UBER
SUPPLY-DEMAND GAP

Chu Bao

In Uber Request Data.csv, we have 6 attributes:

(left to right)

Request id: A unique identifier of the request

Pick-up point: The point from which the request was made

Driver id: The unique identification number of the driver

	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
...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

6745 rows × 6 columns

Status of the request: The final status of the trip

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

Data Formatting

bring date data to one uni-format

```
<class 'pandas.core.frame.DataFrame'>
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)
memory usage: 316.3+ KB
```

	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
...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

6745 rows × 6 columns

```
<class 'pandas.core.frame.DataFrame'>
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   datetime64[ns]
 5   Drop timestamp     2831 non-null   datetime64[ns]
dtypes: datetime64[ns](2), float64(1), int64(1), object(2)
memory usage: 316.3+ KB
```

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	2016-07-11 11:51:00	2016-07-11 13:00:00
1	867	Airport	1.0	Trip Completed	2016-07-11 17:57:00	2016-07-11 18:47:00
2	1807	City	1.0	Trip Completed	2016-07-12 09:17:00	2016-07-12 09:58:00
3	2532	Airport	1.0	Trip Completed	2016-07-12 21:08:00	2016-07-12 22:03:00
4	3112	City	1.0	Trip Completed	2016-07-13 08:33:16	2016-07-13 09:25:47
...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	2016-07-15 23:49:03	NaT
6741	6752	Airport	NaN	No Cars Available	2016-07-15 23:50:05	NaT
6742	6751	City	NaN	No Cars Available	2016-07-15 23:52:06	NaT
6743	6754	City	NaN	No Cars Available	2016-07-15 23:54:39	NaT
6744	6753	Airport	NaN	No Cars Available	2016-07-15 23:55:03	NaT

6745 rows × 6 columns

Data Mining

create new categorized data and patterns for analysis

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request Date	Request Hour	Time of day	Request Weekday
0	619	Airport	1.0	Trip Completed	2016-07-11 11:51:00	2016-07-11 13:00:00	2016-07-11	11	Morning	Monday
1	867	Airport	1.0	Trip Completed	2016-07-11 17:57:00	2016-07-11 18:47:00	2016-07-11	17	Afternoon	Monday
2	1807	City	1.0	Trip Completed	2016-07-12 09:17:00	2016-07-12 09:58:00	2016-07-12	9	Morning	Tuesday
3	2532	Airport	1.0	Trip Completed	2016-07-12 21:08:00	2016-07-12 22:03:00	2016-07-12	21	Evening	Tuesday
4	3112	City	1.0	Trip Completed	2016-07-13 08:33:16	2016-07-13 09:25:47	2016-07-13	8	Morning	Wednesday
...	...	...	...	...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	2016-07-15 23:49:03	NaT	2016-07-15	23	Midnight	Friday
6741	6752	Airport	NaN	No Cars Available	2016-07-15 23:50:05	NaT	2016-07-15	23	Midnight	Friday
6742	6751	City	NaN	No Cars Available	2016-07-15 23:52:06	NaT	2016-07-15	23	Midnight	Friday
6743	6754	City	NaN	No Cars Available	2016-07-15 23:54:39	NaT	2016-07-15	23	Midnight	Friday
6744	6753	Airport	NaN	No Cars Available	2016-07-15 23:55:03	NaT	2016-07-15	23	Midnight	Friday

6745 rows × 10 columns

New attributes:

