UBER's
Demand and Supply

Ravi Shankar Hela

Motivation-Business Goals

Often UBER's customers, face difficulty getting a cab from Airport to City and vice versa. If the drivers cancel the request of riders or if cars are unavailable, Uber loses out on its revenue.

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 your analysis, you should be able to present to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them. For this we formulate following questions to guide our analysis.

1. Is there a Demand Supply Gap?
2. How big is the Gap?
3. Given the data, which events contribute to Gap and to what extent?
4. How does the Gap vary across given factors like Pick UP location, time of the day.
5. Recommendarion to the management to address the Gap

Intro to Datasets

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 6 columns):
Request id          6745 non-null int64
Pickup point        6745 non-null object
Driver id           4095 non-null float64
Status              6745 non-null object
Request timestamp   6745 non-null object
Drop timestamp      2831 non-null object
dtypes: float64(1), int64(1), object(4)
memory usage: 316.2+ KB
```

```
Exploring missing values
Request id           0.000000
Pickup point         0.000000
Driver id            0.392884
Status               0.000000
Request timestamp    0.000000
Drop timestamp       0.580282
dtype: float64
```
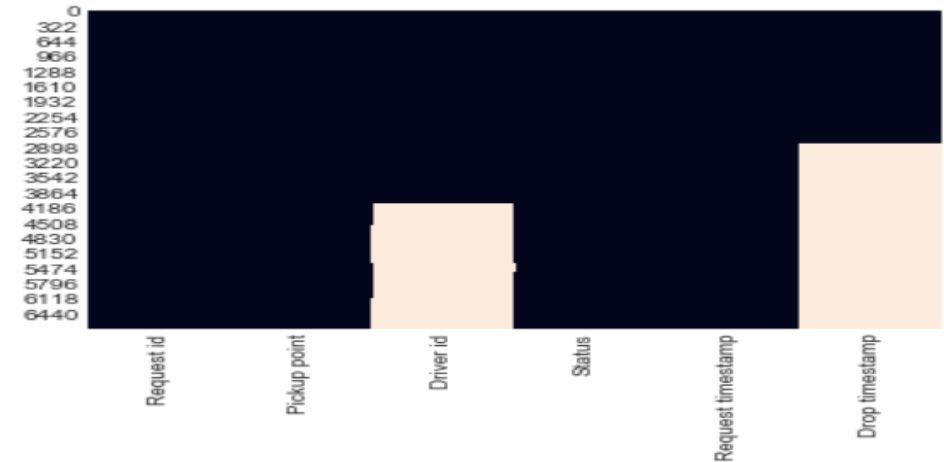
Data Quality

1. Inconsistent strings denoting dates with multiple formats standardized

```
0          11/7/2016 11:51
1          11/7/2016 17:57
2           12/7/2016 9:17
3          12/7/2016 21:08
4       13-07-2016 08:33:16
5       13-07-2016 21:57:28
6       14-07-2016 06:15:32
7       15-07-2016 05:11:52
8       15-07-2016 17:57:27
9           11/7/2016 6:46
Name: Request timestamp, dtype: object
```

```
making dates consistent
0       2016-07-11 11:51:00
1       2016-07-11 17:57:00
2       2016-07-12 09:17:00
3       2016-07-12 21:08:00
4       2016-07-13 08:33:16
5       2016-07-13 21:57:28
6       2016-07-14 06:15:32
7       2016-07-15 05:11:52
8       2016-07-15 17:57:27
9       2016-07-11 06:46:00
Name: Request timestamp, dtype: datetime64[ns]
```

2. Feature like Status and Pickup point converted to category

```
converting Pickup point and Status to category/factors
City        3507
Airport     3238
Name: Pickup point, dtype: int64

Trip Completed      2831
No Cars Available   2650
Cancelled           1264
Name: Status, dtype: int64
```

Missing Data Analysis

