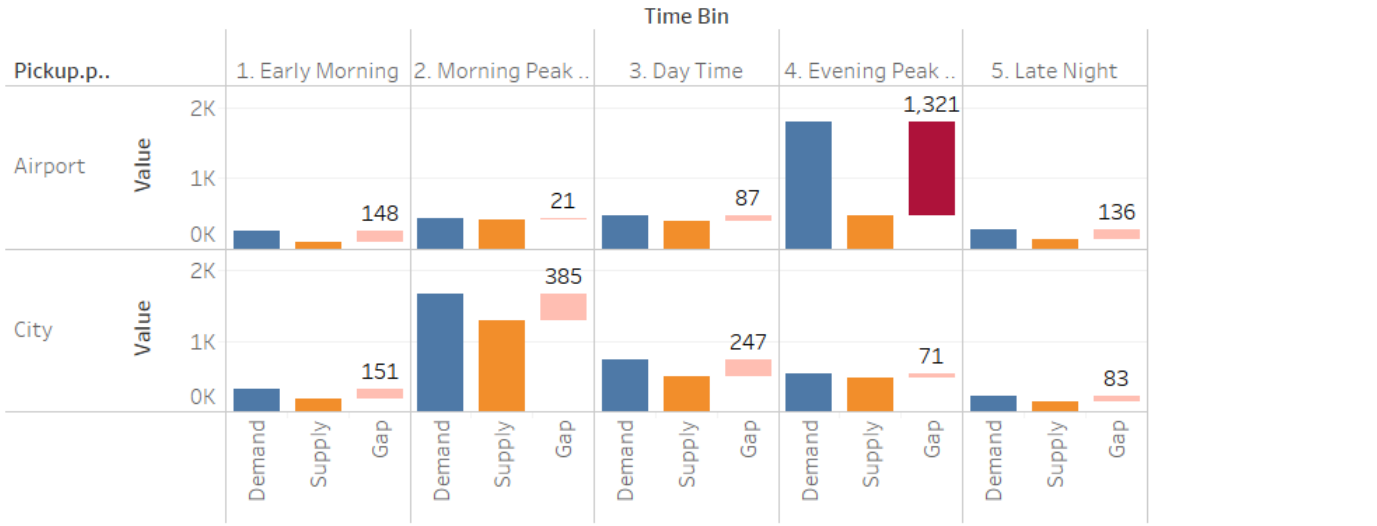

Gap is the most severe at Airport during Evening Peak hours. Gap is also moderate at city during Morning peak hours.

## Observations for Trip starting @ City

- For morning peak hours, the volume of request is very high.
- This High request is not being converted to a business leading to a loss of potential revenue.
- Morning peak hours is the main segment of time where UBER need to concentrate.
- Supply is adequate in the city but still morning peak hours need to be under scanner.
- The Major issue in city is Cancellation of cab and it's very high during morning peak hours.
- Why cab drivers are cancelling the request is major area of investigation.
- One reason could be low demand in airport during morning hours.