- Request Date
- Request Hour
- Time of day
- Request Weekday

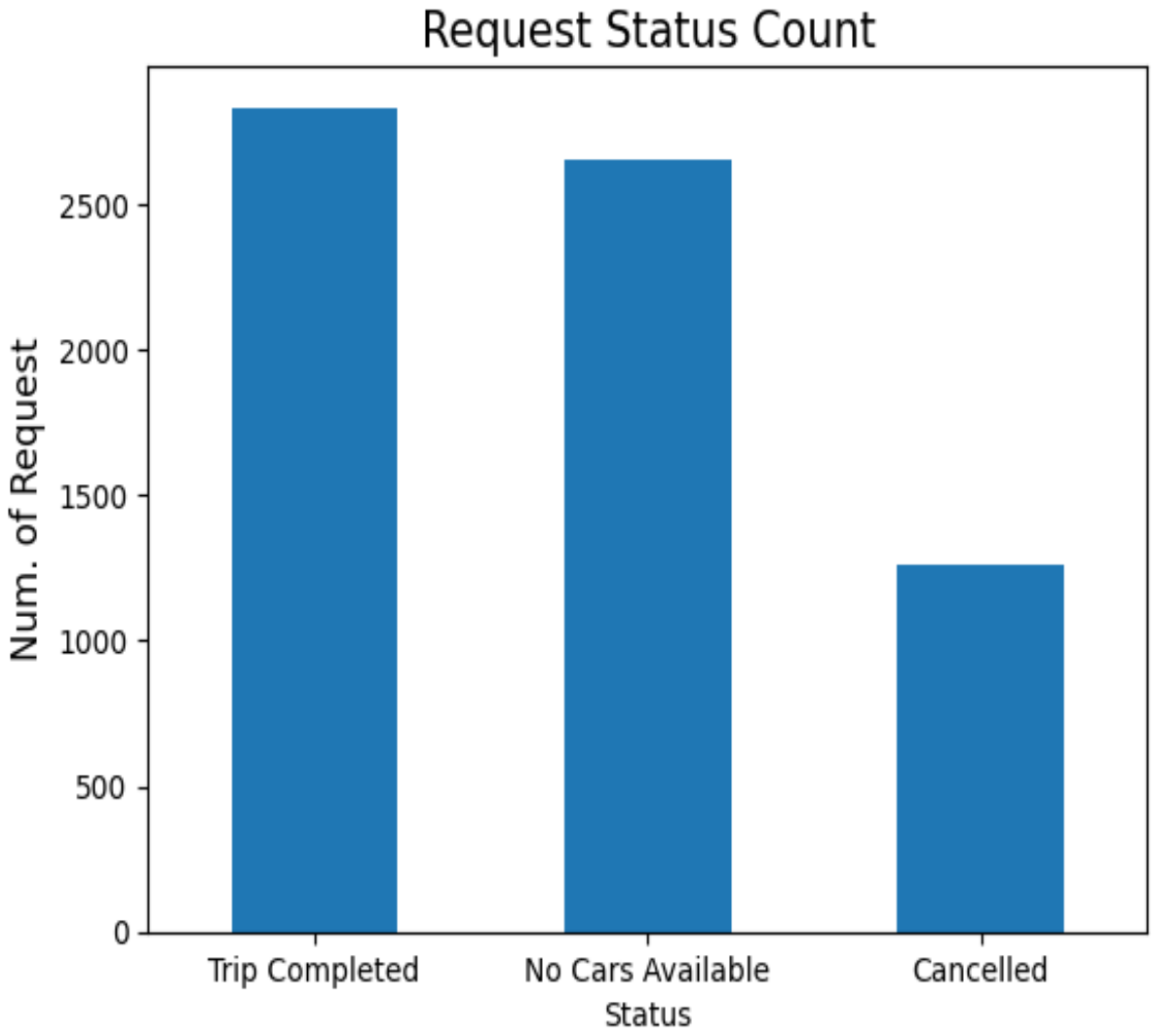
Trip Status

Trip Completed	Trip Cancelled	No Cars Available
2831	2650	1264
42%	39.3%	18.7%

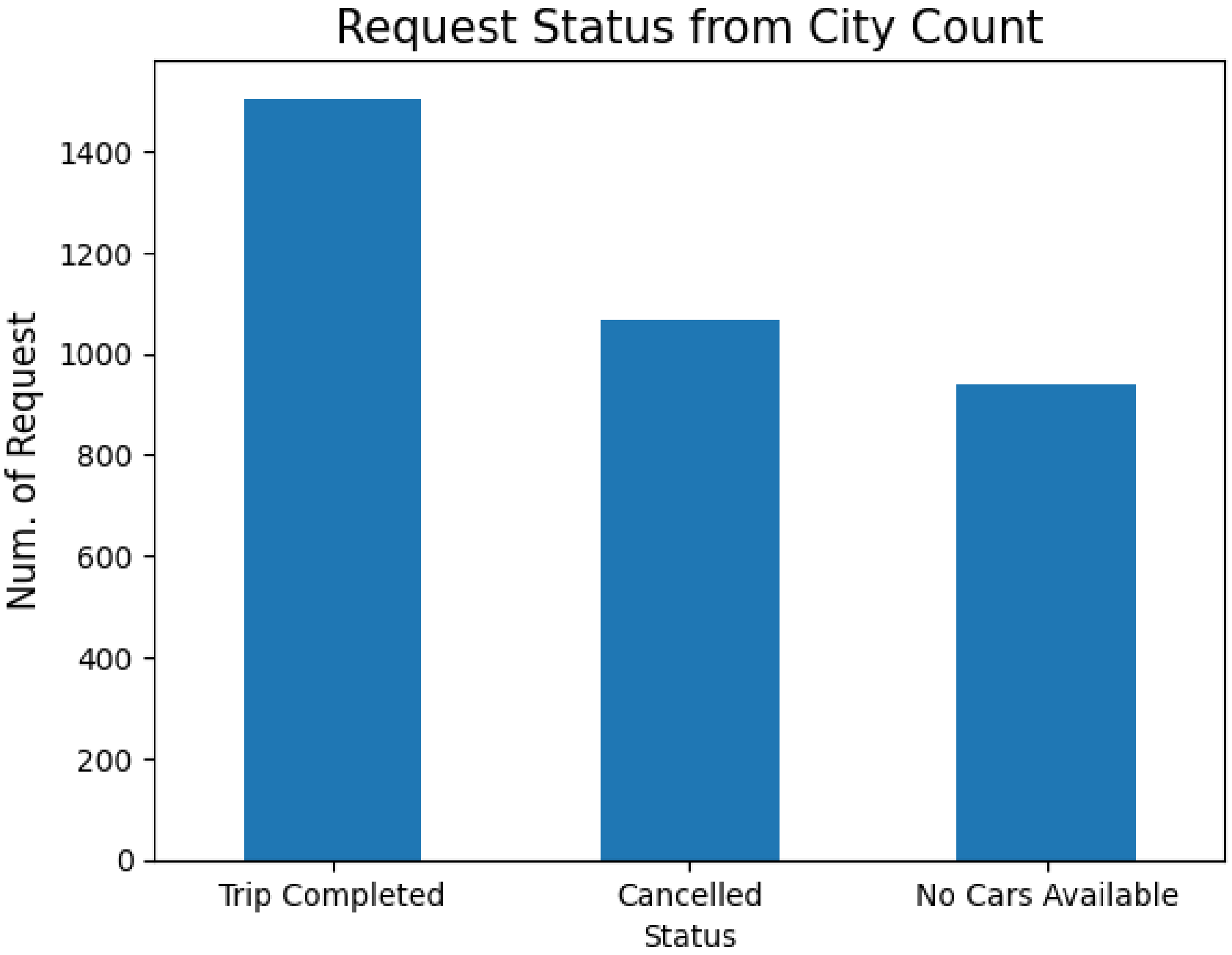
Total number of requests is **6745**

Trip completed (a.k.a. Demand) is roughly **2831**, roughly **42%**

We can assume that there appears a gap between Supply and Demand up to **58%** due to trip being Cancelled or No Cars Available



