

## Data Visualization Project

### NYC TAXI DATA

Data visualization has been one of the most significant issues of this century. With the help of the improved computer power and processor strength we improved the abilities of data visualization. In this project, I tried to create a Dash App that shows interactive data, with a map that changes accordingly with the given data and changing variables. Further in this report, I will explain how I did data gathering and implementing the app.

Firstly, all data that is needed was given with the assignment and I downloaded all necessary data. The data I downloaded was for January of 2017. After downloading the data, I immediately created an SQLite Database. However, given data was still huge. Therefore, from the given database table, I got 7500 for each day, second, third, fourth and fifth of January. This way I wanted to decrease the wait time and increase process time (code for full data is provided as well within the file). When I had smaller datasets everything worked faster, which made the testing time shorter. After getting everything to database, rest of the work was done with the database and sql instructions. For example; for days slider that shows 4 days from January, whenever slider changes, I send the current slider day information to a function that returns specific data of that day. Upper-left graph on the snapshot shows how fare-amount changes(y) with the distance-travelled(x). On the map you can see the information that has been picked from dropdown menu on the top. When dropdown variable is 'Pick Up', you see regions of most pick-ups; greater the size of circle greater the number of pickups. If you change dropdown to Drop Off, you'll see drop off count for each region. Finally, at the very bottom you see a uniformly distributed graph which shows lastly hovered vendor id's total earned amount(y) over given 4 days(x). When you hover your cursor over 'Fare Amount v Distance' data, when vendor id changes bottom graph changes accordingly.

We can easily observe that in this huge amount of data there are some incorrectly put data. Instead of cleaning outliers and intrinsic errors, I left them in the set so that I can show that the data used can be erroneous. In conclusion; these graphs show that there is a high correlation between trip distance and fare amount. However, there are some edge cases that needs to be explained. Also, even though each vendor has almost equal total earning amount, vendor 2 varies a lot within itself. From the map we can observe that most of the people in northern New York uses taxi. And, we see that airport has good number of taxi pick up and drop off amount.

You can reach to the video of the app from here:

<https://drive.google.com/file/d/1di3miWLufZyBmp3Pd7xd3VgrrS-fGIPW/view?usp=sharing>