## Supply vs Demand GAP Analysis

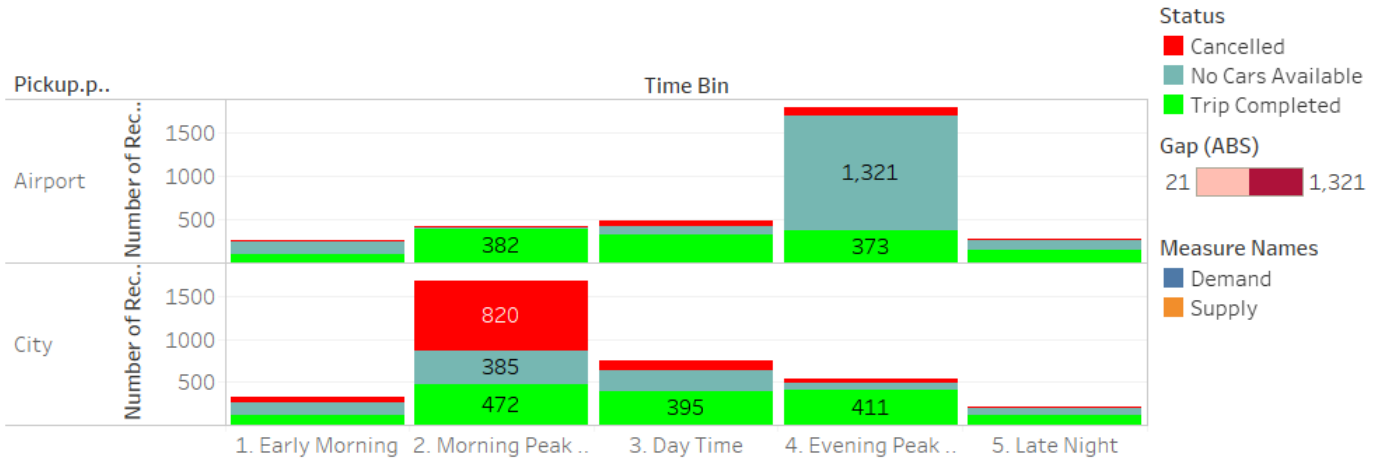


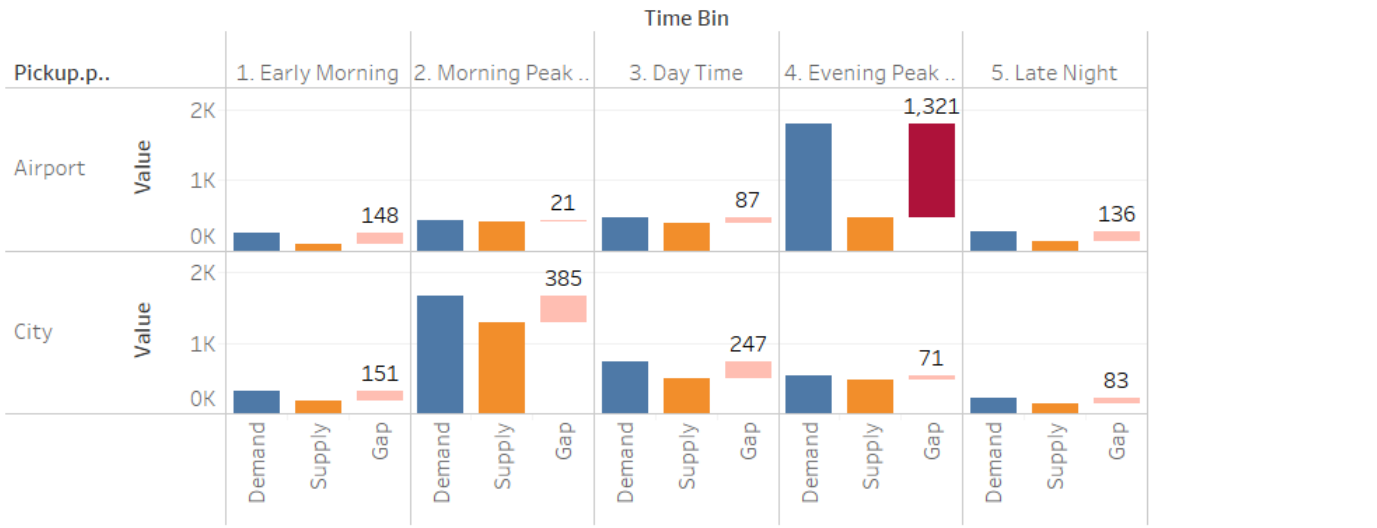
## Observations for Trip starting @ City

- This low demand at airport during morning time leads to high waiting time of cab drivers to get next trip. So cab drivers are rejecting requests for city to airport trips.

## Recommendations for Trip starting at City

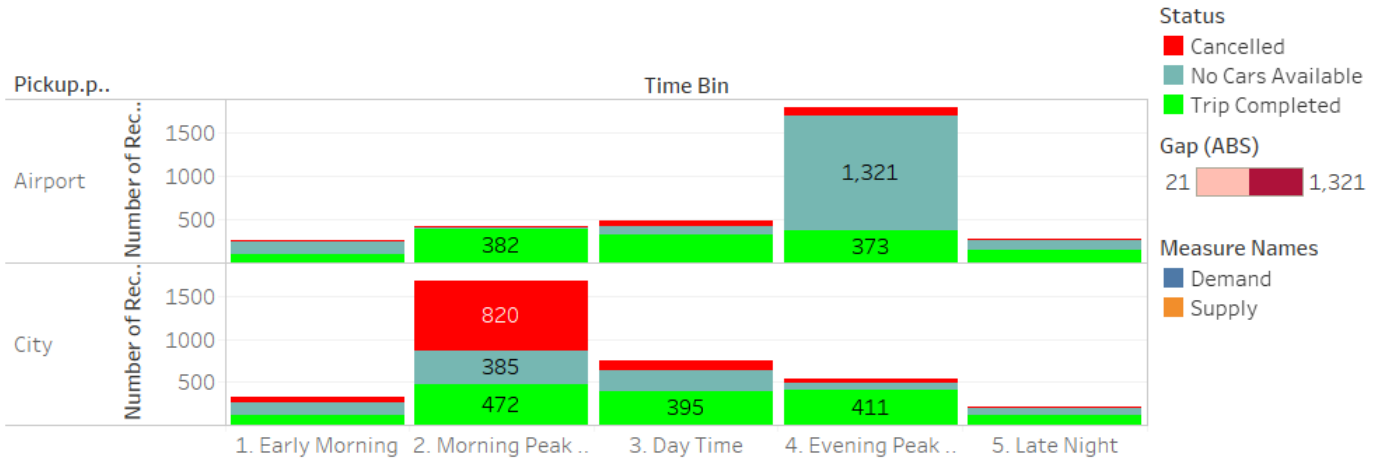
- UBER can rectify this low demand at airport by reducing price cuts for passengers, easy cab booking methods.
- Cab driver can be incentivized during morning hours by providing bonus, by setting a wait time threshold, beyond that time Uber can pay gas cost to drivers to come back to city without a ride.
- One solution could be Uber tie up with Airport services to bring back airport staffs from airport to city. This program can drastically reduce the low demand at airport.



