

Upgrad Assignment

By Sumit Pandey

AIM: The given dataset has information on trip status. The main task of this 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.

Data Cleaning

In Data Cleaning, Checked Duplicate rows if available, using info checked non-null values in each variable, Converted Column Request timestamp and Drop timestamp to Date Time and created derived variables. The column "Drop Timestep" store information about when the trip ended. Since this has nothing to do with the problem statement, we will drop this column.

Convert the date fields to date time and since the date in the source data has different formats.

Now, we will extract the hour form "Request time" this will act as the bins of our data analysis. We don't need extra information as to at which second the request was made.

This resulting data set will be in the form of:

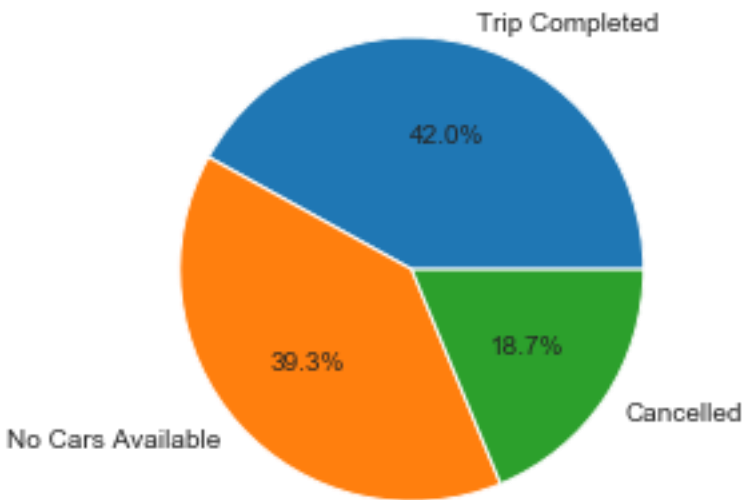
	Request id	Pickup point	Driver id	Status	Request timestamp	Request hour
0	619	Airport	1	Trip Completed	2016-07-11 11:51:00	11
1	867	Airport	1	Trip Completed	2016-07-11 17:57:00	17
2	1807	City	1	Trip Completed	2016-07-12 09:17:00	9
3	2532	Airport	1	Trip Completed	2016-07-12 21:08:00	21
4	3112	City	1	Trip Completed	2016-07-13 08:33:16	8

Creating a new column from the hours column by categorizing it into different categories and a Request day column by fetching the day of week from the request timestamp column

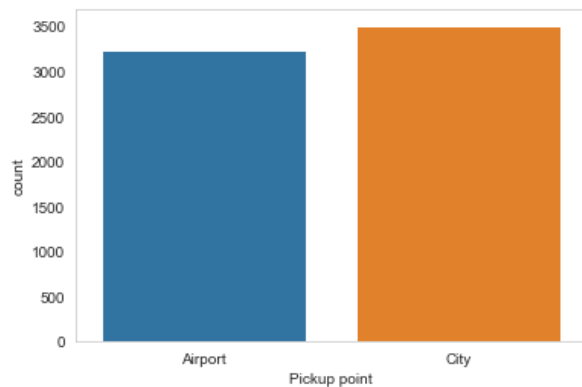
	Request id	Pickup point	Driver id	Status	Request timestamp	Request hour	Request time	Request day
0	619	Airport	1	Trip Completed	2016-07-11 11:51:00	11	Morning	Monday
1	867	Airport	1	Trip Completed	2016-07-11 17:57:00	17	Evening	Monday
2	1807	City	1	Trip Completed	2016-07-12 09:17:00	9	Morning	Tuesday
3	2532	Airport	1	Trip Completed	2016-07-12 21:08:00	21	Night	Tuesday
4	3112	City	1	Trip Completed	2016-07-13 08:33:16	8	Morning	Wednesday

DATA ANALYSIS:

At this point the data is ready for analysis with all the required columns in place.

