Analysis of Oolala Cabs Dataset

FINAL REPORT

Submitted By:

Archit Garg

[architgarg96@gmail.com](mailto:architgarg96@gmail.com)

(+91) 8979778552

**Problem Statement:**

Analyzing the dataset to find whether customers are really facing problem in finding cabs to airport.

**Hypothesis:**

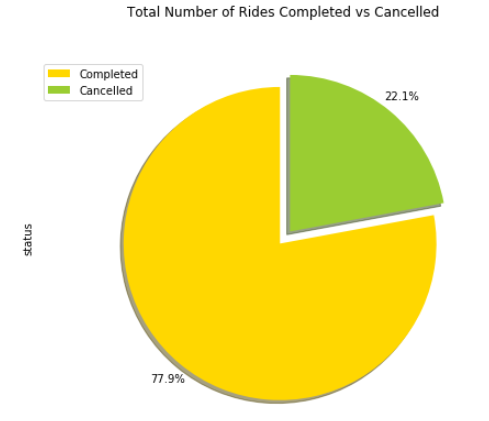
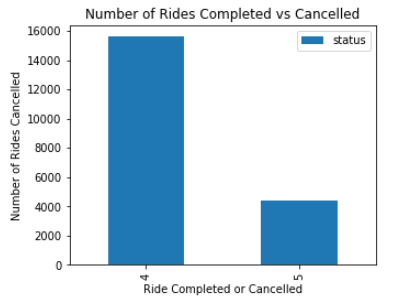
Null Hypothesis- Drivers are not cancelling rides to airport

Alternate Hypothesis – Drivers are cancelling the rides to airport

**SECTION- 1**

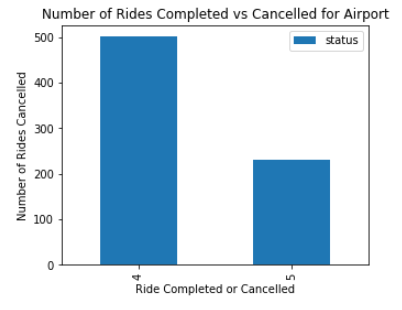
**Data Analysis**

**1.** To compare total no. of rides that are completed vs cancelled



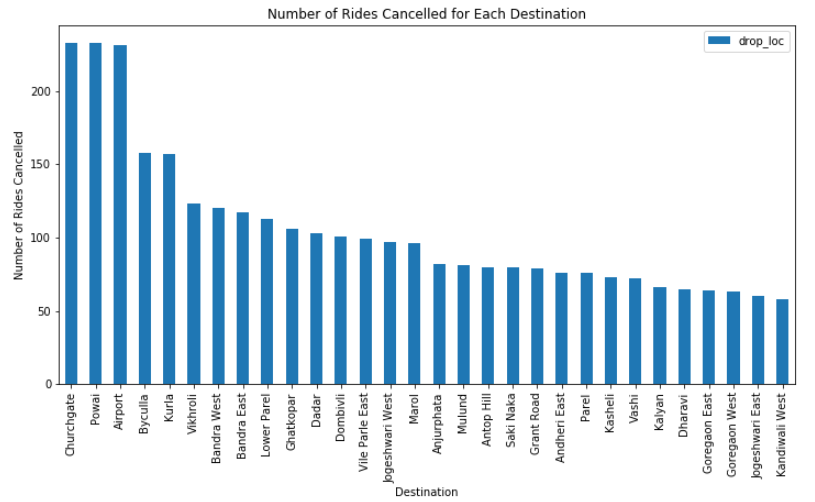
This shows that 22.1% of rides in whole dataset are cancelled out of any reason.

2. To compare no. of rides to airport that are completed vs cancelled

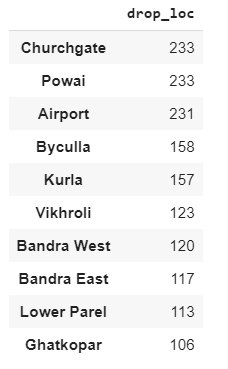


This shows that 31.6 % of airport rides are cancelled which is significantly more than average number of rides being cancelled.