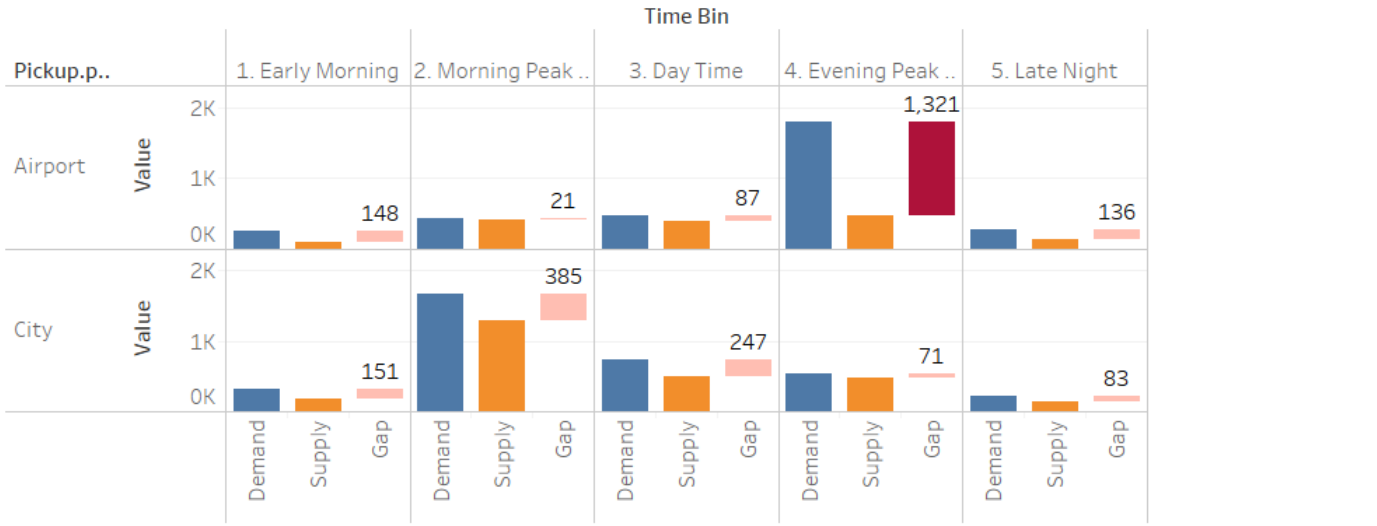



## Observations for Trip starting @ Airport

- Airport has more severe supply and demand gap as compared to city.
- During evening peak hours there is huge request but Uber loosing the business due to non availability of cabs.
- This high positive gap(more demand) is major area of investigation.
- One cause, is very less incoming cabs from city to airport during day and evening time, this leads to low availability of cabs at airport.
- Also plots shows, there is nice demand during day and evening time at city. Which keeps the cab drivers busy at city.
- Cancellation is not a major issue at airport.