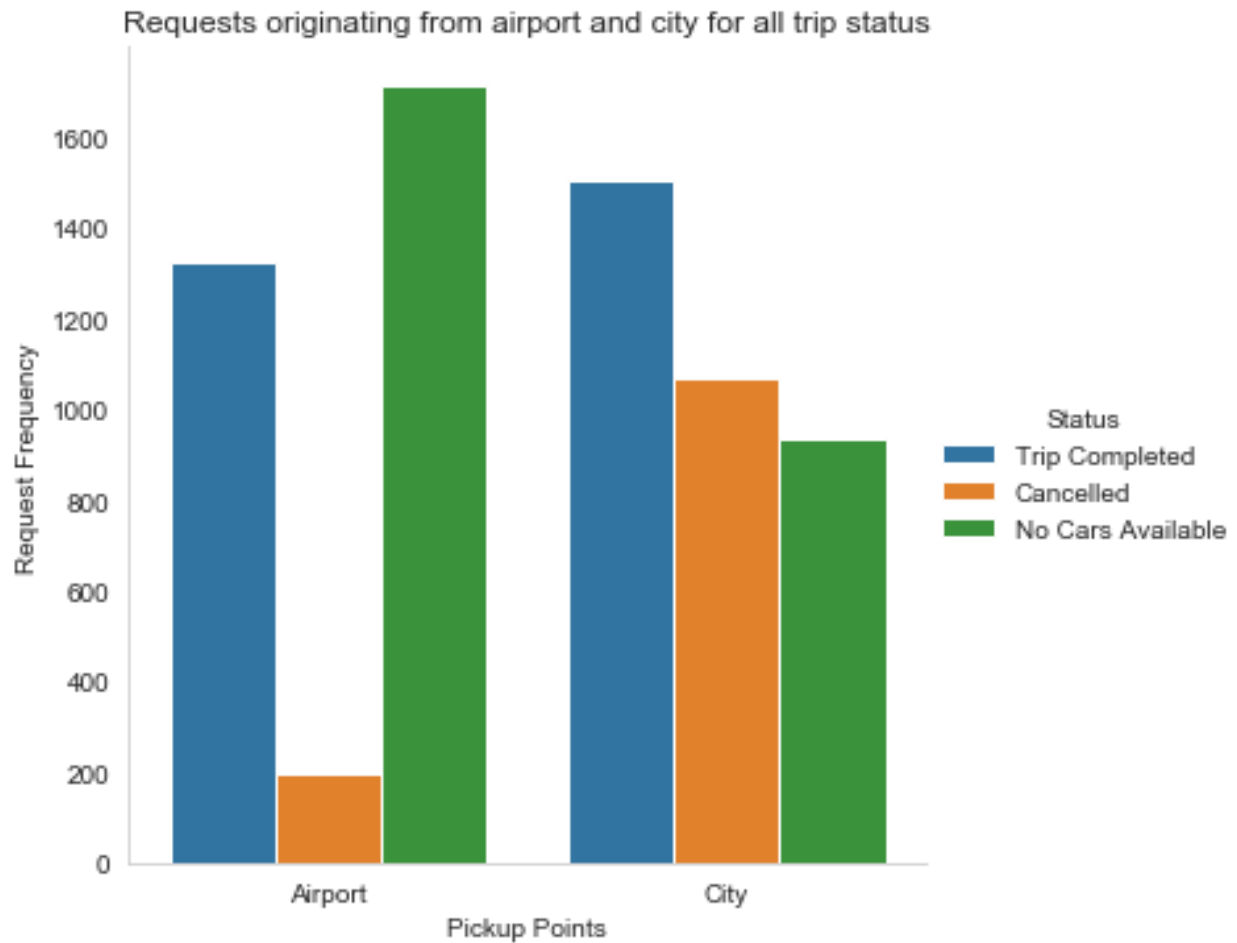


Viewing the frequency of request with their status shows us the above the figure. It is clear that most of the failed request are due to no cars being available at that moment. Still further analysis is need to give a concrete conclusion and the cause for this.