Pickup Point
CITY

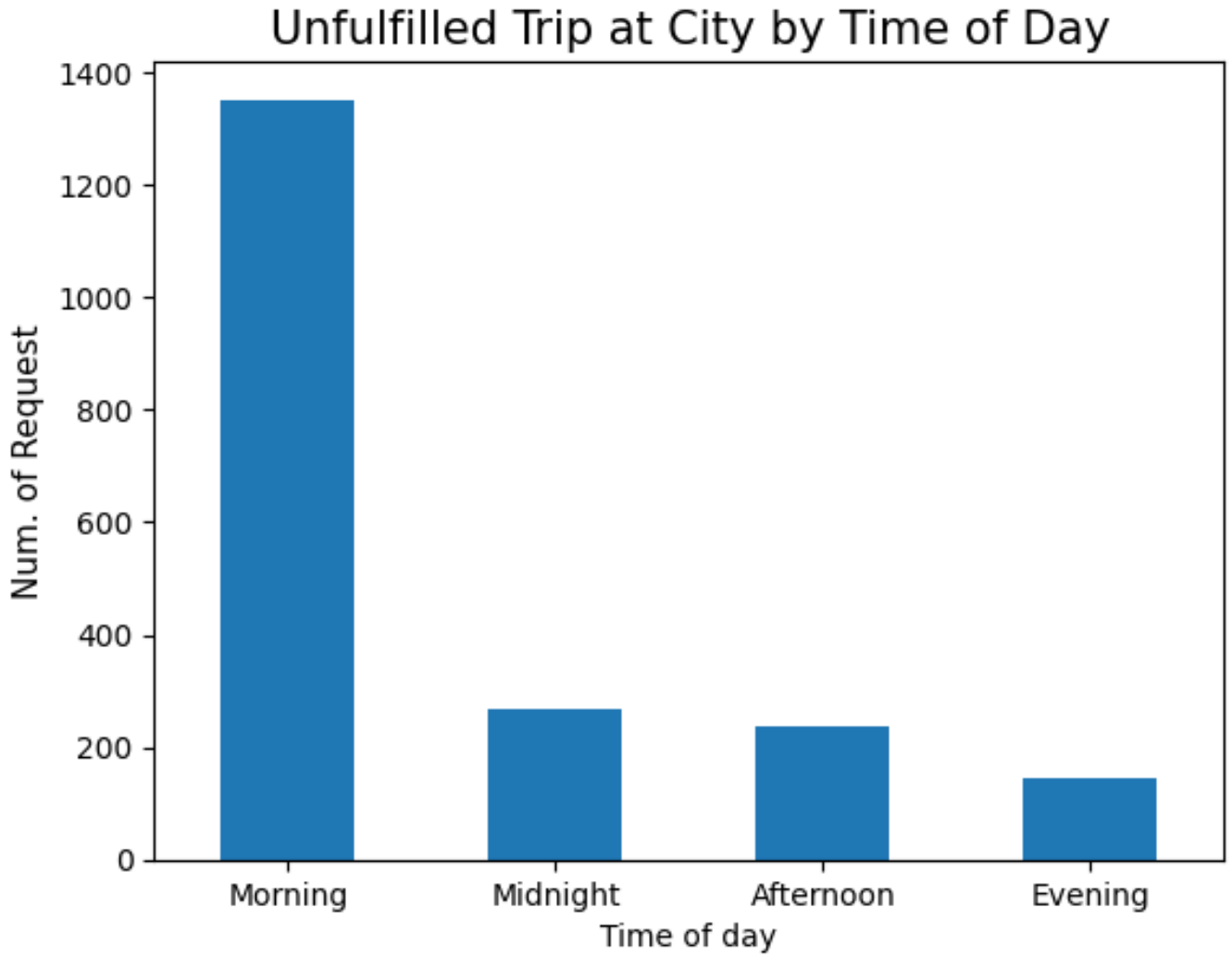


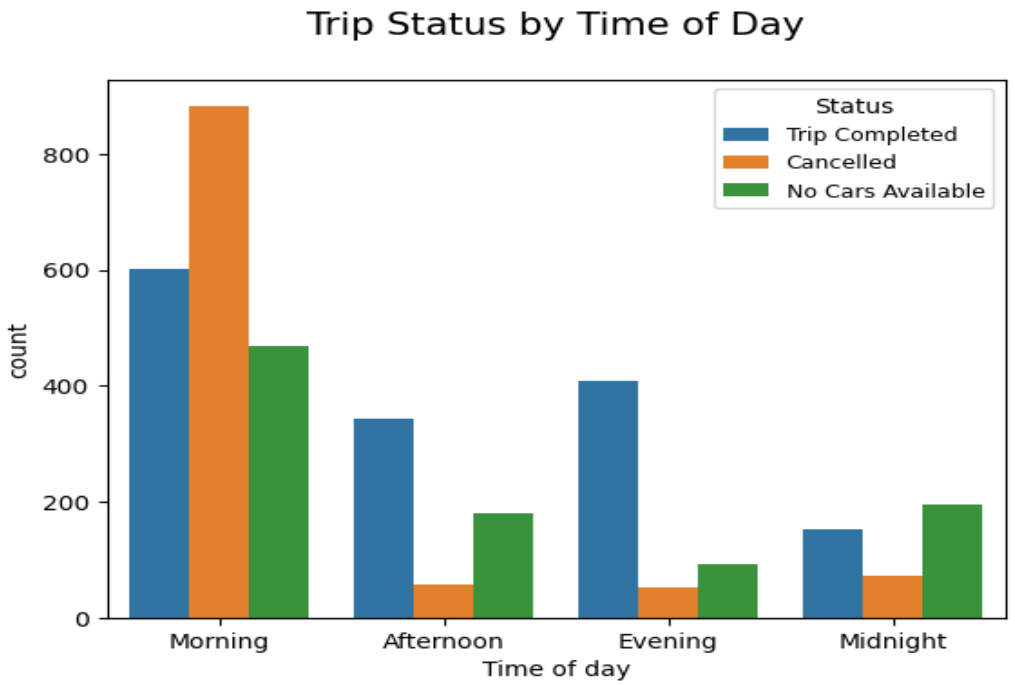
Overall, more requests are being marked as **“Completed”** when evaluated coming from City

Looking deeper in to trips that are being marked as “Cancelled” or “No Cars Available” by Time of the Day

- We can observe that most unfulfilled trips happen in the **morning** (5am-11am), which is **67% of the total unfulfilled requests**.

- Next is Midnight (11pm-4am), Afternoon (12pm-5pm), Evening (6pm-10pm) which are 13%, 12%, and 7% respectively.





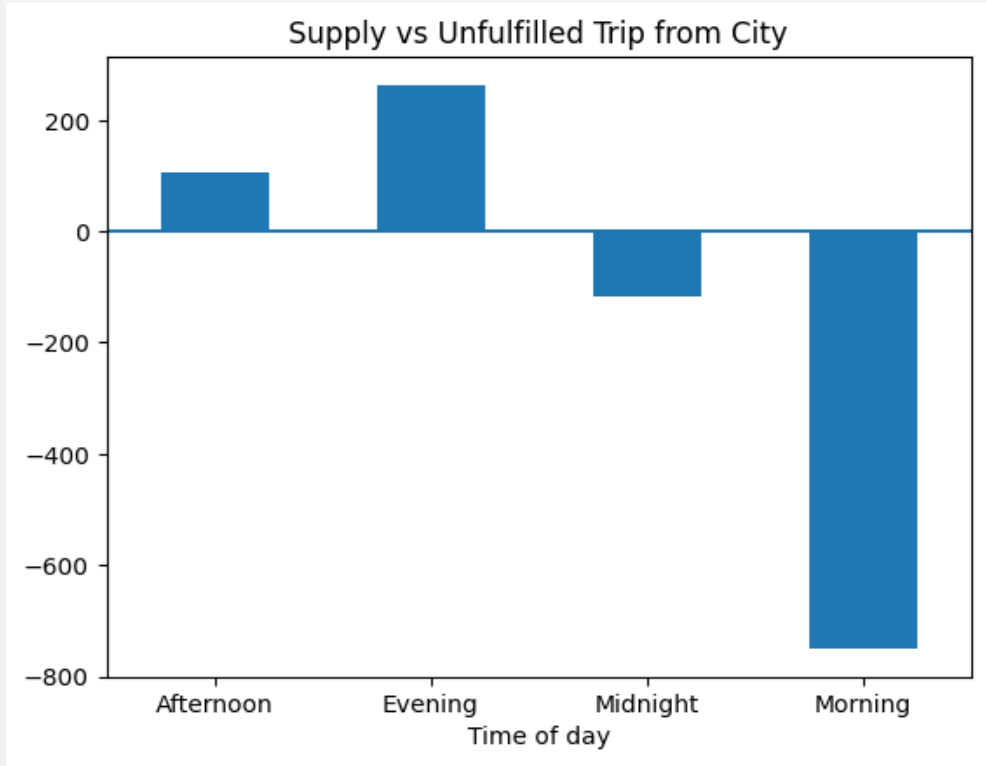
- Morning: 5am – 11am
- Afternoon: 12pm – 5pm
- Evening: 6pm – 10pm
- Midnight: 11pm – 4am

In the **Morning, most trips from City are reported as Cancelled**. Afternoon and Evening displays the closest gap between Demand and Supply as very low trip being cancelled or due to no car availability

