NYC Yellow Cab Data Visualization

Team:

1. Arun Kodnani (ak6384)
2. Pranav Mandala Reddy (pm2696)
3. Rahul Keshwani (ryk248)

1) Project Description

As we know public transportation is one of the most important part of New York City, through this data visualization project we aim at analysing the yellow cabs data provided by NYC Open Data source and generate visualizations that would help a New York cab rider to make a decision of whether to take a cab or prefer using the subway.

2) Data Set

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

Dataset Attributes

Attribute	**Type**	**Values**	**Range**
Pickup Datetime	Temporal, Quantitative	Jan 04, Jan 07	Jan 01 - Jan 31 2018
Drop-Off Datetime	Temporal, Quantitative	Jan 08, Jan 09	Jan 01 - Jan 31 2018
Passenger Count	Quantitative	2,3,4,5	1 - 9
Trip Distance	Quantitative	200.8, 181.9 etc	0.25 - 252.1
Pickup Location	Categorical	Astoria, Battery Park	
Drop-Off Location	Categorical	Astoria, Battery Park	
Fare amount	Quantitative	700, 654 etc	-198 - 888

Tip amount	Quantitative	10, 20 ,30 etc	0 - 310
Tolls amount	Quantitative	20.76, 89.7, 888	0 - 910.5
Total amount	Quantitative	-29, 87, 67 etc	-209.3 - 977.3
Payment Type	Categorical	Credit card, cash etc	4 Categories

3) Analytical Questions

Question 1:

Is there any hourly pattern in the trip frequency of yellow cabs?
Proxy Task: Finding the number of cab requests throughout the day(based on **Pickup time**)

Question 2:

How long does an average New Yorker travel in a cab?
Proxy Task: This can be attributed to the **Trip Distance** for all the cab rides.

Question 3:

How do yellow cabs distribute spatially?
Proxy Task: Identifying localities (such as boroughs based on **Pickup** and **Drop-Off Location**) and calculating the density of cabs in that locality.

Question 4:

Does yellow cab count is affected by the type of day such as weekend or weekdays?
Proxy Task: Calculating number of cab request for each day of the week(**Pickup Datetime**).

Story Design

Storyboard:

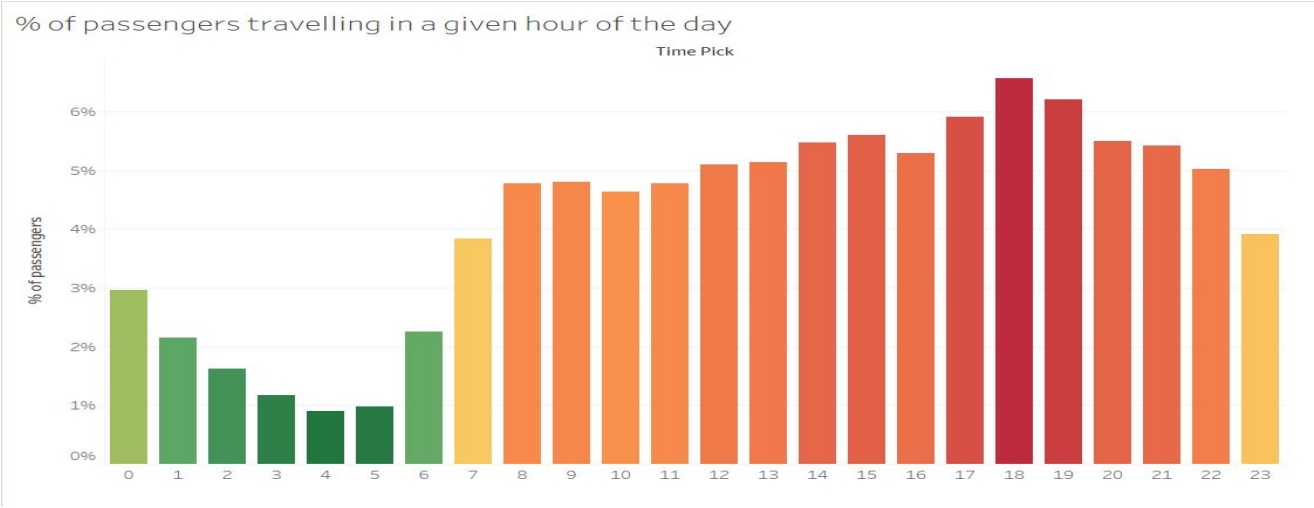
Story: “What if he took a cab instead?”
Storyboard link:
https://docs.google.com/presentation/d/108-Fqr1C3jquG0utmrxSTTDUkn-3-vI6REDlzOsOLXA/edit?usp=sharing