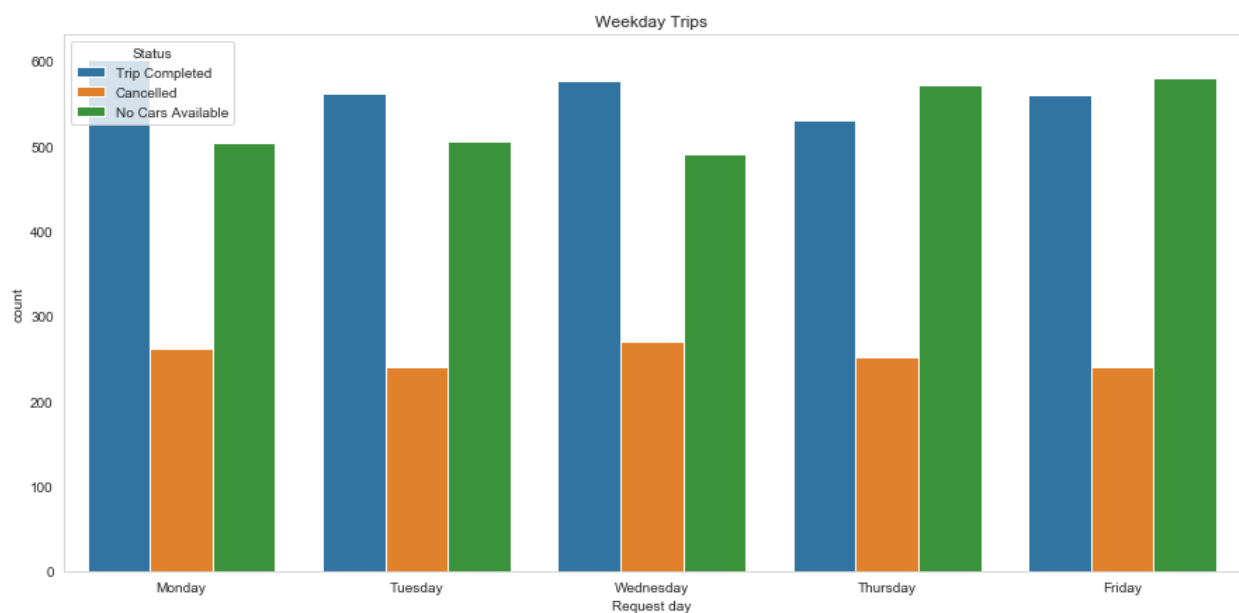


From above plot it can be seen that there are two pickup points **Airport** and **City**.

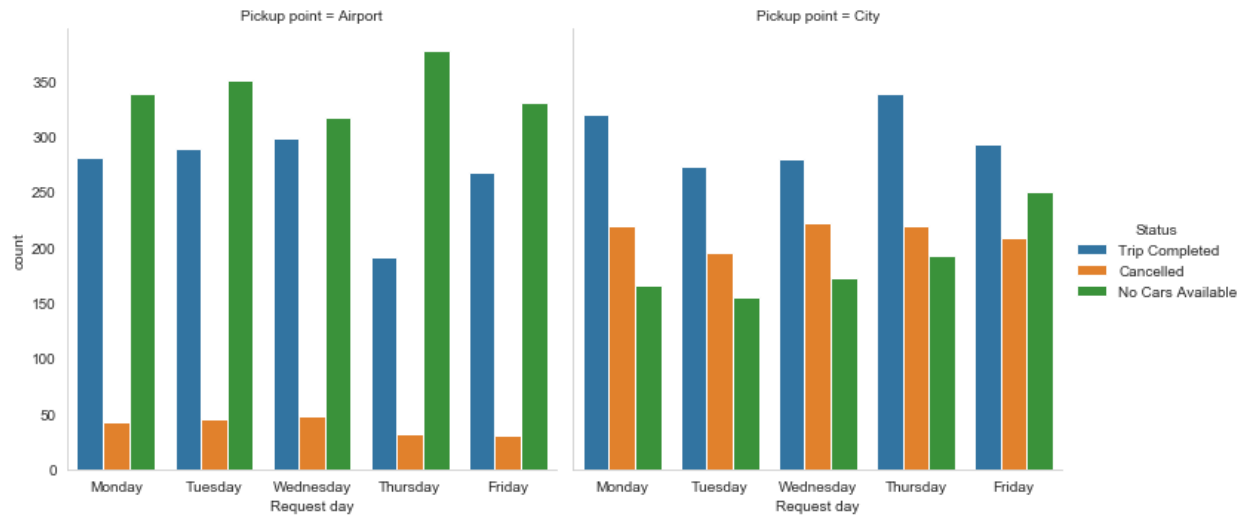
They have almost equal request however city have slightly more requests than airport.



This figure shows us that in cities the greatest number of failed requests are due to cancelled trips. This might be due to personal reason of the driver. The car unavailable will require it plotting against time of the day. We will do this later.

