

# UBER CASE STUDY

Presenter:

Name: Mohit Rathore

Roll No: DDA1710238

## Abstract

The aim of analysis is to identify the root cause of the problem (i.e. cancellation and non-availability of cars) and recommend ways to improve the situation. As a result of my analysis, I will be able to present to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

# Problem solving methodology

## 1. Data Cleaning:

- Set Date & Time data in a consistent format using `parse_date_time` & `strptime` functions.
- Derived new variables for analysis, this includes
  - TripTime: Difference between drop & pickup timestamps in minutes
  - RequestMeet: 1 if status is “Trip Completed” & 0 otherwise

## 2. Univariate & Bivariate Analysis: During Univariate analysis found gap between demand & supply & average trip times between airport & city. During Bivariate Analysis found relationship between following variables using Tableau & R EDA & Visualization techniques

- Found & visualized relationship between no of cancellations & time of day especially during late evening & early morning when major fluctuations occur in Tableau.
- Found & visualized relationship between time of day across all 5 days & identify patterns of demand supply gap.
- Based on these visualizations found root cause of problem & proposed a hypothesis & solution.

## Univariate Analysis

1. Analyzing Demand Supply Gap: Using **Request Meet** column I found ration of total number of cabs canceled + no cabs available to Total records which indicated that only **42 percent** of requests were meet out of which around 40.4 % were meet when pickup point was airport & 42.9 % when pickup point was City.
2. Average Trip Time: Using **Trip Time** column an average Trip Time(one way) between airport & city was found to be **52.4mins**.

## Bivariate & Multivariate analysis

Relationships between variables was found using plots(shown in Results slide). Following Dimensions & Measures were compared:

- a) Total count of status[Cancelled, Unavailable, Trip completed](Dimension) vs Request timestamp by hour of day (Measure)
- b) Total count of status(Dimension) vs Request timestamp by Date(Measure) vs Pickup point[Airport, City](Dimension)
- c) Average Trip Time(Measure) vs Drop Timestamp by hour of day (Measure) vs Pickup point(Dimension)
- d) Average Trip Time(Measure) vs Driver Id(Dimension)

I did basic plots in R using ggplot & all plots above in Tableau, this visualization provided main insights & conclusions I drew from given data.

# Analysis

Tools Used for Analysis:

1. R: I used R for Data Cleaning & Data preparation part, this involved below

- Removing irrelevant columns(Request Id)
- Bringing data in consistent format(Date Columns)
- Deriving new variables(TripTime & RequestMeet)

After data cleaning basic univariate analysis was done in R like finding trip time & calculating supply demand gap.