So we have decide to divide our story in three major sections:

1. **Pretext:** Our protagonist “Rahul” overslept and is now late for the movie “Aquaman”. He prefers traveling via subway, but is wondering if he will be able to make it on time this time? And so, he contemplates that whether **he should take a cab instead?**
2. **Contemplate:** In order to make his life easier, Rahul decides to take a virtual cab tour (through our visualizations).
    a. He first checks the time and wonders,”**Will there be a cab easily available at this hour?”**
    b. He then looks for cabs near his house and asks himself, **“But, how easy is it to get a cab from my location?”**
    c. After not being able to find a cab for a couple of minutes he thinks, **“Oh shit! Is it a weekend or a weekday?**”
    d. After some struggle he finally gets a cab and is currently on his way to the theatre wondering **“if I’ll reach before the movie starts?”**
3. **Conclusion(Open ending):** After having the virtual cab ride Rahul has all the information he needs to make his decision. If you want helped Rahul decide with his decision, don’t forget to check our storyboard link.

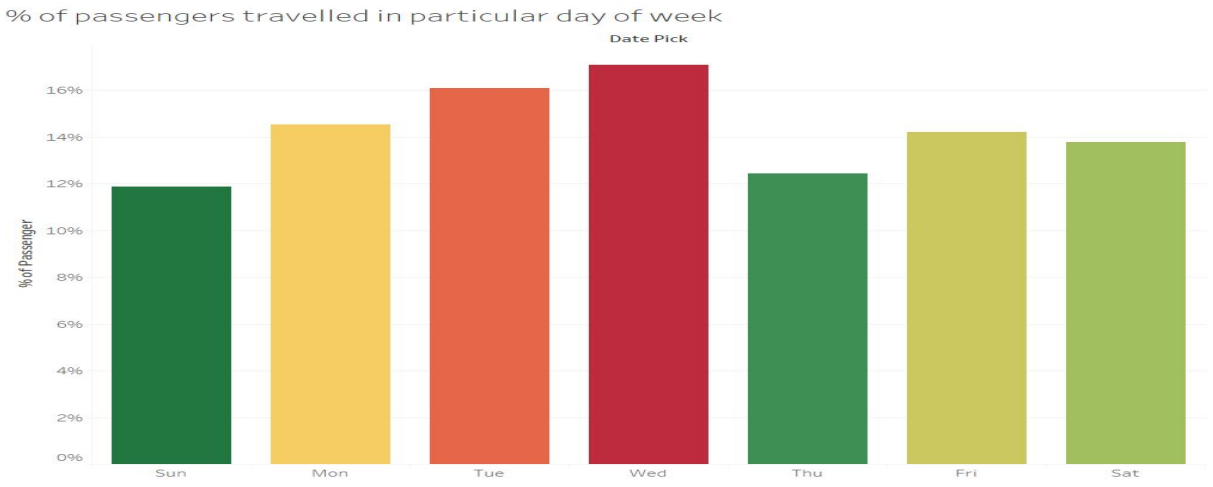
4) Data Analysis and Sketches

1. **Is there any hourly pattern in the trip frequency of yellow cabs?**



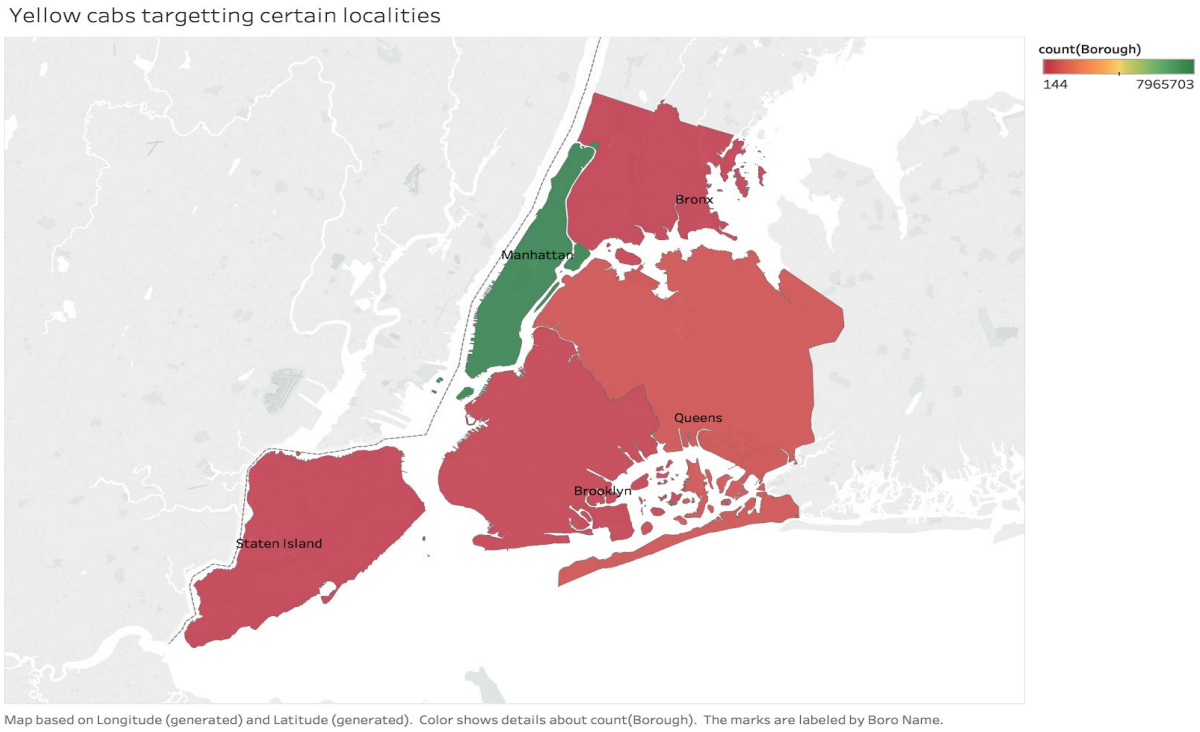
Explanation - As we can see that **evening 6 PM** is the busiest hour for the cab drivers. Also, there are few cabs requested from **Morning 12 AM to 6 AM.**

2. Does yellow cab count is affected by the type of day such as weekend or weekdays?



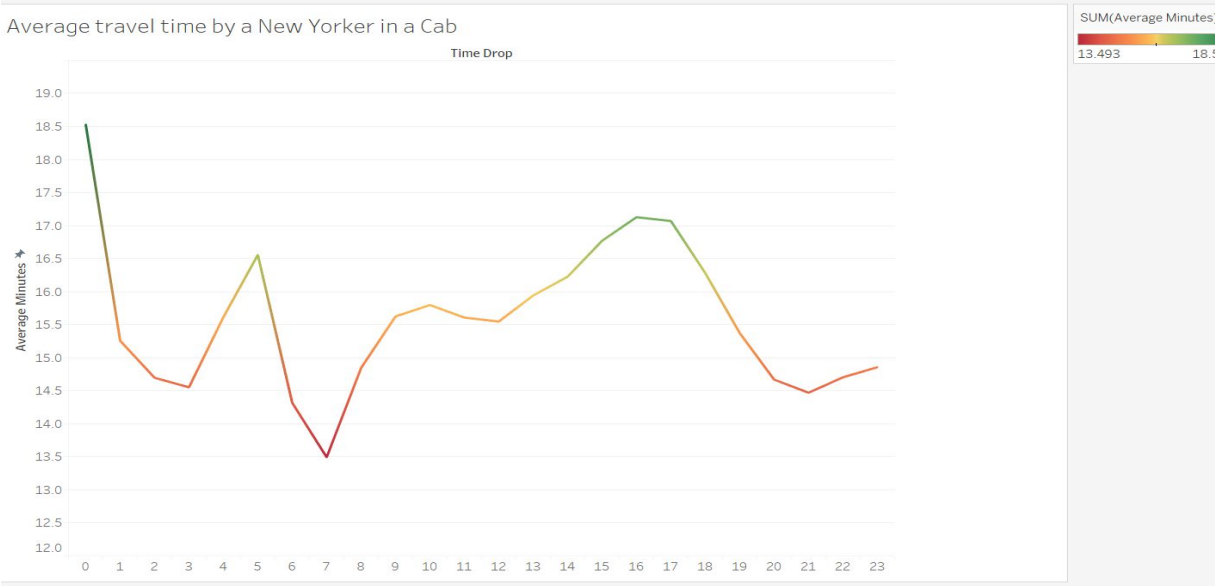
Explanation - Yes, it seems that the cab count is affected by the day of the week. **Sundays** generally have **fewer cab rides** compared to other weekdays. We also see a trend where in the cab rides increases for the first 3 days of the week and then drops or stays almost the same for the next 3 days.

