Assignment summary Report

By:Ravi Ranjan Kumar
Master of Science
University of Arizona

Uber Supply-Demand Gap

Objective:

Supply Demand Gap of Uber Bookings from Two pickup points : Airport and City

DataSet Descriptions:

- There are 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, can be either completed, canceled by the driver or no cars are available

Note: For this assignment, only the trips to and from the airport and city are being considered.

Exploratory Data Analysis

- **Info of Datatype in Dataset**

```
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Request id         6745 non-null   int64
 1   Pickup point       6745 non-null   object
 2   Driver id          4095 non-null   float64
 3   Status             6745 non-null   object
 4   Request timestamp  6745 non-null   object
 5   Drop timestamp     2831 non-null   object
dtypes: float64(1), int64(1), object(4)
```

- **Stastical info of Dataset**

	Request id	Driver id
count	6745.000000	4095.000000
mean	3384.644922	149.501343
std	1955.099667	86.051994
min	1.000000	1.000000
25%	1691.000000	75.000000
50%	3387.000000	149.000000
75%	5080.000000	224.000000
max	6766.000000	300.000000

- **Snapshot of Dataset:**

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47

- **Missing value present in dataset**

```
Request id              0
Pickup point            0
Driver id            2650
Status                  0
Request timestamp       0
Drop timestamp       3914
dtype: int64
```

➢ **In data set , missing of Driver id column and Drop timestamp column are 2650 and 3914.**

Exploratory Data Analysis

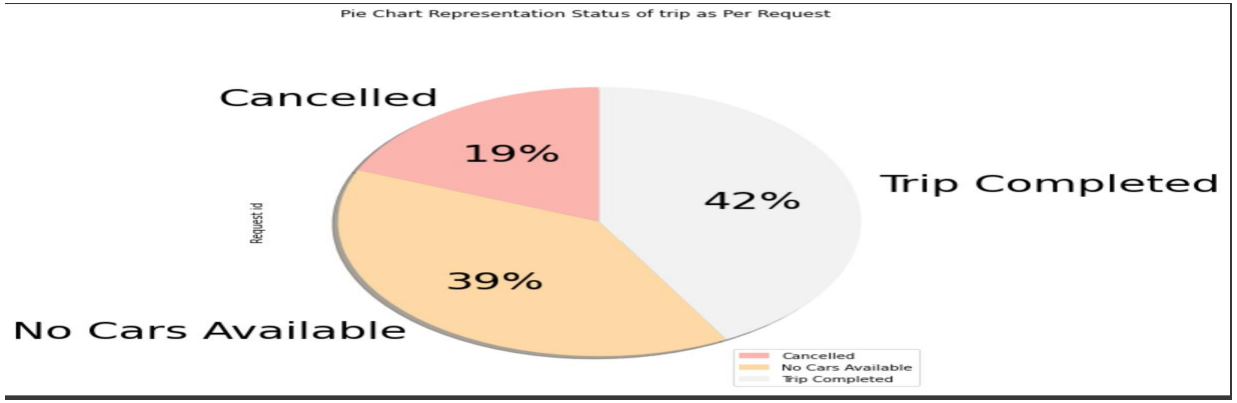
- Status of Trip

Status	Request id
Cancelled	1264
No Cars Available	2650
Trip Completed	2831

"Driver id" column has 2650 missing as shown above but this is equal to the "No cars Available" trip status, hence there are no missing values "Drop Timestamp" columns has 3914 missing values which is the sum of "No cars Available" and "Cancelled" (2650+1264=3914).