2. Tableau: I used tableau for most EDA & visualization as shown in Result slides.

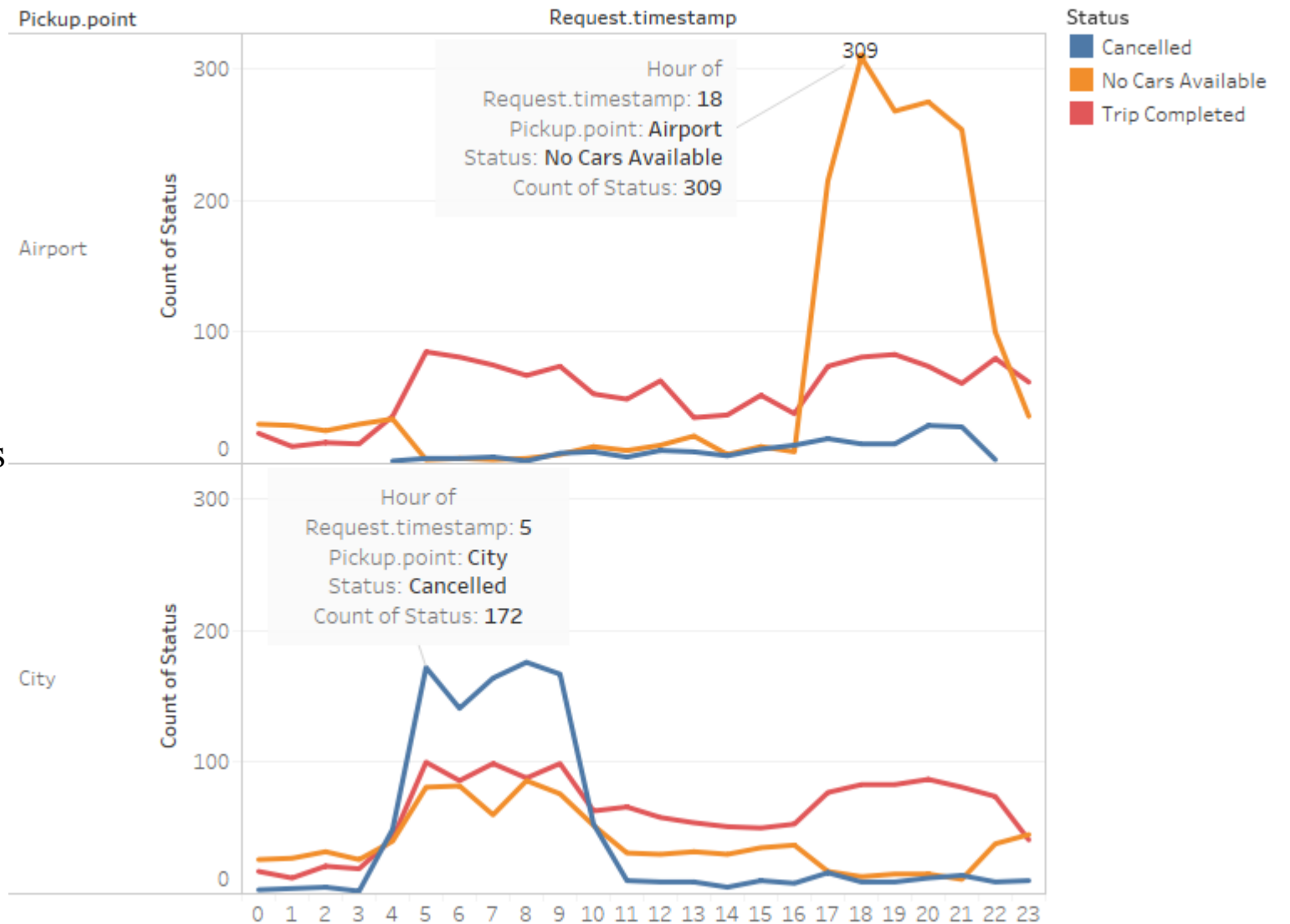
# Results

Plot 1:

First Plot clearly shows a spike in count Of “No Cars Available” during evening

Second Plot shows a similar spike in “Cancellations” during early morning hours

Count of Booking Status by Hours of Day

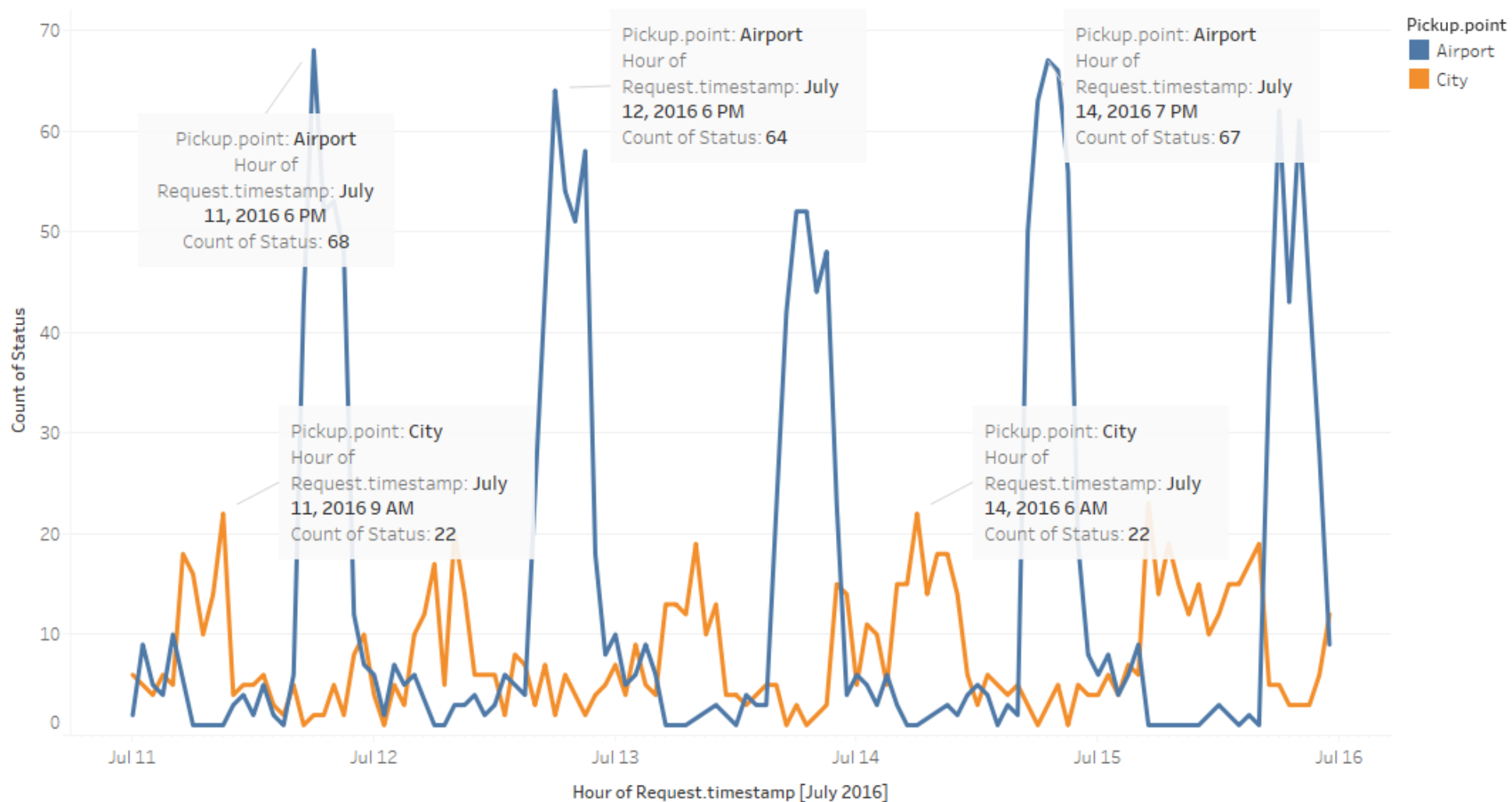


The trend of count of Status for Request.timestamp Hour broken down by Pickup.point. Color shows details about Status.

# Results

Plot 2:  
This Plot shows a  
Similar Pattern  
throughout the week  
for unavailability  
of cabs.