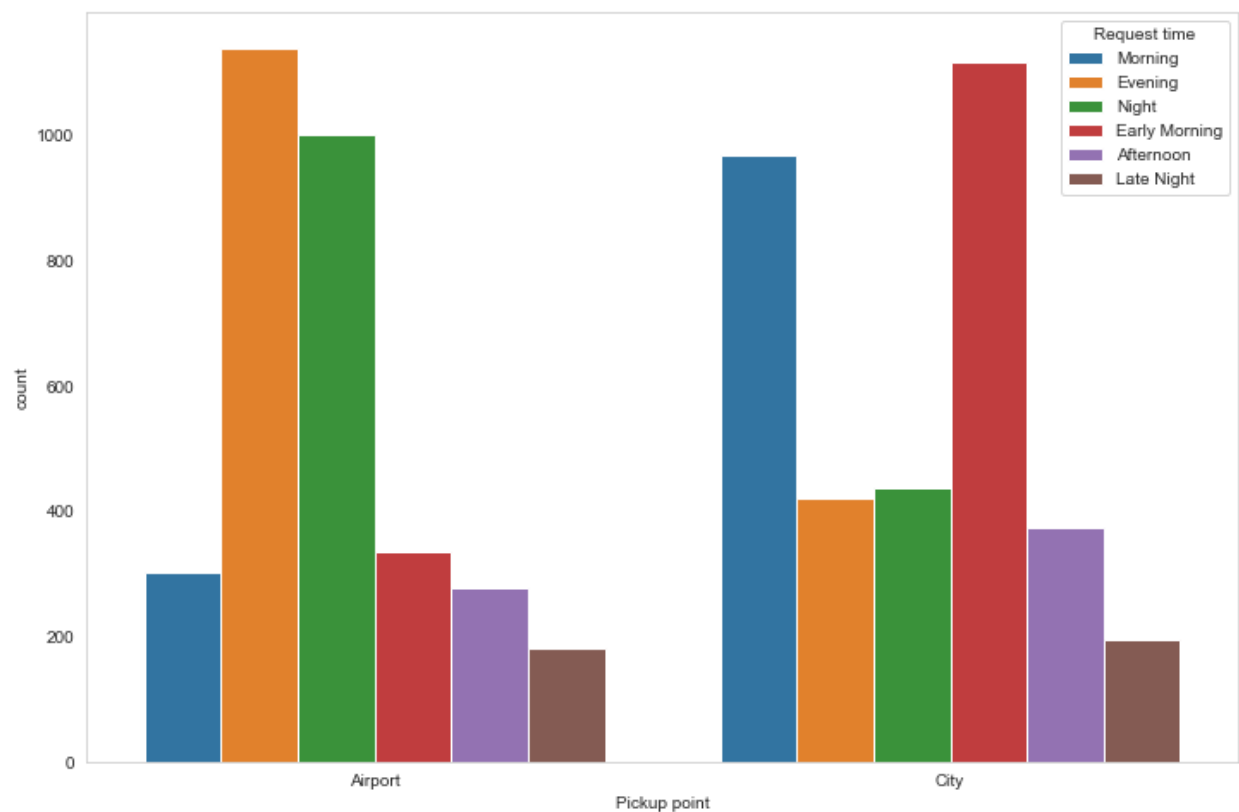


From above plot it can be see that the cancelled trip frequency is almost constant.