Morning, we see the gap widen to **750 trips (Demand is too high)**

Evening performs the best with **263 trips (Supply matches Demand)**

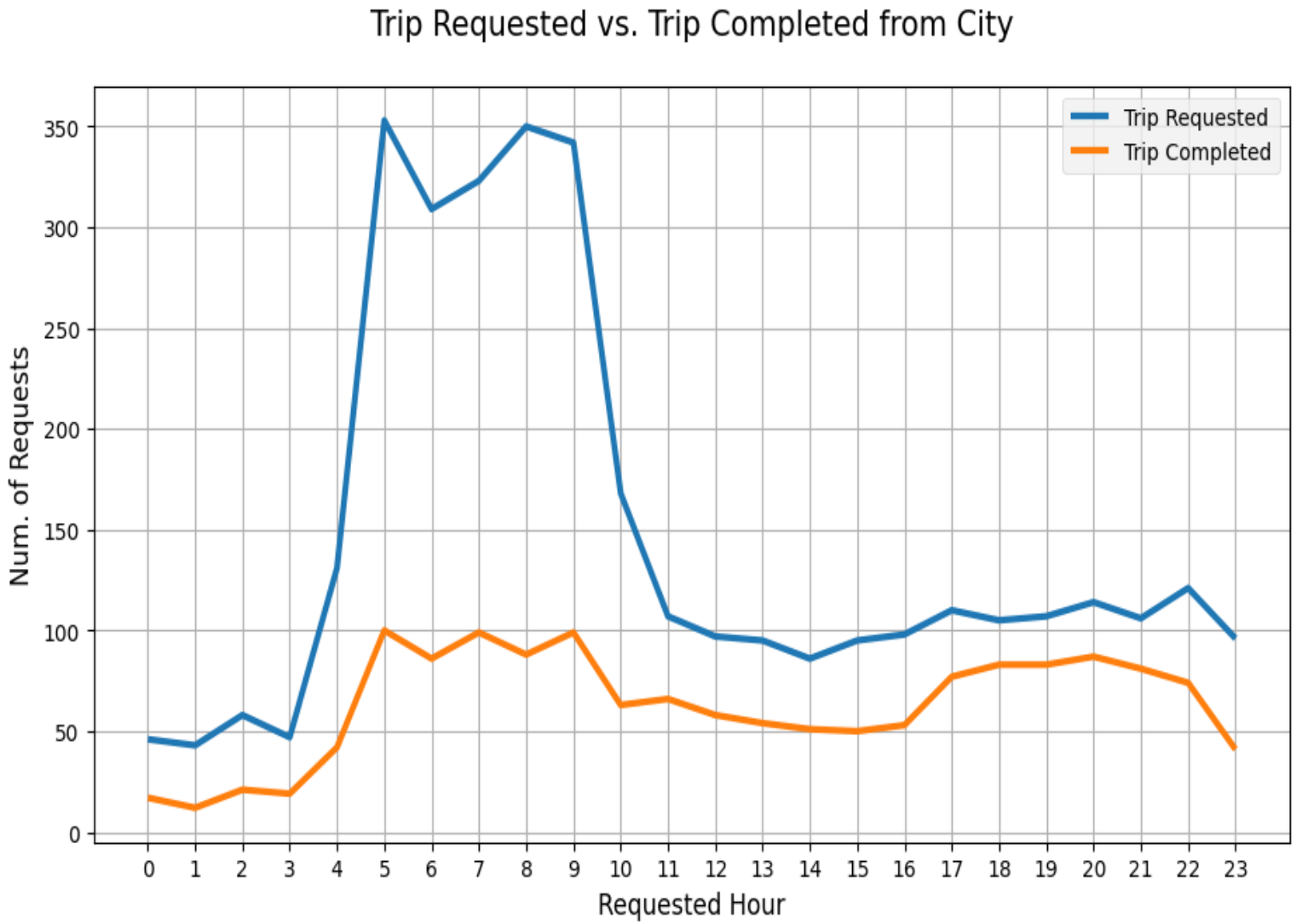
Status	Cancelled	No Cars Available	Trip Completed	Supply Demand Gap
Time of day				
Afternoon	57	181	343	105
Evening	53	92	408	263
Midnight	73	196	152	-117
Morning	883	468	601	-750



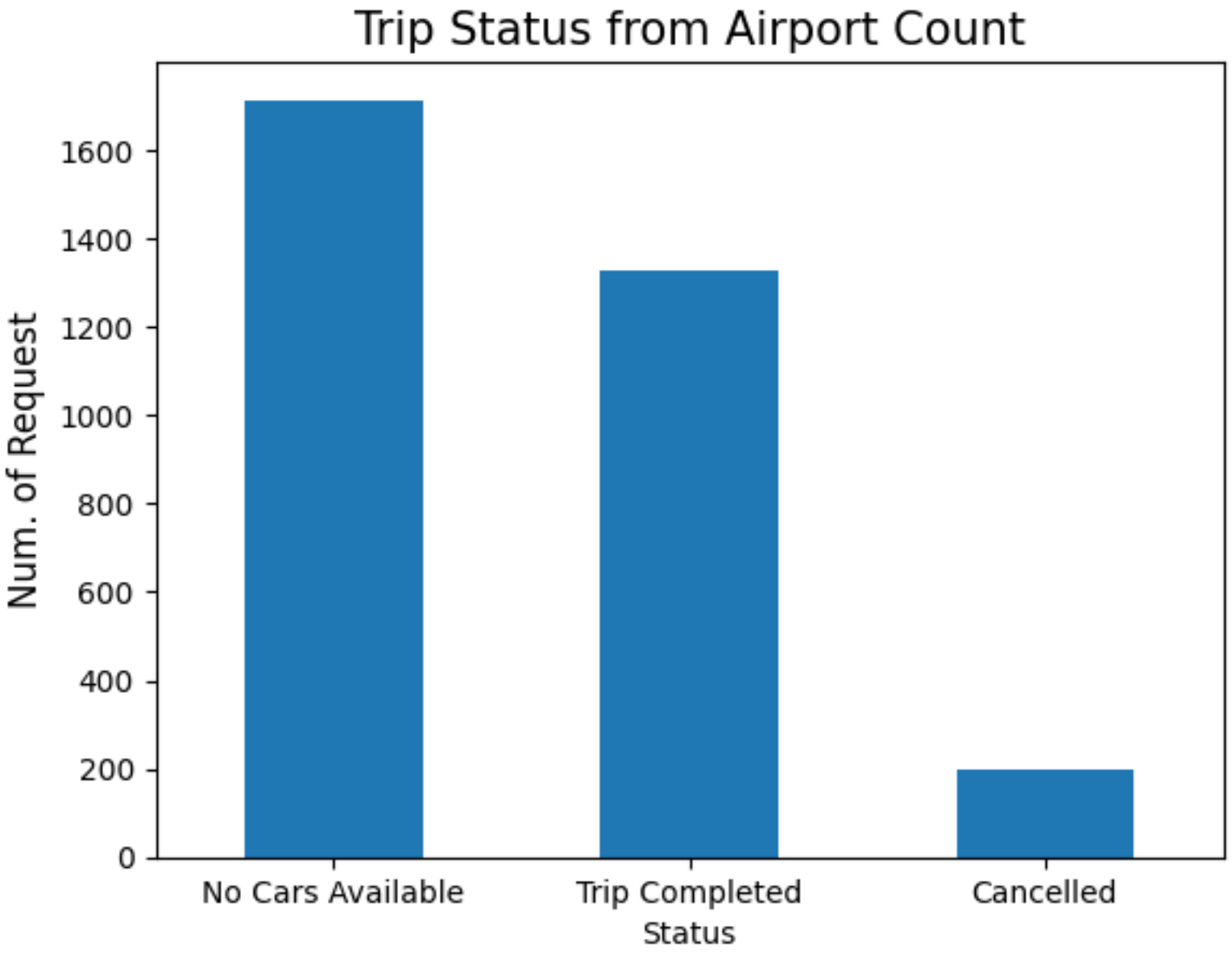
Here is a trend between number of requested trips and completed trips coming from **City**

- **From 5am to 9am, City** are experiencing extreme high Demand for Uber, although supply of drivers does increase, it is still inadequate compared to number of requests during that time

- This corresponds with the “Supply & Unfulfilled Trip from City” chart where the **Morning is the time with highest gap between Supply and number of requests happen**.