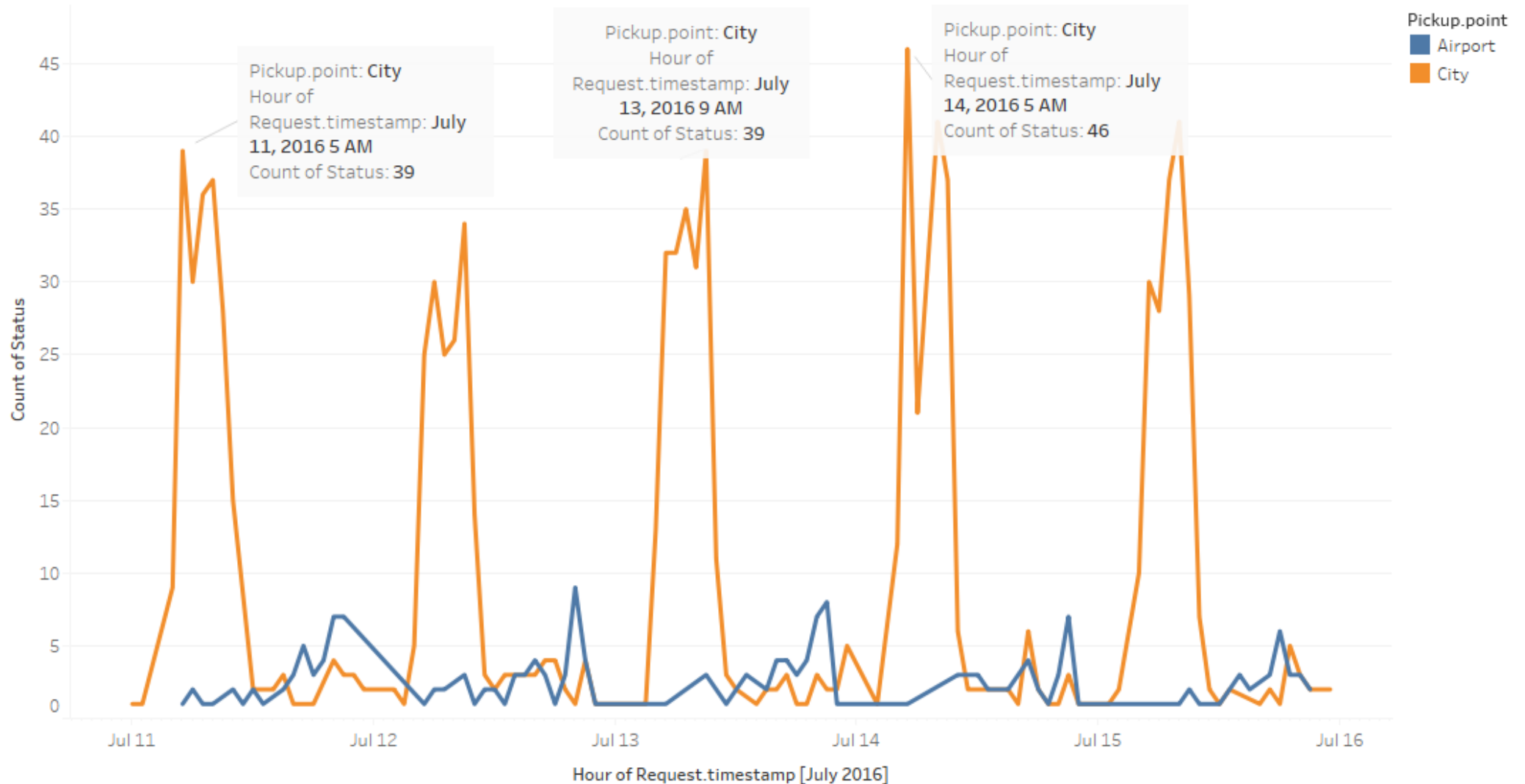
Cars unavailability throughtout week



The trend of count of Status for Request.timestamp Hour. Color shows details about Pickup.point. The data is filtered on Status, which keeps No Cars Available.