Therefore, there is no missing values in the data set

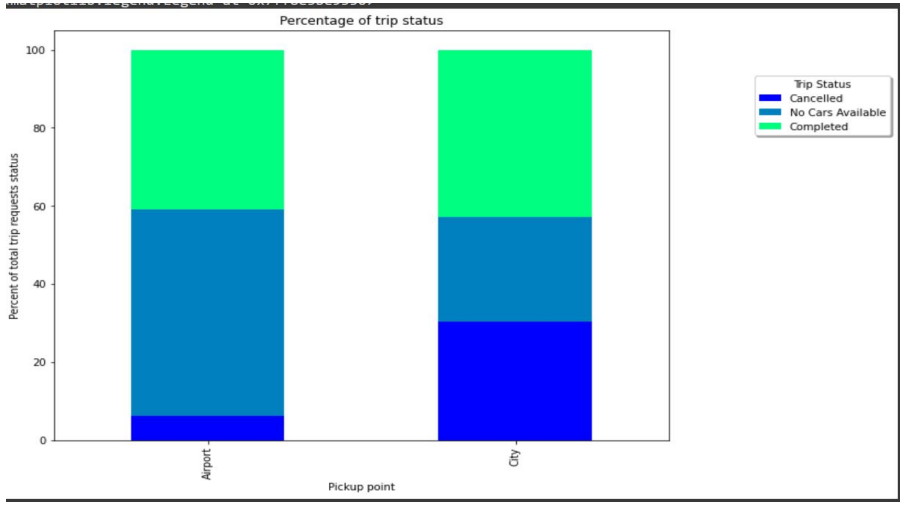
- **Demand and supply gap**



Above pie chart shows that There is a very big supply demand gap Only 42 percent of the total trip requests are being completed There are a total of 19 percent trips that are being cancelled by the driver There are 39 percent trip requests which are being turned down due to unavailabiltity of cars.

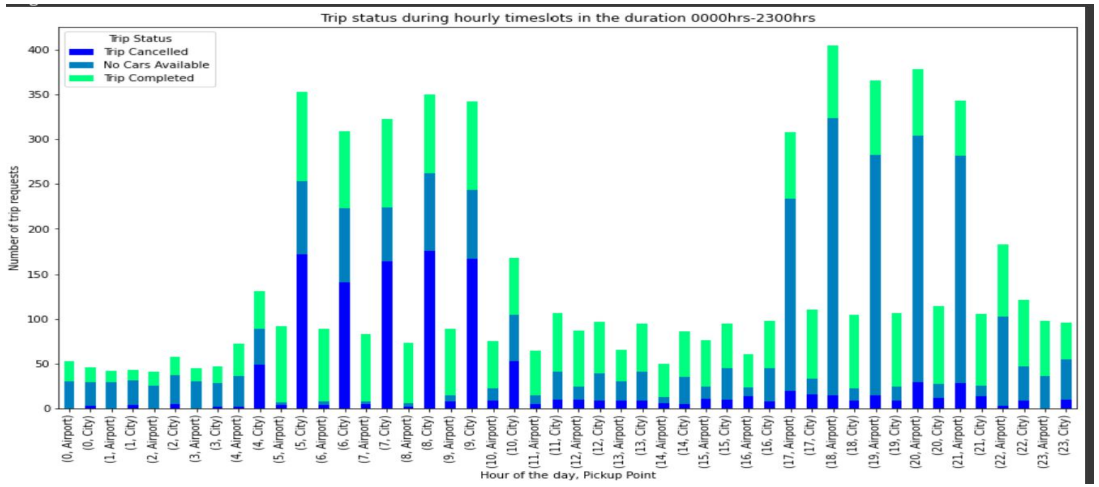
Expioratory Data Anaiysis

- **Trip Status w.r.t pickup point**



More problematic Pickup point is Airport with 59% (53%- non availability of cars and 6%- cancellations) supply demand gap City has 57 % (27%- non availability of cars and 30%- cancellations)supply demand gap

- **pickup point on hour basis w.r.t number of request**



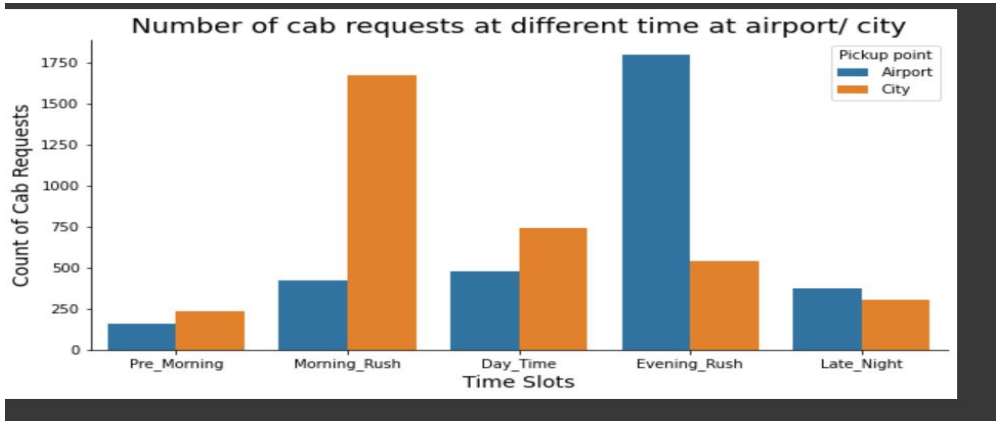
Here we can see that

There is surge number of trip completed and trip cancelled as per request in the morning(5-9AM) from pick up point city.
There is large number of car unavailable at pick up point Airport at night(5-9PM)