```
Exploring missing values
Request id          0.000000
Pickup point        0.000000
Driver id           0.392884
Status              0.000000
Request timestamp   0.000000
Drop timestamp      0.580282
dtype: float64
```
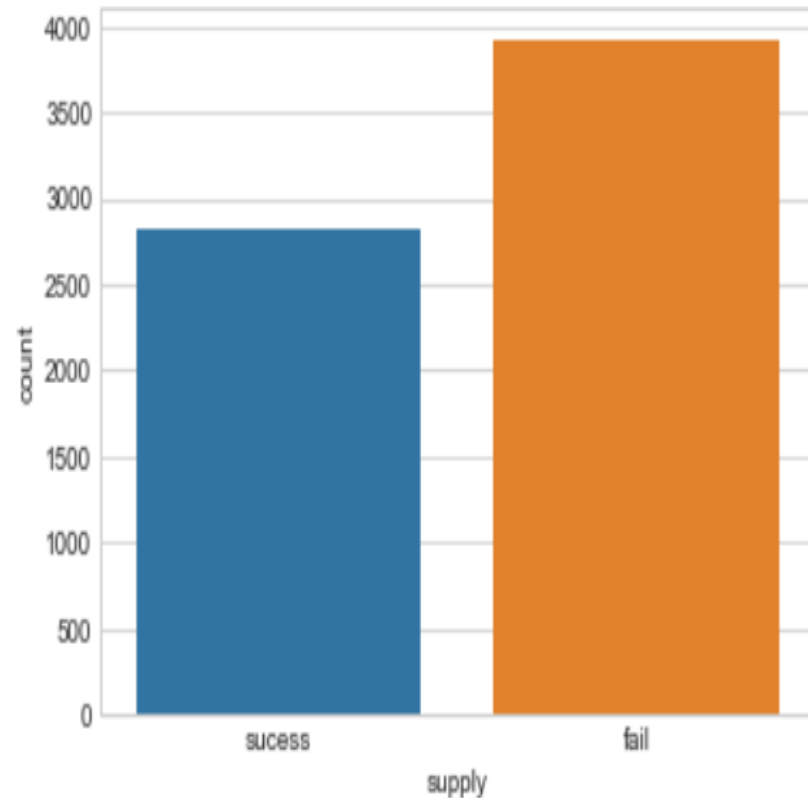
Driver id → Failed cases(Cancelletion)

Drop timestamp → Failed cases(Cancelletion, Cars Not available)

EDA – Insights from Dataset

Created a new feature "supply" to lump failed cases together as "Fail" and trip completed as "success". Using this new feature we determine demand supply gap.