Trips cancelled throughout week



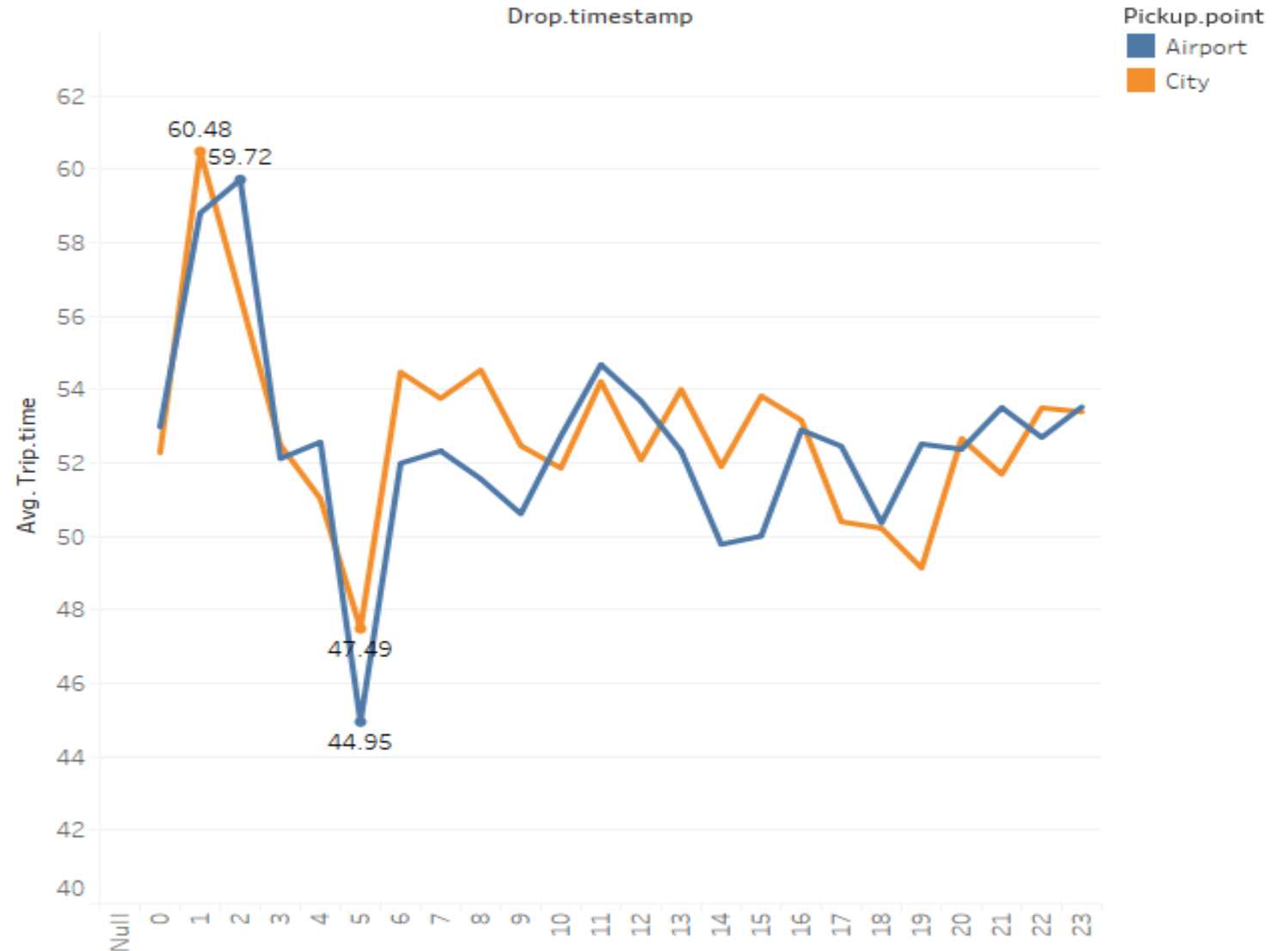
The trend of count of Status for Request.timestamp Hour. Color shows details about Pickup.point. The data is filtered on Status, which keeps Cancelled.

