DSC 530 Data Visualization

Project Proposal

- Abhishek Manoj Kumar

New York City taxi datasets

I obtained the datasets for different taxi services in New York from kaggle and NYC. Different taxi services involve Uber, Yellow cabs and Green cabs.

URL:
https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city/kernels

cab:http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

Semantics:
Date and time: (Quantitative) States the date and time at which the cab was booked.
Latitude and Longitude: (Quantitative) Specifies the exact location of pickup and drop-off.
Base: (Categorical) The licencing company code.
Passenger: (Quantitative) specifies the number of passengers.
Trip distance: (Quantitative) distance travelled by the passengers.
Fare amount: (Quantitative) Amount charged for the particular trip.
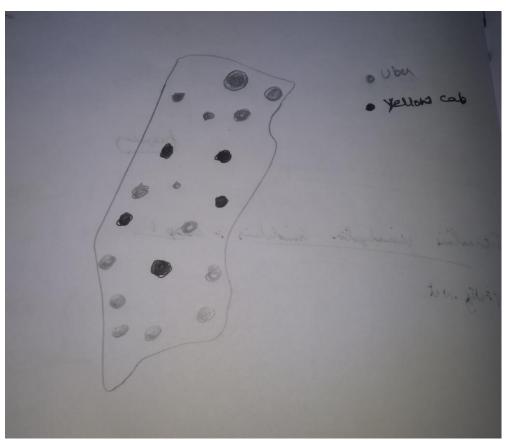Tip amount: (Quantitative) Tip given by the customer.

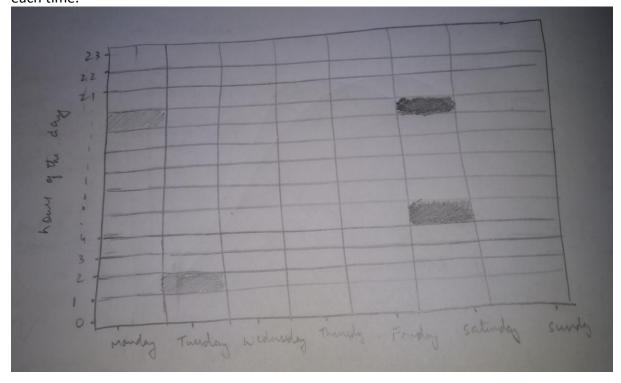Questions I will try to answer using visualization:
1. What cab services are most used based on part of the city i.e. is there a relation between the cab service and parts of New York.
2. Did any cab service observe any decline or increase in business due to other cab services or fair.
3. Does the distance effect people's choice of cab service i.e. certain cab services might be cheaper for longer distances and certain for shorter distances which can have an impact on the cab service chosen.
4. What are the trends observed in tips given to cab drivers i.e. are certain times and days of the week observe better tips than the rest of the week.
5. Does time have an impact on passenger count? I.e. when do people travel alone and in groups.
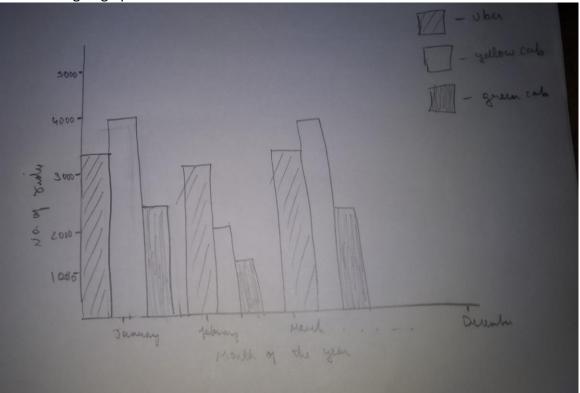
Initial designs:

1. To answer the question 1 I will use the bubble map on New York City map showing the density of cabs and how each cab service provider is spread out throughout the city. The different color of the circle specifies the different cab services
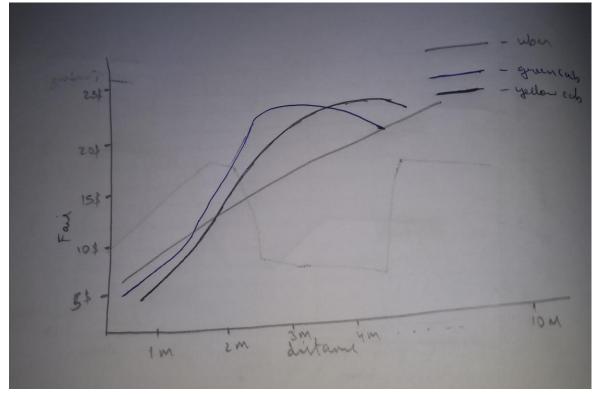
2. To answer the question 1 I am also using the Heat map that showing the trends of rides used at each hour of each day. The sequential color will show the no of rides at each time.
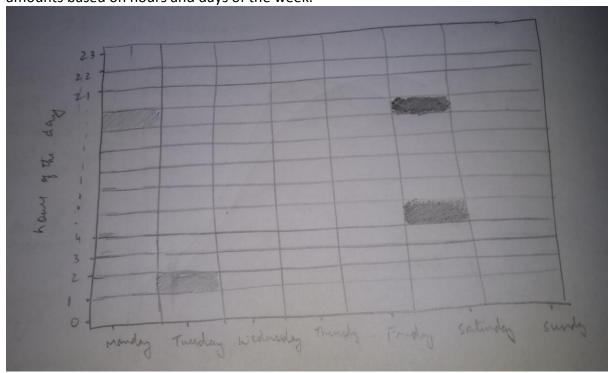
3. To answer the second question I am going to use the Grouped bar chart showing the number of rides each month and observing the General trend if the particular cab business is going up or down or is constant.
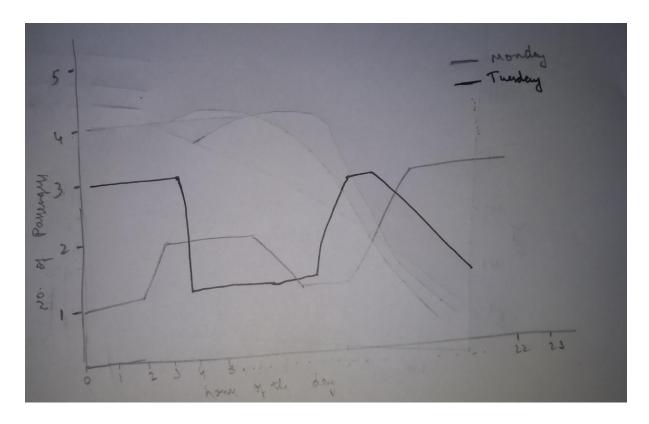

4. To answer the question 3 Line chart showing the trend between fair and distance. So does fair have impact on the chosen service, as we don't have uber fare data, but uber shared there estimate fare which we will be calculating based on the distance travelled in other cab services.

5. To answer the question 4 I am again using the Heat map that showing the tip amounts based on hours and days of the week.

6. To answer the question 5 I am using the Line chart showing the number of passengers on particular hours of the days. People travel in groups in certain times and this information can be useful to show trend which cab is good to use as shared ride.

Monday
Tuesday

5
4
3
2
1

No. of Passenger

0 1 2 3 4 5 ....   22 23

hour of the day

Interactions:

The chart are interlinked in the fallowing way, the heat map is showing the particular day and hour, so by selecting any day in the heat map, the map to answer question1 will also get updated and also same thing is highlighted in Tip heat map. As the bar chart is showing the overall trend of each month, but selecting the particular day and hour in heat map, the bar chart will also show the trend of which service is used most. There will be drop down for particular month, so by selecting the month every chart will be updated. The visualization # 6 is also linked to heat map.