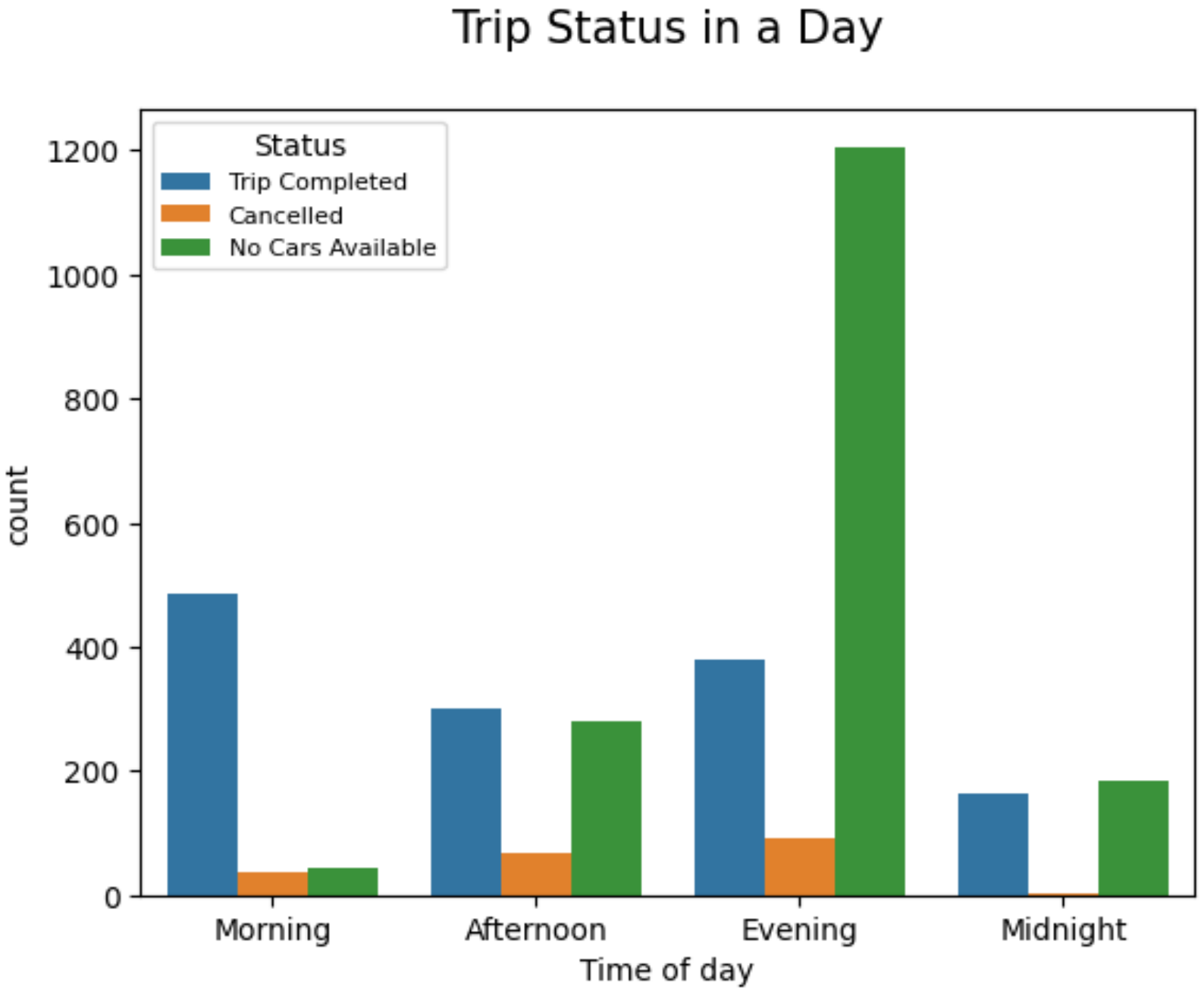


Pickup Point
AIRPORT



- From Airport, most requests are marked as No Cars Available with over 1600 requests.
- This means there is seriously high gap between Demand and Supply of Uber

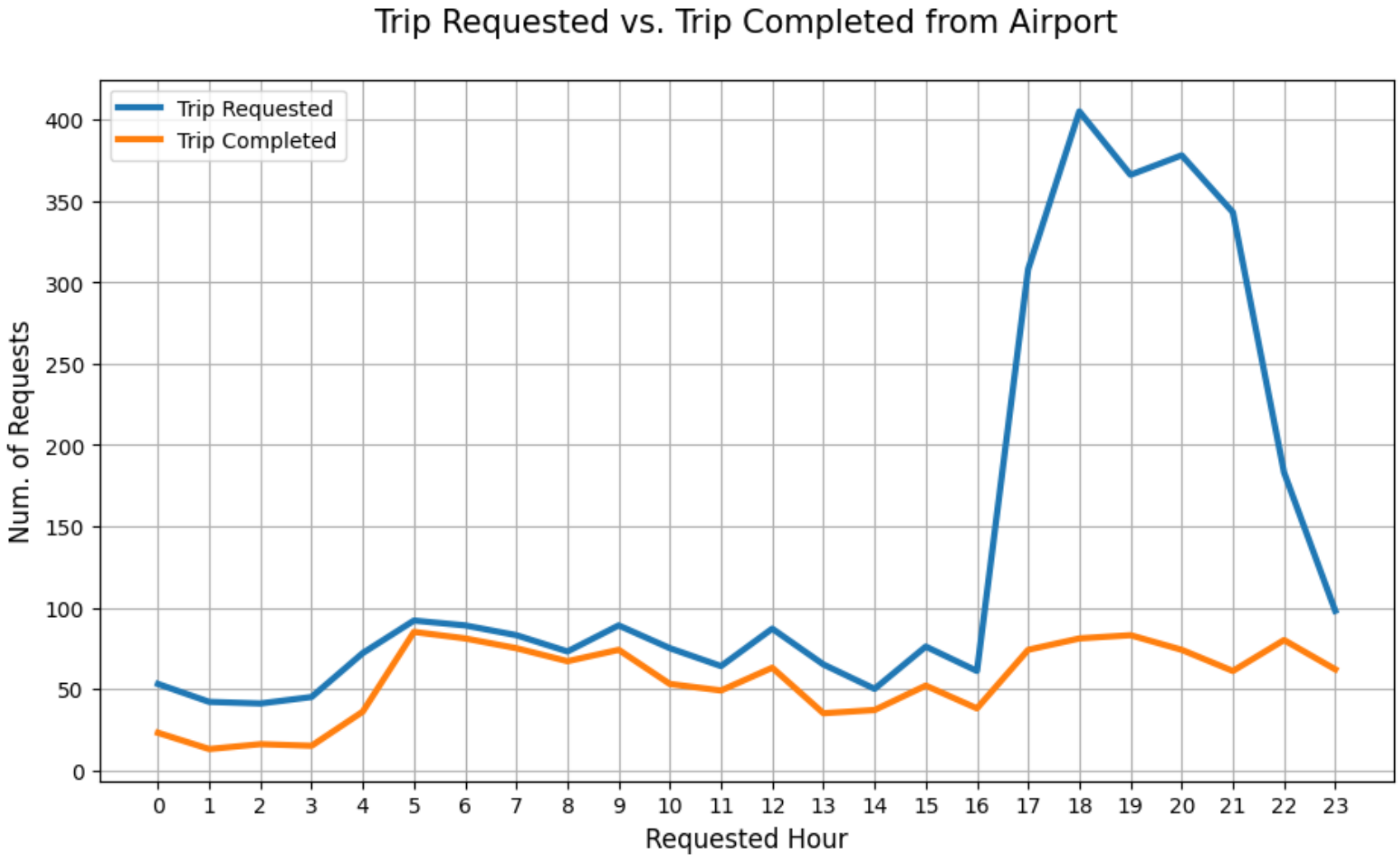
- Most unfulfilled requests happen during Evening (6pm – 10pm), where requests are eventually marked as “No Cars Available”.
- **Morning** experiences lowest volumes of those type of requests because it has highest number of trip being marked as Completed
- **The other time of the day** experiences low volume of requests, however, “No Cars Avai.” requests are just as high as “Trip Completed” requests. **This means there is also an issue with Demand far exceeds Supply**