Plot 3:  
This Plot shows a  
Similar Pattern  
throughout week  
for Cancellation  
of cabs for cabs  
from city

Plot 4:

Plot showing fluctuations in Average Trip time during early morning hours,  
From sudden rise to sudden dip

Average Trip Time of Drivers from Airport & City

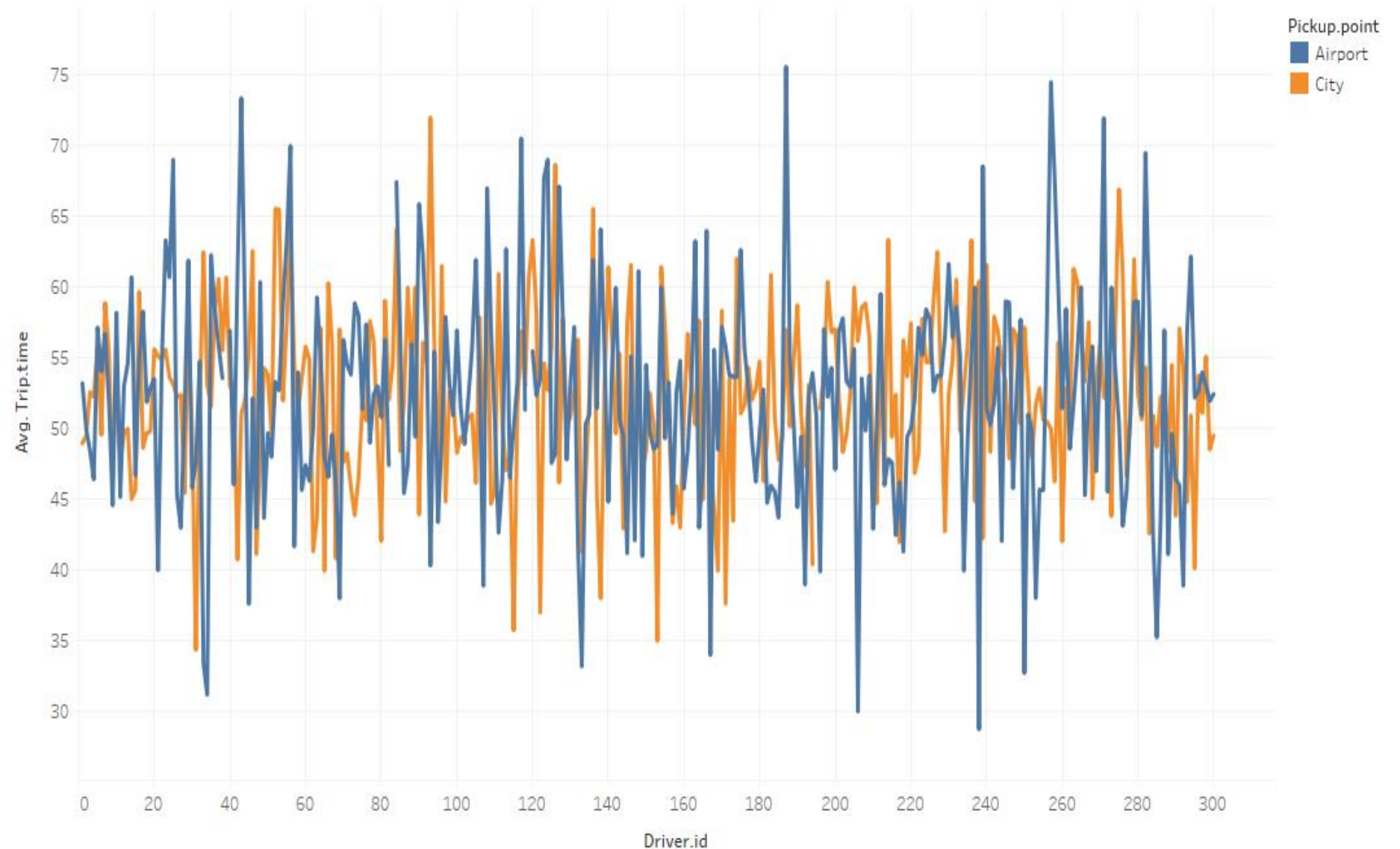


The trend of average of Trip.time for Drop.timestamp Hour. Color shows details about Pickup.point.

Plot 5:

Plot showing fluctuations in Average Trip time for different Drivers, Avg. Trip Time for pickup from Airport is more fluctuating compared to City Pickup which may indicate more Waiting Time at airports

Average Trip Time by Driver Id



The trend of average of Trip.time for Driver.id. Color shows details about Pickup.point.

# Conclusions

## Observations:

1. Availability of cabs from Airport to City drops after evening(after 6 PM to 11 PM) as shown in Plot 1 & Plot 2 .
2. Number of Cancellations from City to Airport increase in early morning hours (4AM to 9AM) as shown in Plot 1 & Plot 3.
3. Average Trip Time fluctuates sharply between 1-5 AM as shown in Plot 4.
4. Only 42% of cab requests were fulfilled.
5. Average Trip Time from Airport to City is larger compared to Trip time from City to Airport as shown in Plot 5.
6. There is an increase in Cancellations, Cab Unavailability & Trips Completed from City to Airport during early morning hours(4AM to 9AM) as shown in Plot 1.

## **Hypothesis** based on above observations:

1. Supply of cabs from airport to city may be reducing(6PM to 11PM) due to increased waiting time at night which suggests a flight pattern as domestic flight traffic reduces during night leading to higher waiting time at airport, so drivers avoid taking trips to airport during these hours.
2. Number of Cancellations from City to Airport increases during early morning hours(4AM to 9AM) suggest sudden increase in demand during these hours due to domestic flight patterns & reduction in waiting time, now more drivers are willing to go to airport. This also justifies increase in number of trips completed from City to Airport during these hours.
3. Average trip time sharply drops from 60mins to 40mins between 1- 5 AM probably due to reduction in waiting time. As as demand suddenly increases during early morning waiting time reduces to zero. Also there is less traffic during early morning hours as Avg. Trip time climbs back to 50-55mins from 40mins after 7 AM and remains in this range till late night.
4. These fluctuations in late night & early morning seems to be independent of drivers as seen on Plot 5 Avg. Trip time of all drivers is more or less same rather there is a general trend of supply demand gap throughout the week(Plot 2 & 3).

**Recommendations** based on hypothesis:

As this is clear issue of supply demand gap Uber can reduce this gap using following strategies:

- a) As many drivers avoid going to airport during late hours due to increased waiting time, if they are paid for waiting time on airport only during these hours they will be willing to go to airport during late hours as their waiting time losses are compensated.
- b) Starting special cabs only between airport & city may reduce this gap as drivers now will not have any choice except these trips.
- c) Adding more cabs specially during early hours is must(4AM to 9AM) will be helpful as demand clearly rises during this time.
- d) Uber can further study international flight patterns to as these flights arrive in late night & early hours. This can help better plan cab moment during these hours such that waiting time is reduced & international flight passengers get a Uber cab from airport. They can also increase cab fare for these international passenger segment which may compensate for waiting time loss of drivers.