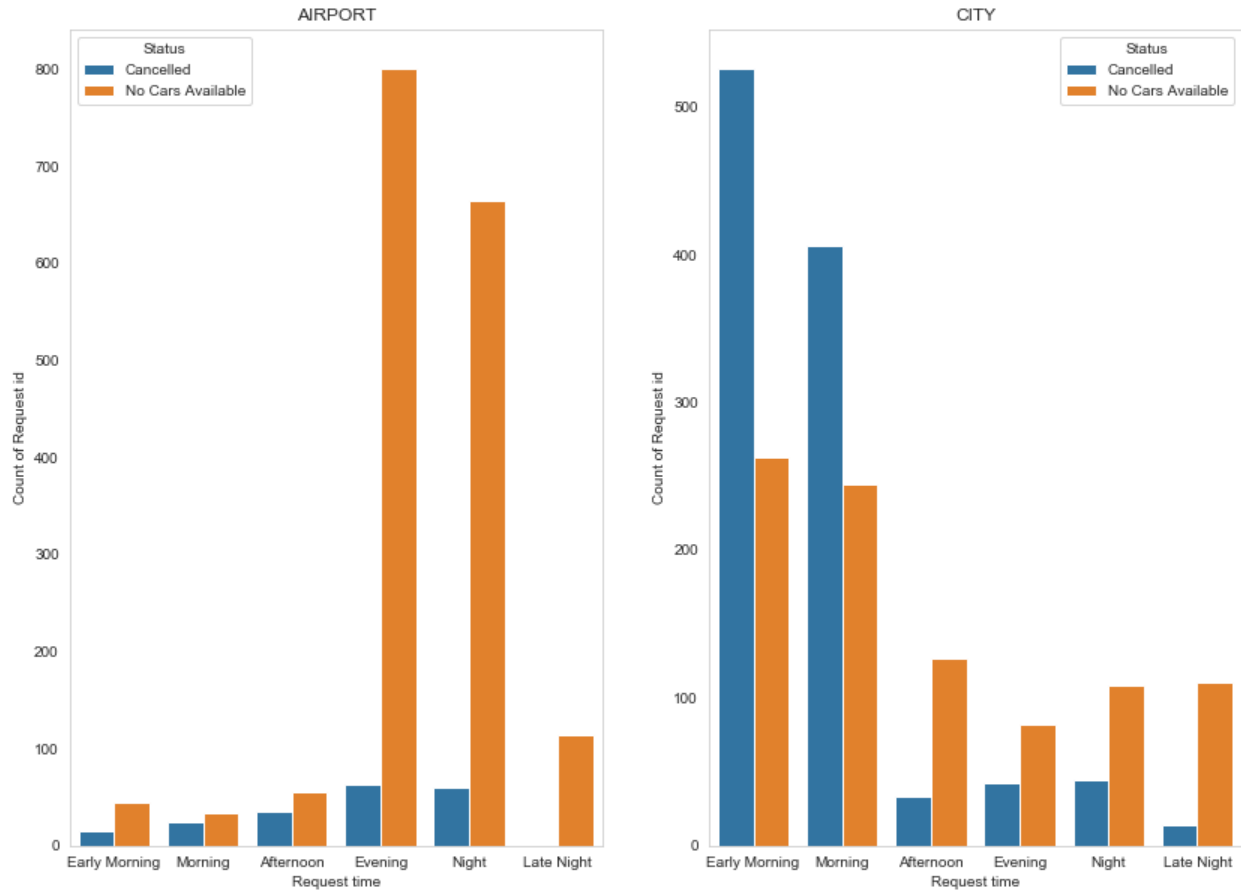


From above plot it can be seen that -

- For trips from city - airport, except for Friday, cancelled trips are more for all other days of week
- For trips from airport - city, for all days there are a greater number of No Cars Available

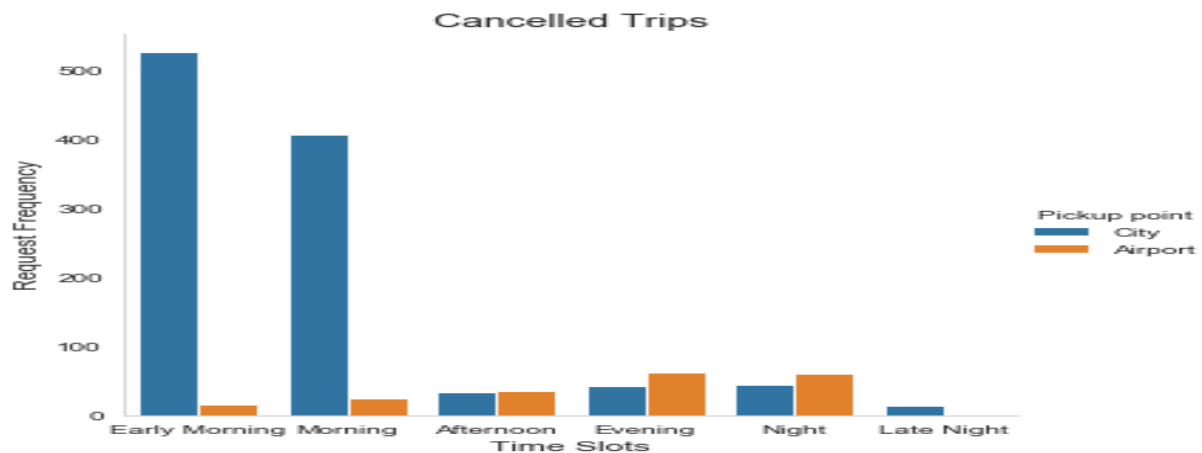


From above plot it can be see that **Airport** has higher request in **Evening** and **City** has higher request in the **Morning**.



The airport data is highly skewed in favor of no car being available this might be due to the fact that the airports are generally far away from the city and drivers generally do a one-way trip to airport.

This figure plots the status (cancellation) against time. We can see that most of the trips in the city are cancelled in the early hours. At airport the same is for in the late hours this may be due to the fact the drivers aren't able to get a return booking.



This graph clearly confirms our hunch that the drivers aren't available near airport during late night. This results in failed trips.

Conclusion

1. Provide incentives for the airport trips during peak time.
2. Assigning few extra cabs specially for the airport trips.
3. Fixing a base price for driver's idle time in the airport or to come back to the city without any passenger.
4. Impose penalty for cancellation of requests by the drivers.
5. Set a threshold for the maximum cancellation per day.
6. Promote continuous trip to airport with incentives.
7. Promote advance booking to airports and at the same time keeping drivers updated with the flight schedule will help them plan their work and they can accept the request as per their work plan.