3. How do yellow cabs distribute spatially?



Explanation - From this map visualization we can clearly see that **Manhattan** has the highest number of cabs and on the other hand **Staten Island** has the fewest number of cabs.

4. How long does an average New Yorker travel in a cab?



Explanation - The above line graph shows if there is a drop that indicates that it takes less time to travel and if there is a spike, the average New yorker travels more.

5) Implementation

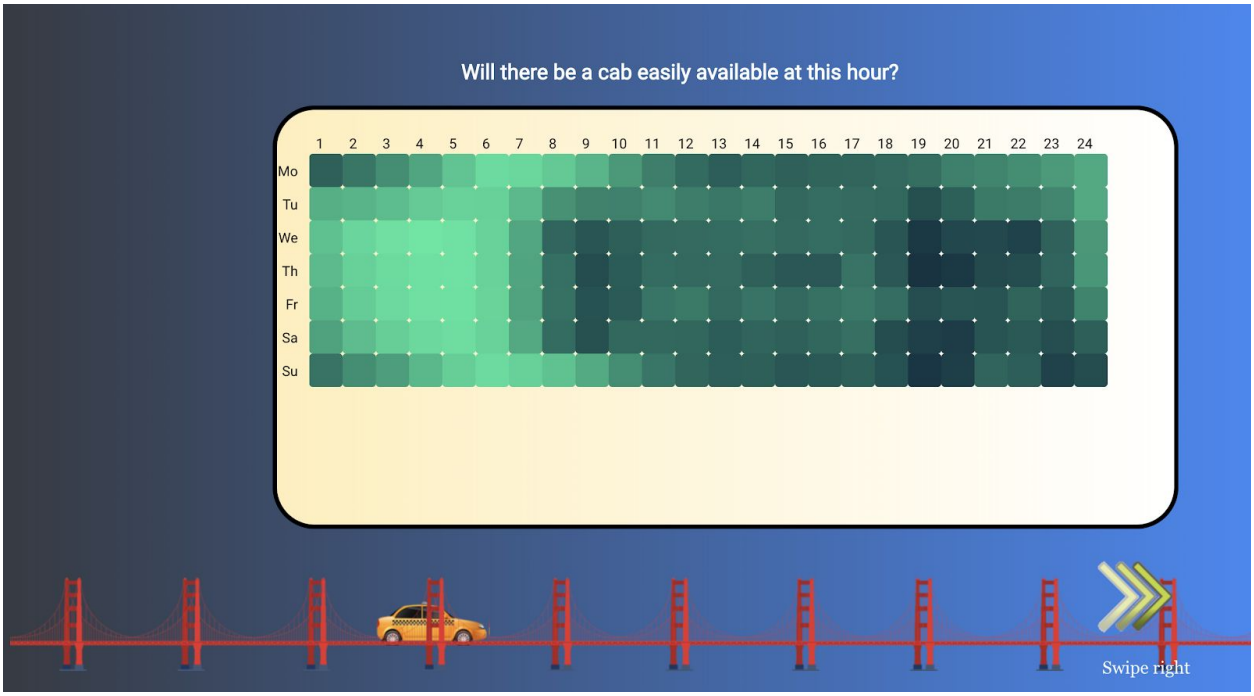
GITHUB LINK:
https://github.com/NYU-VIS-FALL2018/storytelling-group-10-nyc-cab-data-visualization

DEMO PAGE:
https://nyu-vis-fall2018.github.io/storytelling-group-10-nyc-cab-data-visualization/