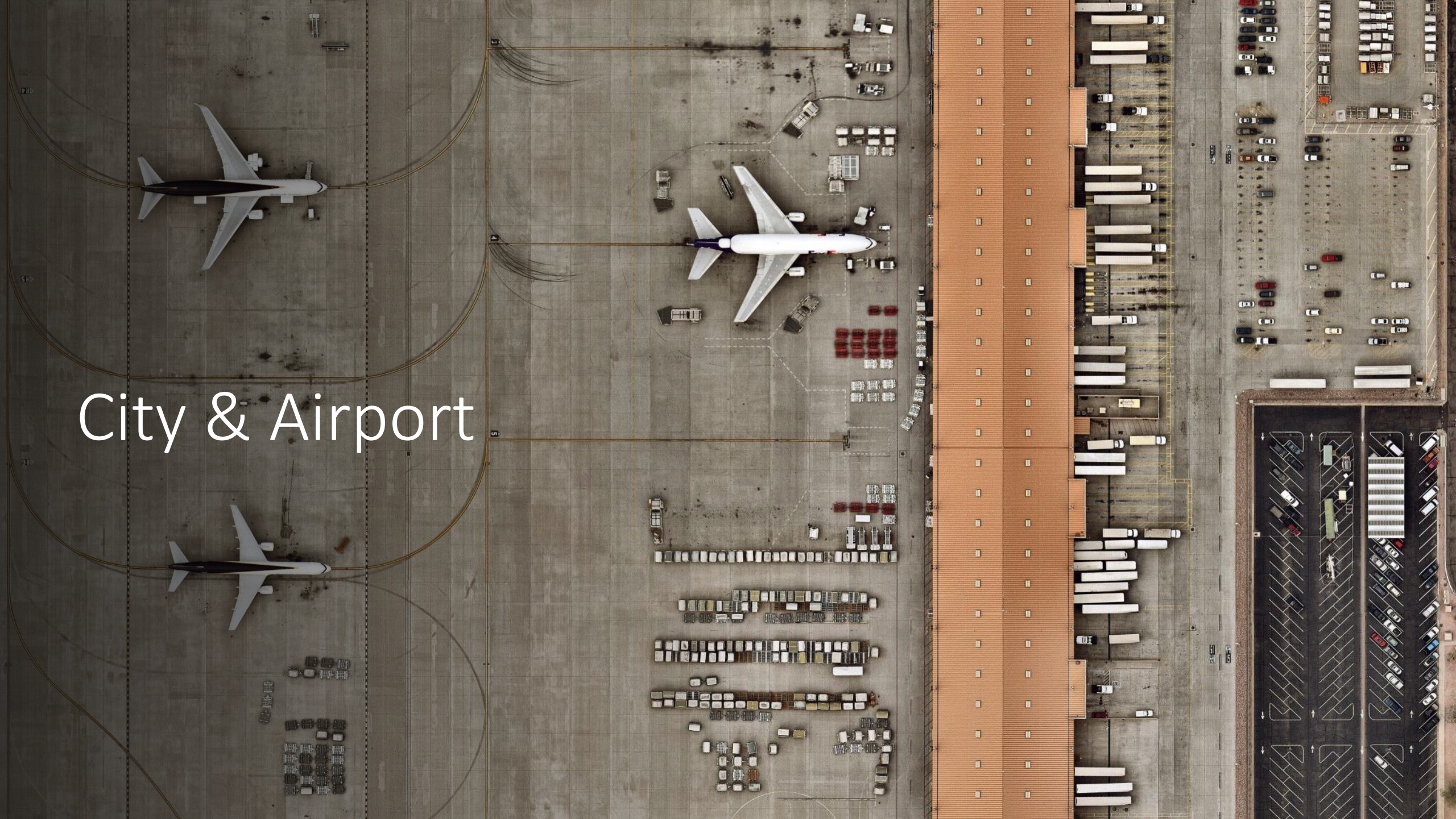


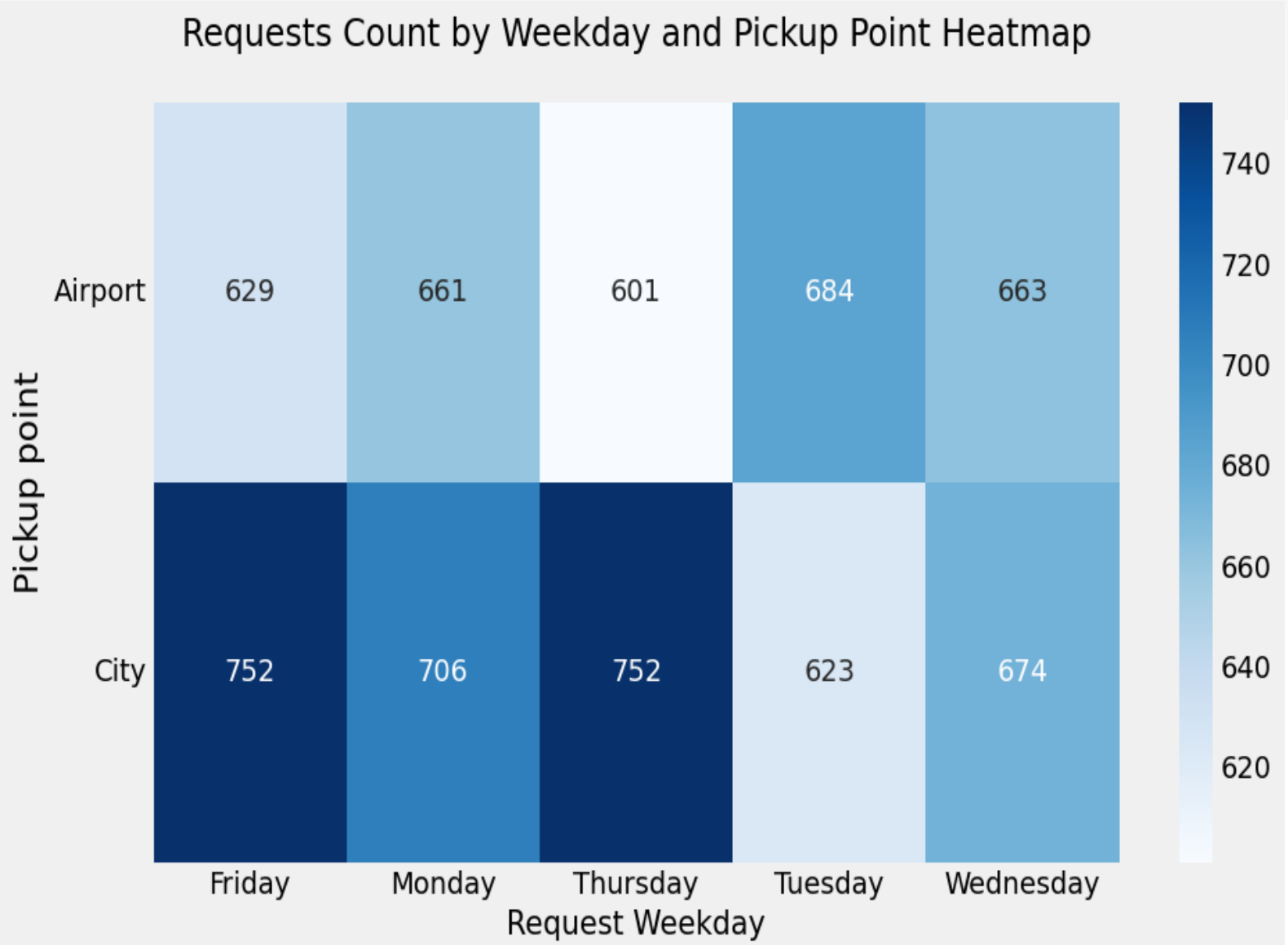
This graph shows, supposedly, the Supply and Demand for Uber coming from Airport by Hour

- We can see the **high uptick in Demand** for Uber starting from **4pm**, lasts until **9pm**

- This time range explains for extreme high number of “No Cars Available” requests in previous slide since **Demand exceeds Supply by 3 folds**