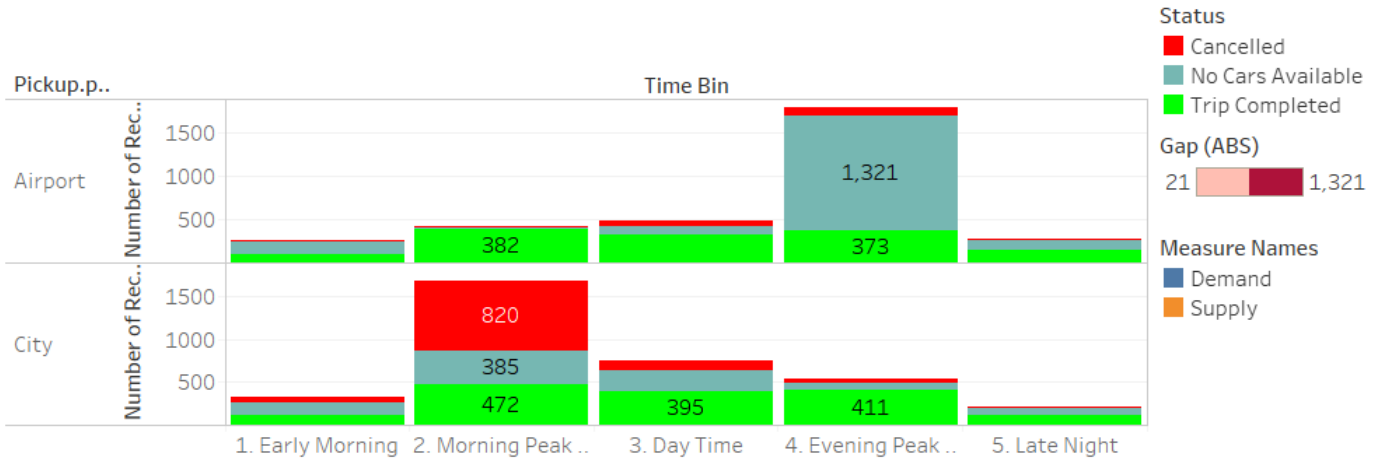


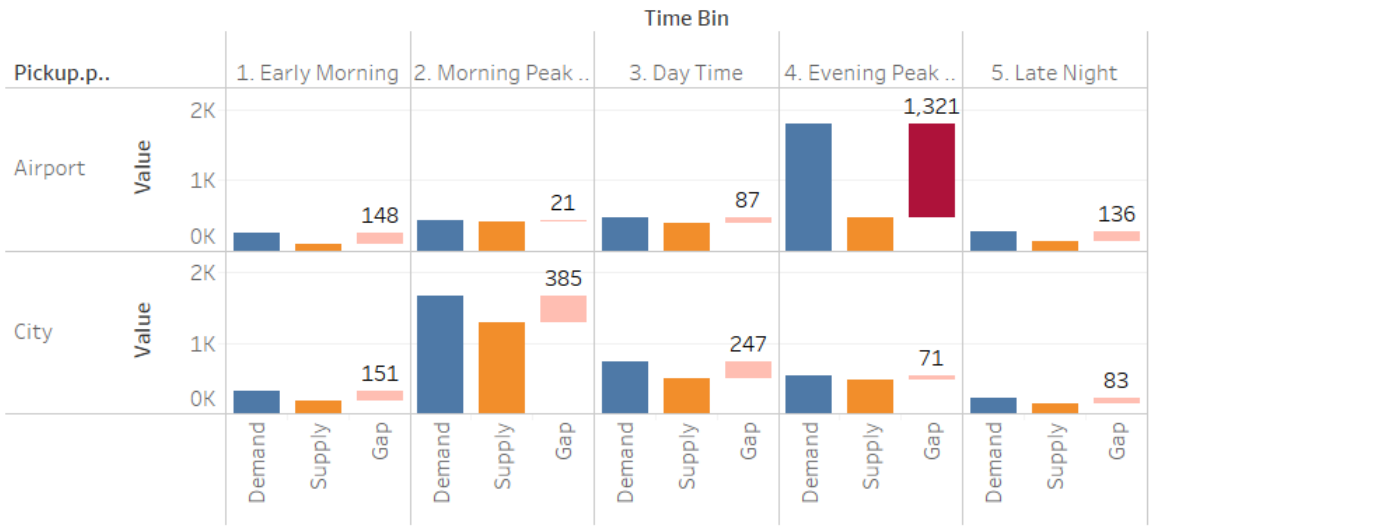

## Supply vs Demand GAP Analysis



## Recommendations for Trip starting @ Airport

- Drivers can be given a bonus to complete a trip from Airport to city at pick hours.
- Offer day time cost benefit to customer for city to airport trips so more cabs can be accumulated in airport before evening peak hours.
- Give cash incentives to riders and cab drivers for share rides from airport to city in evening hours.
- Uber can pay fuel cost to drivers who comes to airport without any passenger at evening hours.
- Uber should look into a long term solutions to increase more cabs into its fleet. This will solve it's supply and demand issues.



# Thank you..

- Tools Used
  - RStudio for data import and cleaning. Modeling, Analysis and Visualization.
  - Tableau for Visualization.
  - Microsoft Power Point for presentation.
- Assumptions
  - Cancellations are done by drivers not by riders.
  - Definition of Demand= Trip Completed+ Cancelled + No Cabs Available
  - Definition of Supply= Trip Completed + Cancelled.
  - City cabs may have different trip destination but for our analysis all trips originating from city is to airport.
  - Cabs at airport will wait till they get a trip.