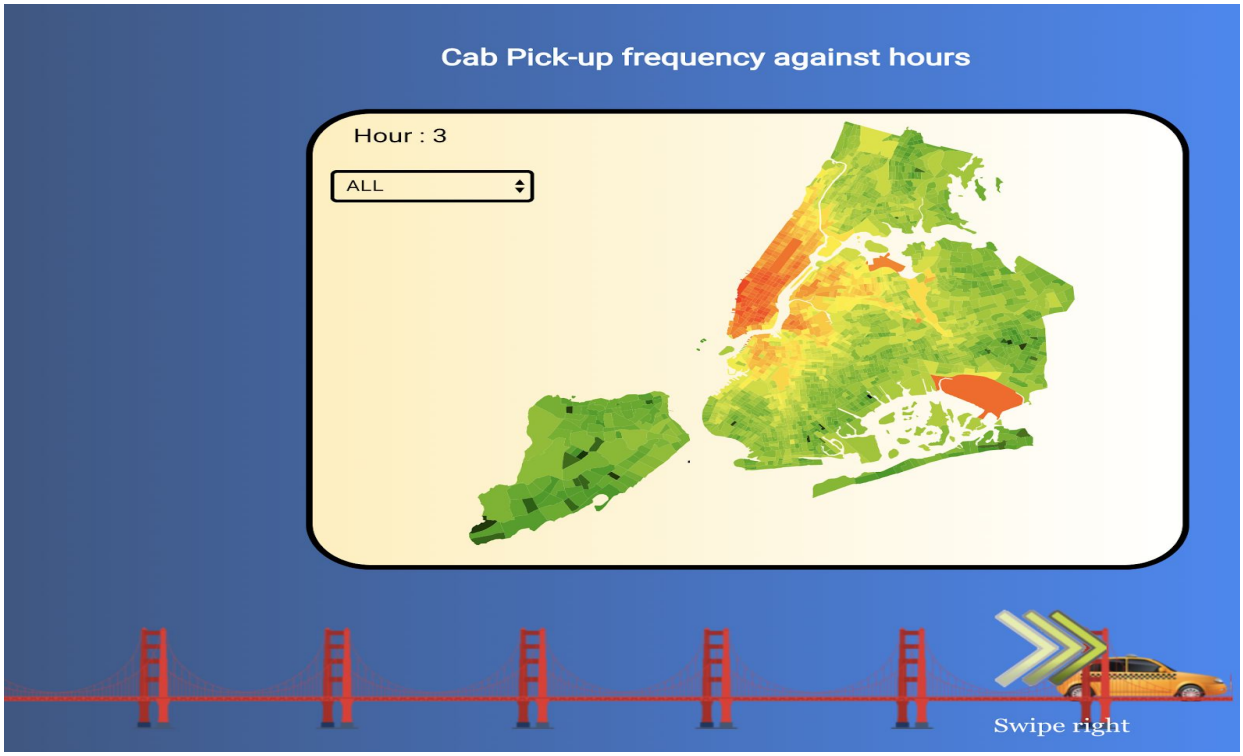
CHANGE LOG:

1) Modified our project goal a bit in order to make it more specific as suggested by the Professor.

2) We also modified our visualizations for Question 1 and Question 4. Instead of making 2 different bar graphs, we have now implemented a **"Heatmap".** In the Heatmap, the x-axis denotes the hour of the day, the y-axis denoted the day of the week and the pixel intensity denotes the availability of the cab.



3) The map visualization for the question **“How do yellow cabs distribute spatially?”** is now updated and can be used to visualize the distribution of yellow cabs spatially along with how the distribution changes by the hour.



4) Lastly, a pickup and dropoff location filter is added for the user to visualize the data according to his/her own source and destination locations.

FEEDBACK:

1) The first feedback that we received was modifying the project goal and to update the visualization for Proxy task 1 and use ratio of cab riders to the number of cabs instead. Well, we could not find a dataset which gives us the number of cabs by the hour of the day. Hence, we are unable to incorporate the change of calculating the ratio of cab riders to cabs.
2) The second feedback that we received from the professor was adding interactivity to our graph so that the user can visualize the data according to his/her source and destination location. We have successfully implemented this feedback and added interactivity to all our graphs.