Exploratory Data Analysis

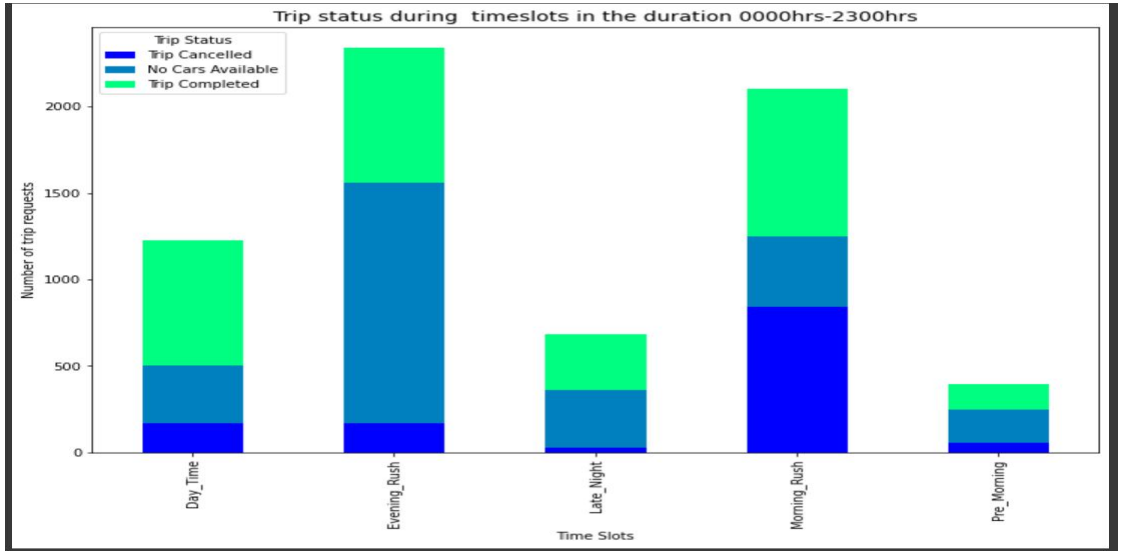
Let's divide the hours in slots for various time period of the day

- 2am - 5am: Pre_Morning
- 5am - 10am: Morning_Rush
- 10am - 5pm: Day_Time
- 5pm - 10pm: Evening_Rush
- 10pm - 2am: Late_Night



The morning rush is due to the cab requests in the city and the evening rush is due to cab requests at the airport.

- **TripStatus w.r.t diffferent time slots**



A large number of requests are denied in the morning.

By comparing both the graphs, it can be suggested that the less number of cabs are going to the airport from city and thus very less number of cabs are available at the airport in the evening.

Understanding the supply-demand gap

- Demand: Total number of requests for cab rides Supply: Total number of requests completed
- We will look at the difference between the supply and demand for both the locations.

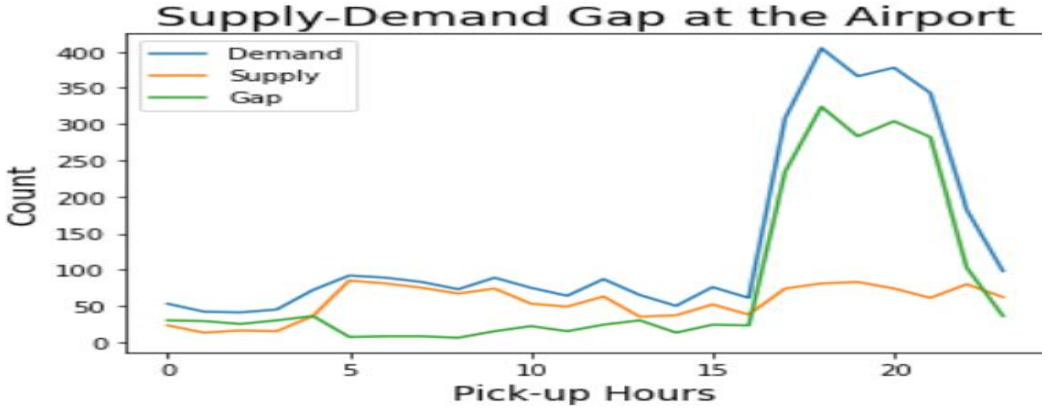
Count of Status of trip in different timeslots w.r.t number of request