```
Is there a Demand Supply Gap and if there is how big?
fail       3914
sucess     2831
Name: supply, dtype: int64
fail       0.580282
sucess     0.419718
Name: supply, dtype: float64
```
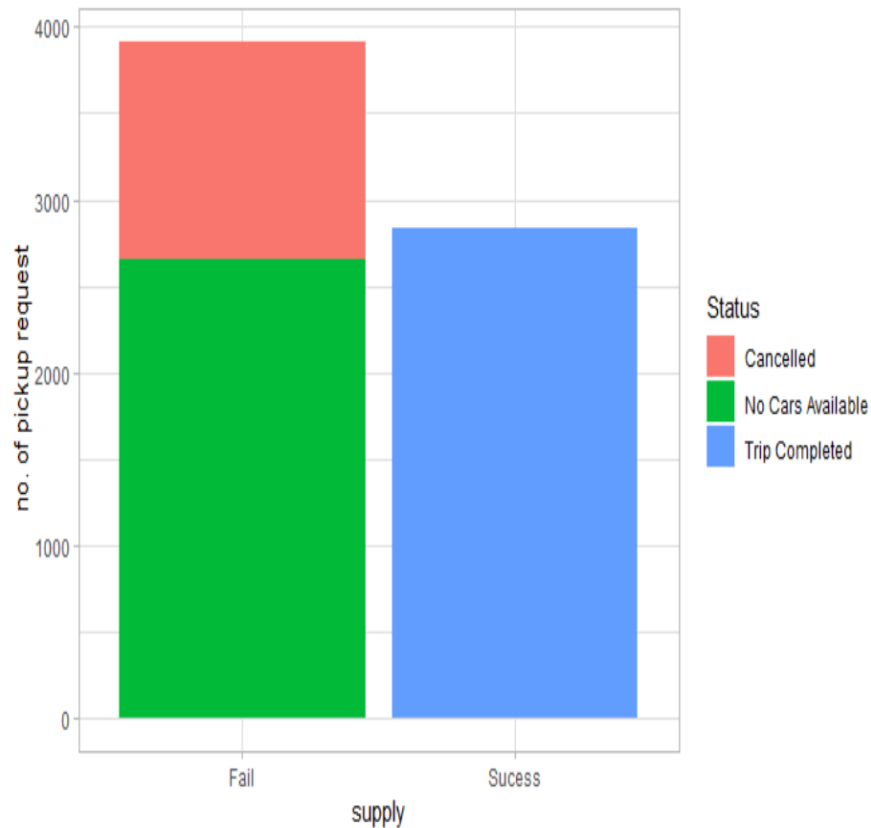
Yes, there is a demand supply gap

Overall Fail rate = 58%

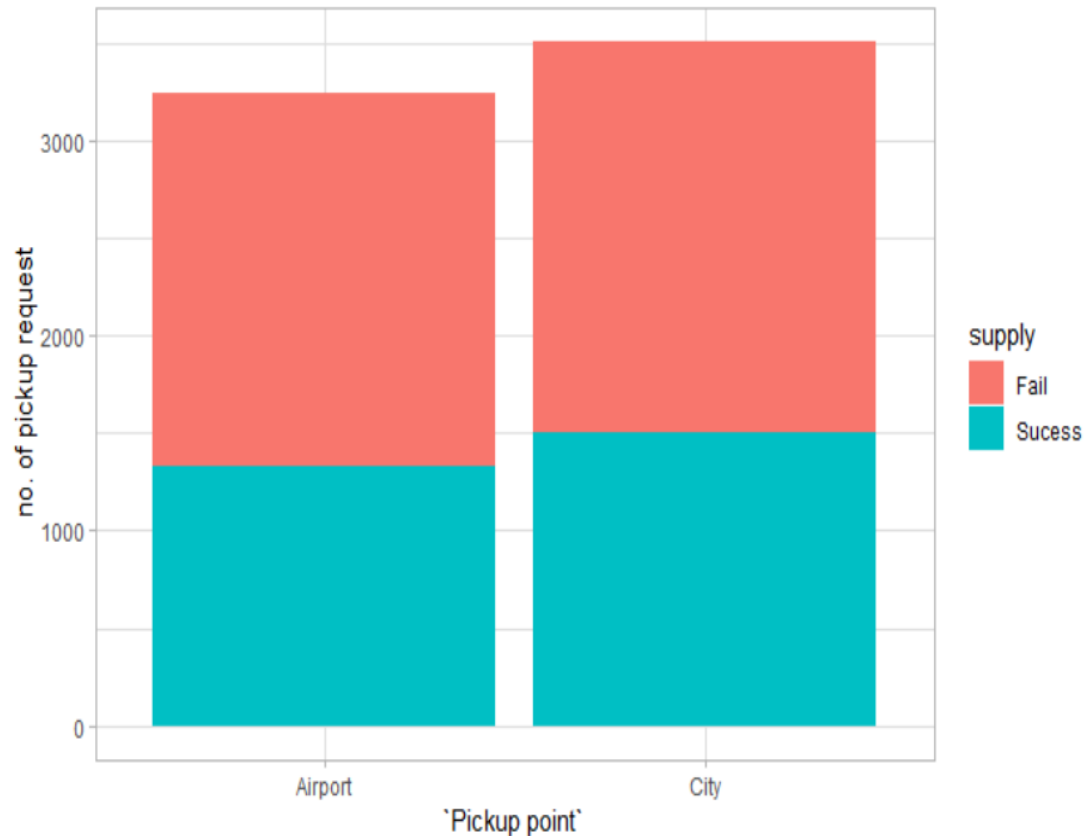# EDA – Insights from Dataset