3. To find No. of rides cancelled for each drop location

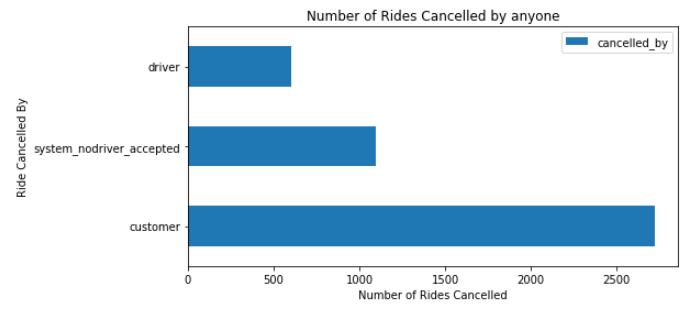


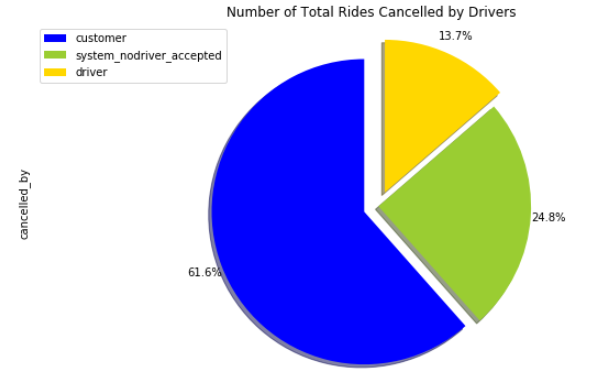
Amongst all the rides that are cancelled for any drop location airport has 3rd highest number of rides being cancelled.



In this data set we can see that no of rides to airport that were cancelled are 231.

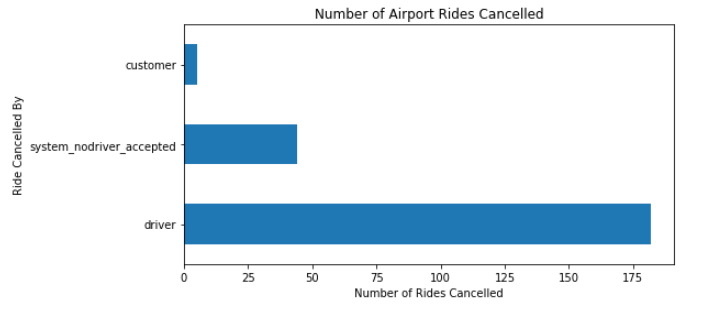
4. Let us compare that no of rides that are cancelled by driver, customer or due to system for all drop locations

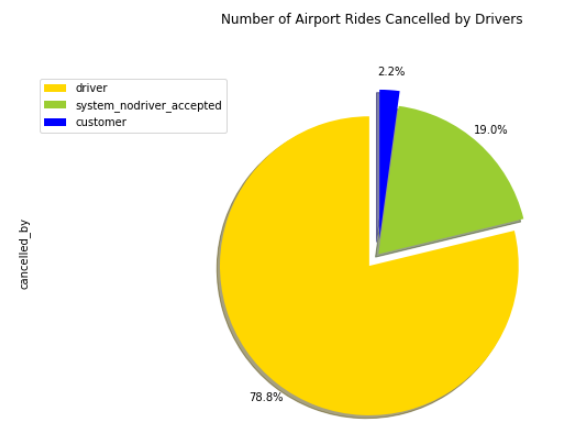




This clearly shows that for any general drop location customers cancel 61.6% of rides while drivers cancel only 13.7% of rides.

5. Let us compare that no of rides to airport that are cancelled by driver, customer or due to system for all drop locations

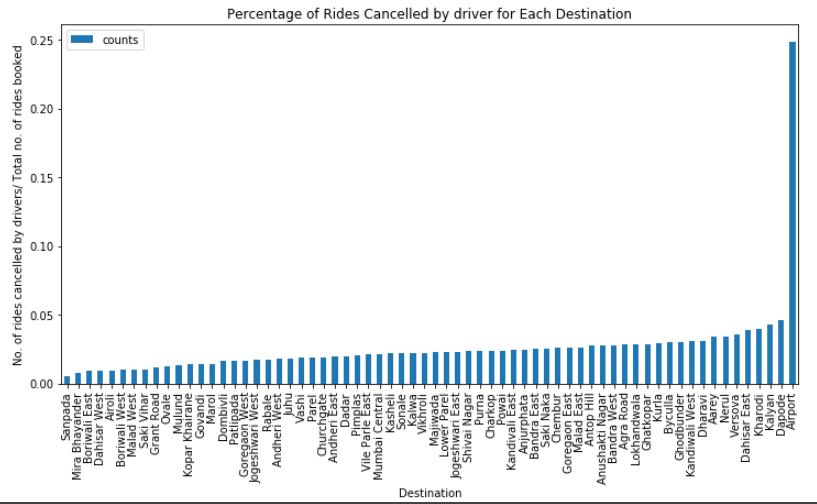




Through these fig. it is clearly visible that out of all cancelled of airport that were cancelled 78.8% rides are cancelled by drivers whereas above we have seen that for any general drop location drivers cancelled only 13.7% rides. Thus, we can see that there is something unusual about rides to airport.

Let us, analyze further. There might be a case that number of rides booked for airport are significantly less.