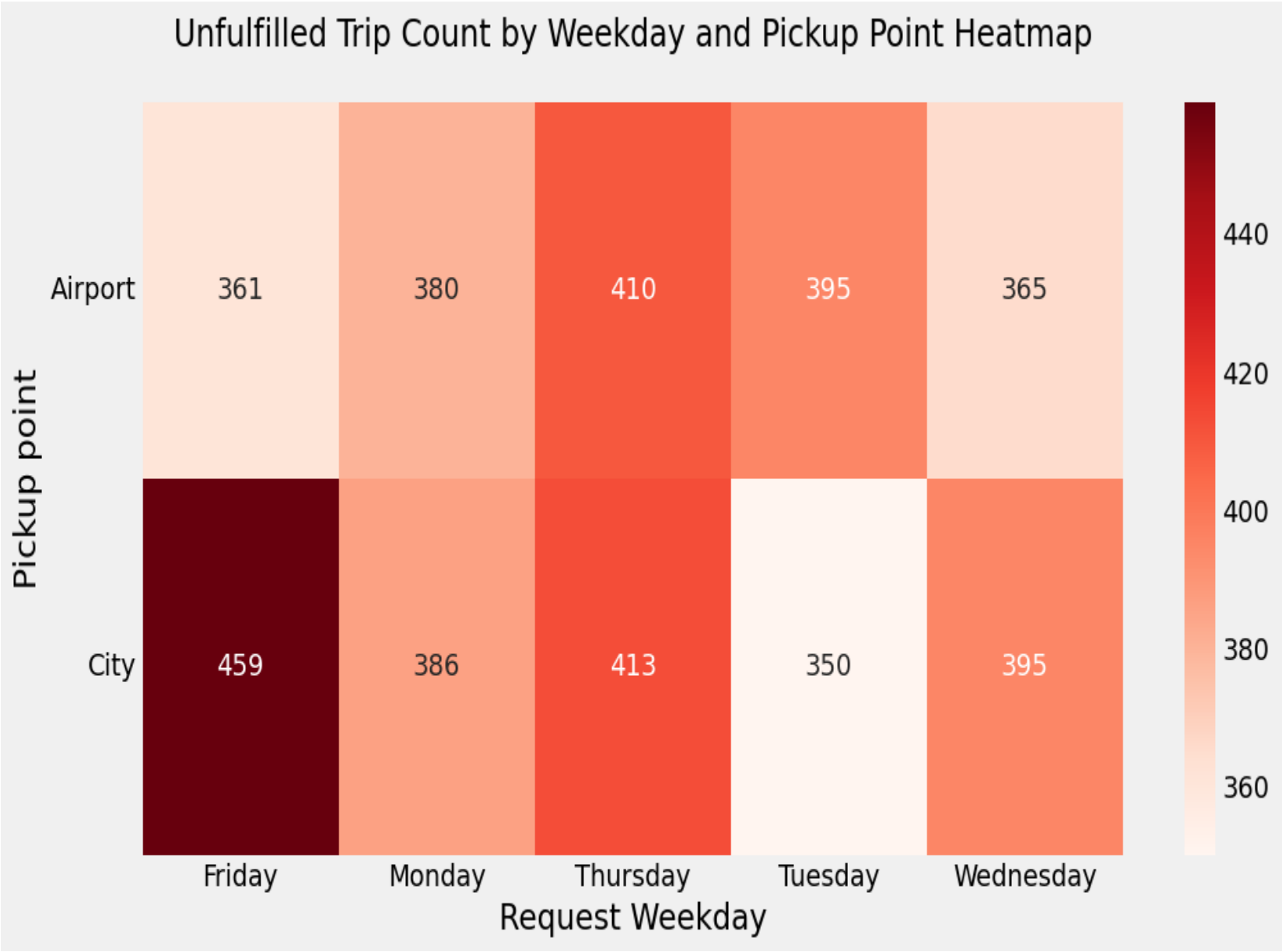

City & Airport



- **Friday, Monday and Thursday** appear to have packed high number of requests **(Demand)** for Uber rides from **City**.

- The highest requests is 752, lowest is 601, only 25% increase from the lowest.

****Important note** -- There are only 5 dates being recorded in the data set



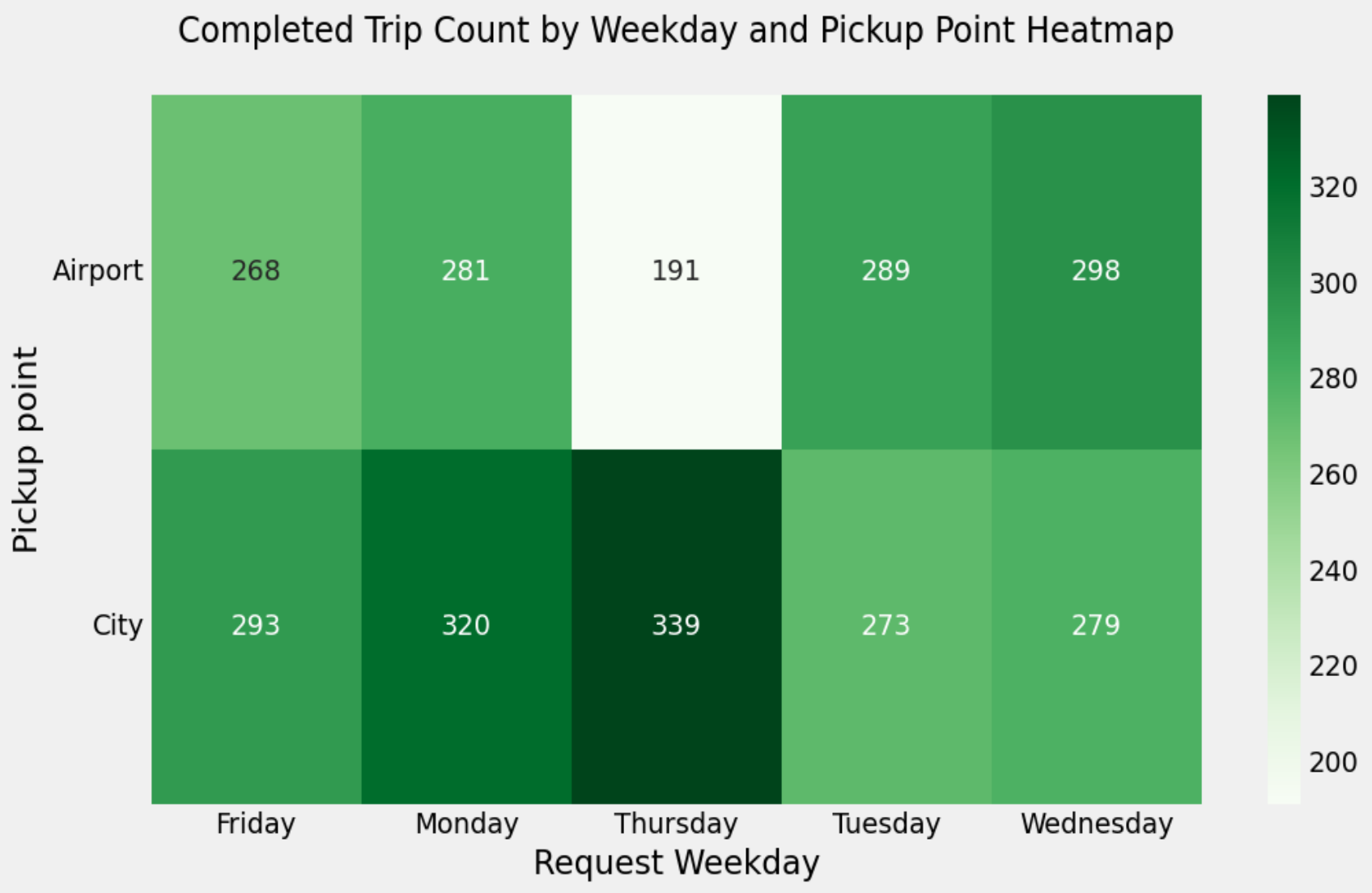
Unfulfilled trip is trip has status marked as either “Cancelled” or “No Cars Available”

The lowest unfulfilled trips is recorded on Tuesday with **350** trips

Highest unfulfilled trip count recorded on Friday going from City with **459** trips, which is 31% change from the lowest.