```
Checking the nature of the fails
supply   Status
fail     Cancelled            0.32
         No Cars Available    0.68
sucess   Trip Completed       1.00
Name: supply, dtype: float64
```

No Cars Available contributing
to **68%** of Service failure

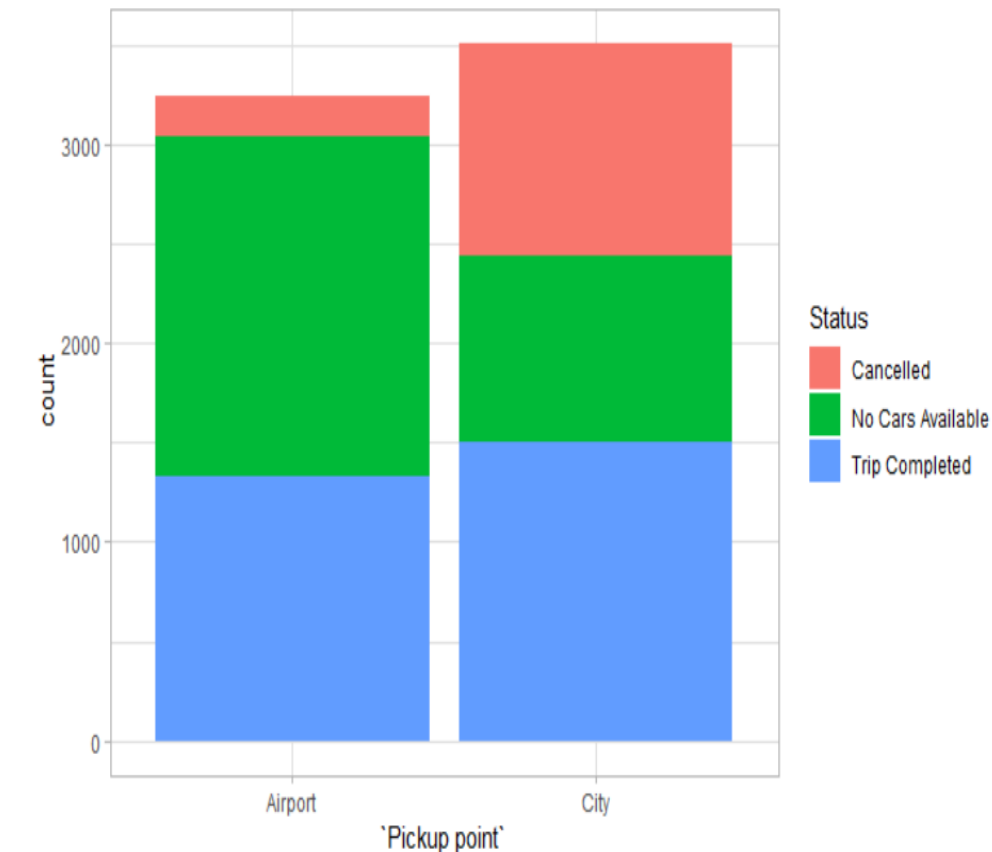# EDA – Insights from Dataset

Its difficult to say that Pick point influences the outcome of a success or failure.
In general the effect of Pick up point seems neutral.

We will explore later if this changes when we consider other factors along with Pick up point.

In general, **Pick up point** is Neutral to

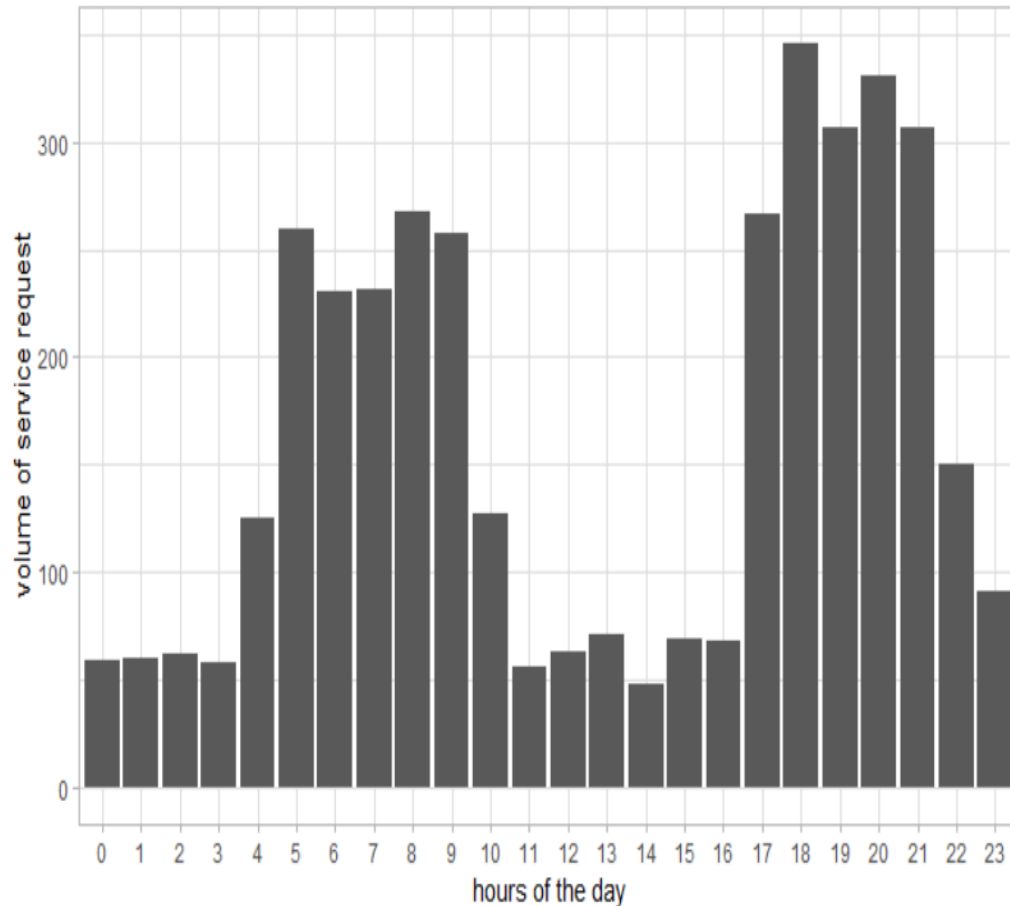Service failure or Is It?

EDA – Insights from Dataset

Even though failure was evenly distributed across pickup point, the nature of failure is different