6. So, let us compare:



As, we can see for most of the destination/ drop locations the ratio of rides cancelled by drivers to total no. of rides booked for that destination is less than 5 % while in the case of rides booked for airport the percentage of rides cancelled by driver is almost tending to 25% which is significant no. of rides being cancelled by drivers for the airport.

Thus, I finally conclude that customers are **facing problem in finding a ride to the airport and drivers are cancelling the rides to airport more as compared to those for other drop locations.**

**Null Hypothesis has been rejected.**

**SECTION-2**

Coming onto the cause of this problem:

Let us evaluate what all can be the causes:

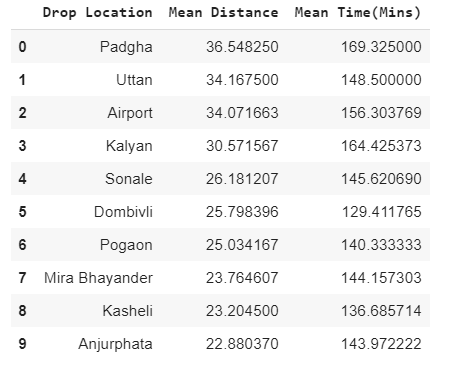
Following can be causes of problem according to me.

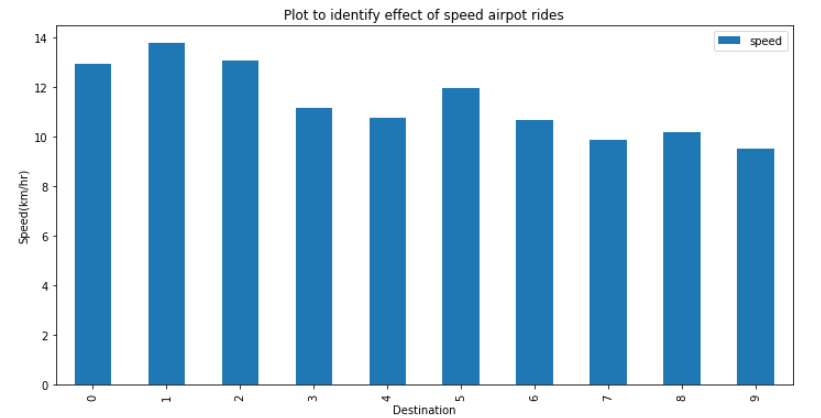
1. Time taken for a ride to airport might be more than time taken for other drop locations.



Here, we can see that there are some locations for which time taken in ride is even more than what was taken by a ride to airport. But we haven’t faced such problem in any other drop location. Thus, this shall not be the reason for drivers cancelling ride to airport.

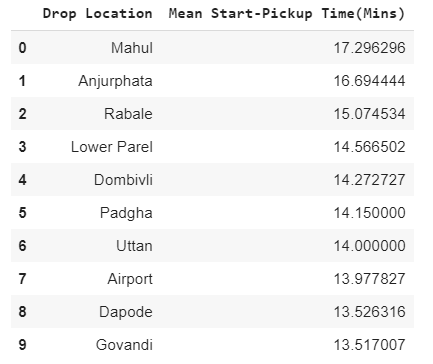
2. There can be case, when for approximately same distance as of a ride to airport- the time taken is more in airport ride. Let, us examine the below dataset:





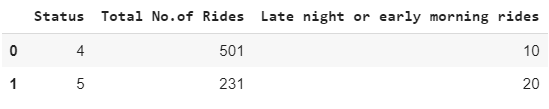
As, we can clearly see that irrespective that there that time taken for covering approximately same distance is almost same. Thus, there is no significant relation amongst them.

3. Comparing the case that there is more time taken in start-pickup for the rides to airport



Time taken in start-pickup for airport rides is also approximately same as for other locations. Thus, there is no relation of this with airport rides being cancelled by drivers.

4. Taking another case in consideration that there might be more no. of late night or early morning rides to airport as compared to day rides to airport



Through this data it is clearly visible that late night or early morning rides have marginal effect because they are very few in number.

5. Coming to that there might be the case that drivers who are going to the airport are not getting return rides, so they might have to return back without customer from airport so they might have been facing loss due to it.

Through dataset we obtained,

No. of drivers who had gone to drop a customer to airport and got customer for return journey from airport= 21

No. of drivers who had gone to drop a customer to airport but we don’t have their any data of return from airport i.e. they might have come back without a customer= 480

Thus, this can be major cause for drivers cancelling the ride to airport.

Through this data available if we calculate simple probability that drivers come without passenger = 21/501 = 0.0419161

Which is not even 5%, therefore 95% of drivers (from dataset available to us) either return without passenger or we are unaware of their further trips.

**SECTION-3**

1. This is Final Report

2. Python script notebook is attached in the zip file

Colaboratory link for the Python Notebook is:

<https://colab.research.google.com/drive/1xpU5g3KZMADAbwcdgZmgYrb1fT0NhiM2>