	Cancelled	No Cars Available	Trip Completed	%cancelled	%NoCarsAvail	%TripCompleted	%Failed_req	failed requests	total_requests
Day_Time	203.0	566.0	873.0	12.0	34.0	53.0	47.0	769.0	1642.0
Evening_Rush	178.0	1530.0	938.0	7.0	58.0	35.0	65.0	1708.0	2646.0
Late_night	0.0	0.0	0.0	NaN	NaN	NaN	NaN	0.0	0.0
Morning_Rush	905.0	471.0	970.0	39.0	20.0	41.0	59.0	1376.0	2346.0
Pre_Morning	234.0	271.0	334.0	28.0	32.0	40.0	60.0	505.0	839.0

Supply -Demand Gap

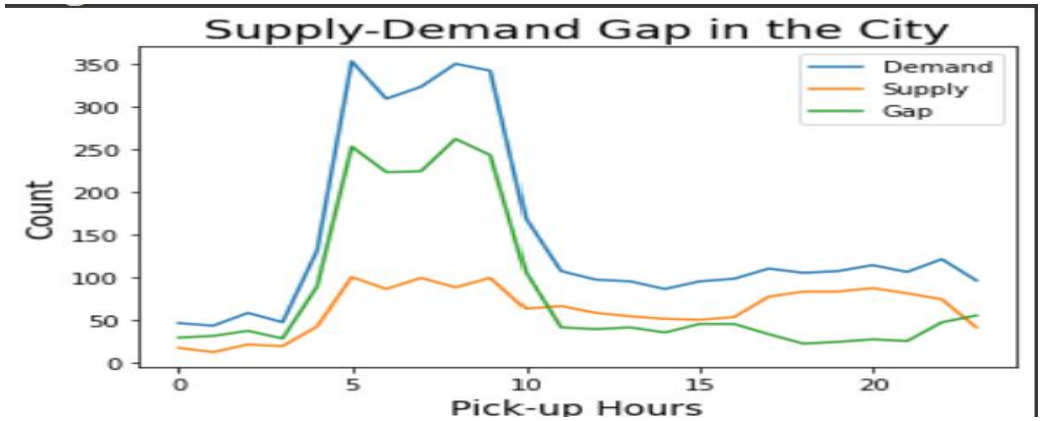
Demand-supply gap : it is difference between demand and supply at different pickup point

Deamand supply gap at the Airport w.r.t different pick up hours



Above graph shows that there is huge gap between supply and demand in the interval of 5pm to 9pm at Airport.

- **Demand supply gap at the city w.r.t different pick up hours**



Above graph shows that there is huge gap between supply and demand in the interval 5Am to 9Am at city pick up point

Data Insight Observastion

01 Considering timeslots with highest supply demand gaps from airport to city i.e. 5 to 10 pm only 70-80 requests each side are completed from City-Airport and Airport-City, it means the reason "no cars available" is not completely due to half of those cars are in transit but it indicates that all 300 cars are not at service even when there is such a high demand of total approx. 500 requests from both points.

02 Total 300 cars 300 unique driver ids i.e. 300 cars in total and the highest requests at one timeslot from all pickup ponts is as high as 500 recorded at 1800hrs.

03 Airport : Upon analysis, the evening time slot seems to be most problematic for pickup points as airport as the requests are not served due to unavailability of cabs. The unavailability of cabs at the airport in the evening slot may be due to the less number of cabs travelling from city to the airport (less number of request in the city at evening).

04 City : As per the analysis, the morning time slot is most problematic where the requests are being cancelled. Most probably the requests are being cancelled by the drivers the reason being they have to wait for a long time or return empty, as their are few request for cab at the airport in the morning

Recommendation

01 Provision of more cars because demand goes as high as 500 at late evenings.

02 Addressing the issue of cars not being at service to customers at all, evident from the fact that only 120-170 requests are completed even when there are total 300 cars existing for service

03 For bridging the demand supply gap from airport to city in evening, more cabs should be present at the airport or cab pooling facility should be started.

04 In morning, Uber can provide return compensation to the driver if had to return empty, this will result in less cancellation of requests by drivers in the morning.

Dataset : Uber Request Data

NotebookLink: https://colab.research.google.com/drive/1YtFKXy1IU29z7Q_6YxW226fyx2DuDSkP#scrollTo=lIUR1ZmzcWW0

Thank you!!!