In City we get more Cancellation

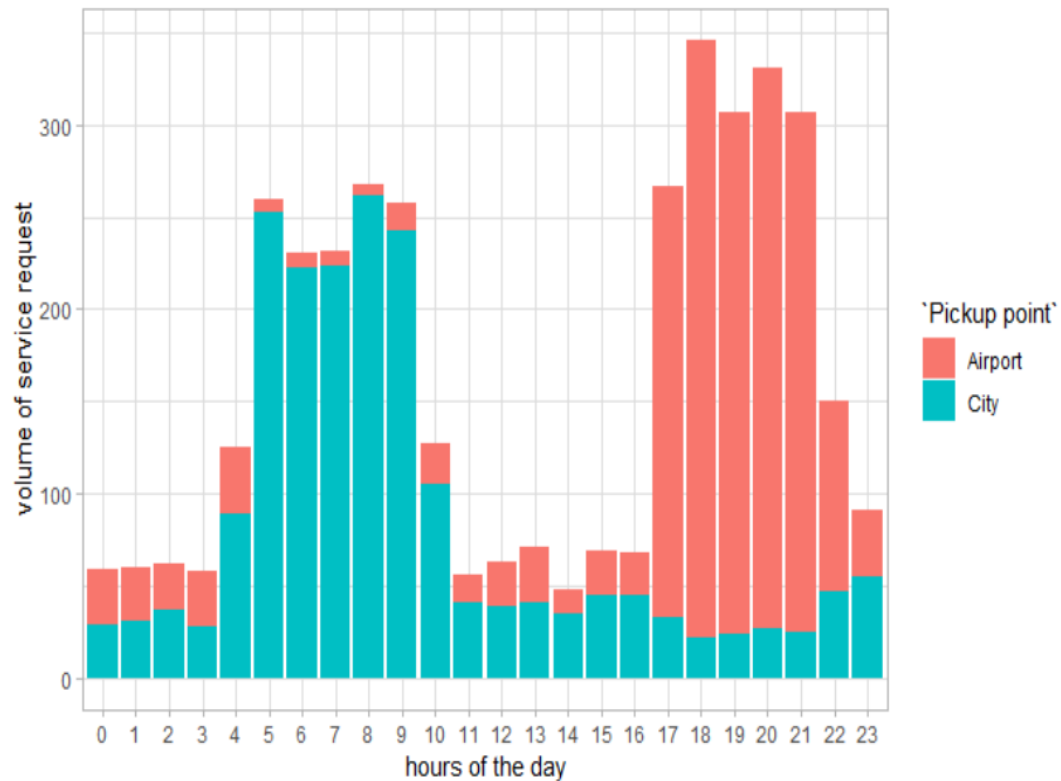In Airport we get Cars unavailable

EDA – Insights from Dataset

There are two demand peak where. The first one is 4:00 am to 9:00 am and second one is 5:00 pm to 10:00 pm

We have two **Demand Peak**
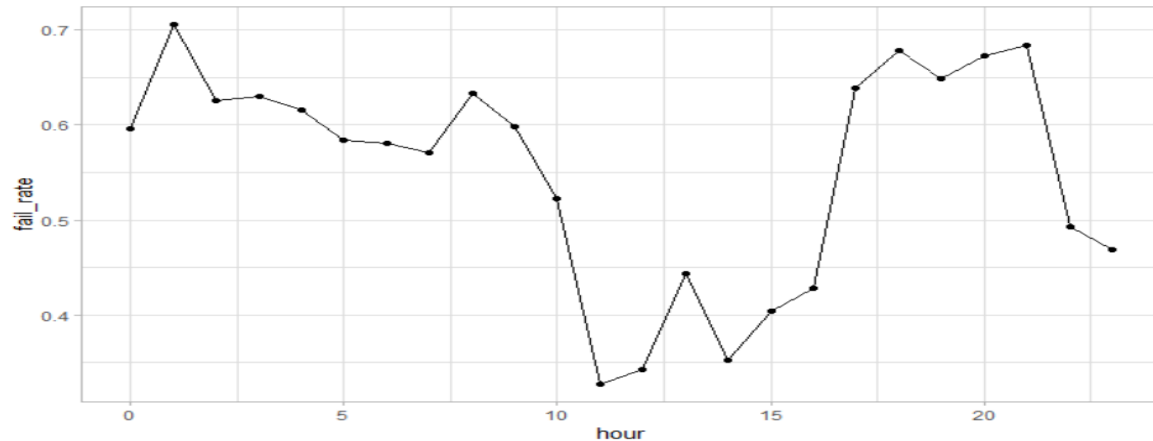
4:00AM-9:00AM

5:00PM-10:00PM

EDA – Insights from Dataset

The first one is 4:00 am to 9:00 am is for City Pickup, and second one is 5:00 pm to 10:00 pm is for Airport Pickup

We have two **Demand Peak**

1st Peak in City

2nd Peak in Airport
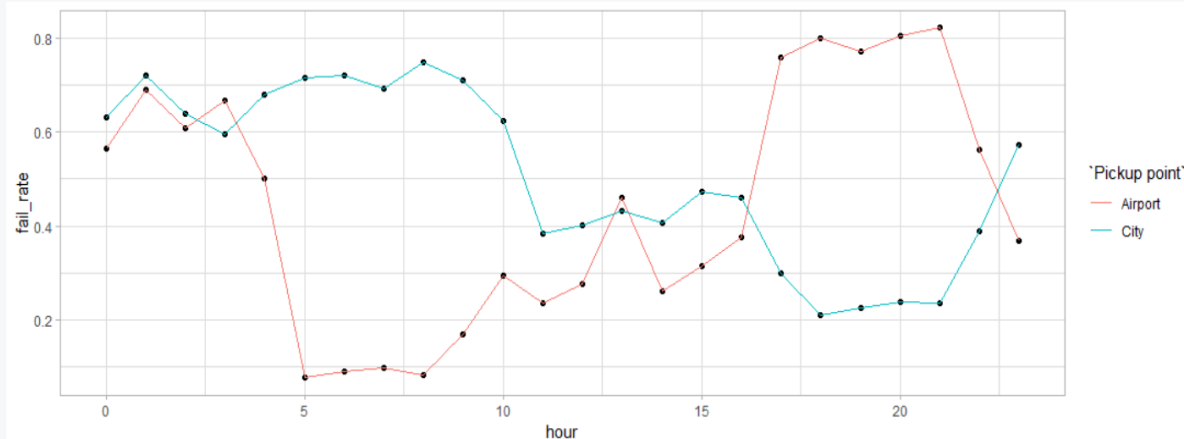
EDA – Insights from Dataset

We see that the Fail Rate is highest in late night

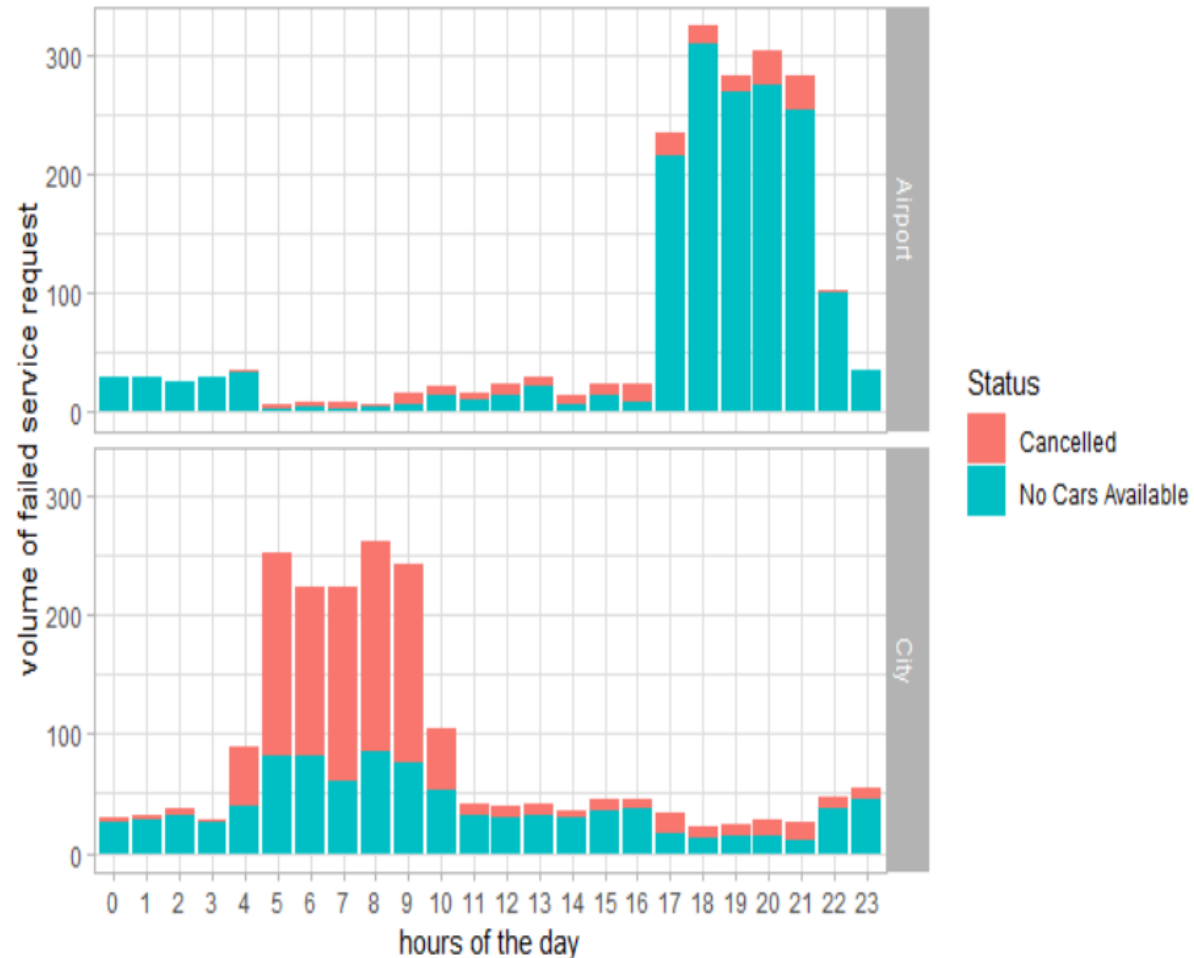The Fail comes down but stays relatively high during early hours till 10:00 AM.