Based on given data, most Trip which got marked as either “Cancelled” or “No Cars Available” happen most on **Friday** coming from City, **Thursday** is the second highest coming from both City and Airport.

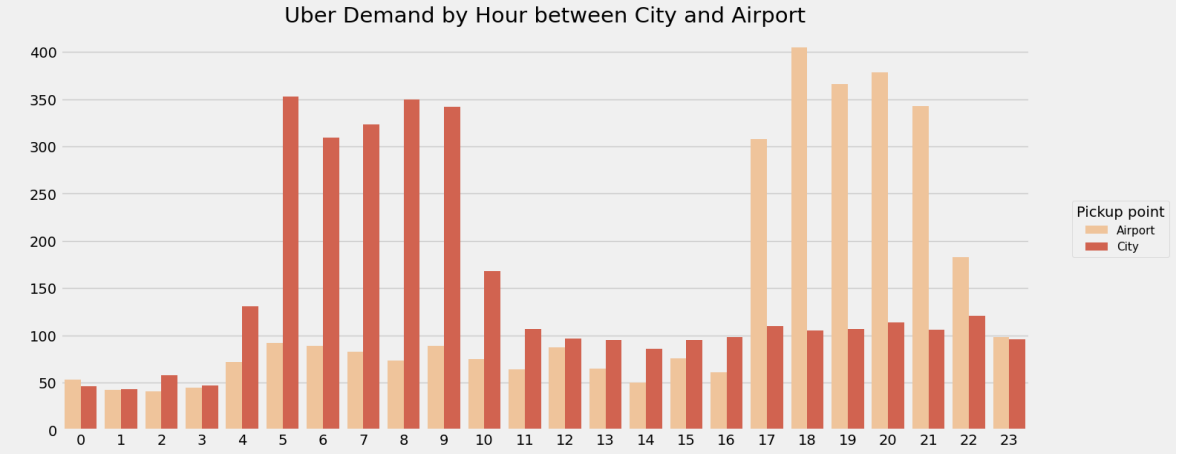
This means, on those mentioned weekdays, more than 50% of requests went unfulfilled.

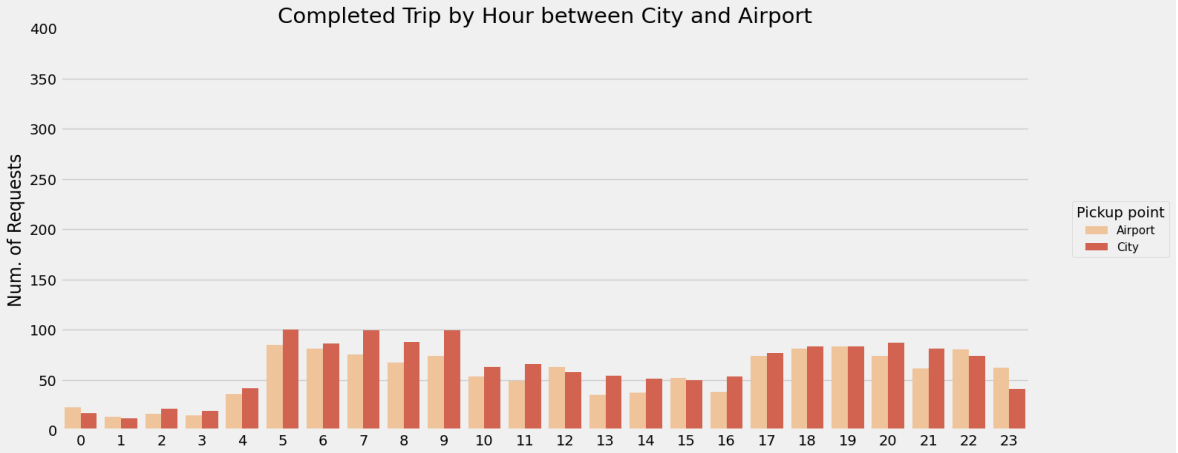


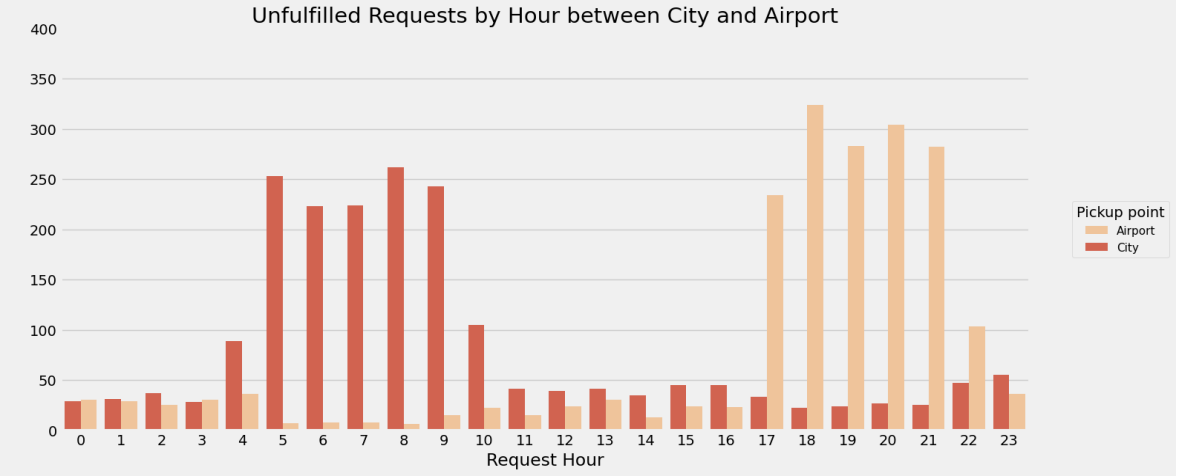
Lowest completed trip happens on Thursday but coming from Airport

Highest completed trip also happens on Thursday coming from City

**There are only 5 dates which are being reported in the given data set. Therefore, it would be inadequate to conclude anything given insufficient data on weekday







This is an overview of visualization between two pickup point (City and Airport) by each Hour

- **From 5 to 9 am** there is high demand for Uber from City going to Airport, the Supply (trip completed) is no where near meeting with the demand, cab drivers also take some advantage of dropping off customers at the airport and pick some up from airport going to the City

- That's why number of **Unfulfilled Requests** are also high from 5am-9am in the City, this reflects on *"Unfulfilled Requests by Hour"* (bottom chart). Whereas this type of requests is highest during 5pm to 9pm going from Airport to City

- **From 17 to 21,** there is high demand for Uber from Airport to City, the Supply is also no where meeting up with the Demand. There are still some trip coming back and forth between City and Airport.

- That's why during this hour, going from the City to Airport is much easier than going from Airport to City due to high demand and limited Supply (drivers)

The Why behind

The obvious issue for the gap between Demand for Uber and the existing Supply of Uber is **due to limited number of drivers**. We can see this clearly at the *Completed Trip by Hour between City and Airport* (prev. slide). Although the **Demand is high during identified peak hours, Supply (number of drivers) does not increase high enough to meet the demand. Plus,** it takes, on average, **52 minutes** to go from Airport to City, vice versa. Drivers are not being rewarded for travelling 52 minutes empty car just to pick up another request

Recommendation

Incentivize – flexible rates can be applied at different hour, higher during peak and more competitive during down time

Reward – drivers who take more requests during peak hour will be rewarded accordingly as they do not worry about driving back and forth with empty seats.

More drivers – hire part-time drivers, this can somewhat solve the problem as there is still rooms for more employees in order to meet the Demand, especially during peak hours