Fail rate is also high during evening 5:00PM to late evening 9:00PM.

The pattern conforms to Demand Peaks. However, the morning fails is influenced by City and night fails influenced by Airport.

Failure Rate **Matches**
Demand Peaks

EDA – Insights from Dataset

The high Fail rate in the Morning is influenced by City Demand Peak where we observe more Cancelletion.

The late evening-night Fail rate is influenced by Airport Demand Peak and is almost attributed to Cars not Available.

City pickup fails due to
Cancellation
Airport pick fails
Cars unavailable at demand peak

Summary Insights:

Is there a Demand Supply Gap? *Yes*

How big is the Gap?

Overall the failure rate is 58%.

There are two demad peak where Fail rate is high in the range of 60-70%. The first one is 4:00 am to 9:00 am and second one is 5:00 pm to 10:00 pm

Given the data, which events contribute to Gap and to what extent?

Demand peaks contributes to most of the failures (exception to this is late night between 12:00 AM to 4:00am)

How does the Gap vary across given factors like Pick UP location, time of the day.

The demand peak across Pick up point is significant. The mornining Peak is for City and late evening and night peak is for Airport.

For the morning peak of the City major cause of failure is Cancellation whereas for night peak of Airport almost all failures are due to unavailability of Cars.

Recommendation:

The morning spike in failure rate is mostly cancellation for City Pick up.
We can hypothesis is that Driver's are reluctant to come to Airport from City. As we observe in Slide 11(red filed portion) there is very low demand for Airport pick up during that time. The diver will have to wait for a pick up or he may have to drive back to city where he gets a booking. This means loss for driver in term of Time and Fuel).

Management may incentivise drivers who accept these pick ups by charging premium on Customer Fare and passing the differential to drivers.
The night spike in failure rate is mostly unavailability of cars for Airport Pick up.
Owing to low demand from City where We can give same argument as above and same recommendation.
Other hypothesis could be that we have few drivers who do night shift. May recruit more drivers with higher incentives. Obviously charging the extra money to passenger fare.

Beside this we may also promote other avenue like allowing private cars to register in Ola for pooling. Obviously this might require regulatory permissions.