Restaurant Analyzer and Recommender

Group No. 5 Offline

Anshul Sharma, as10950
Avaiyang Garg, ag6026
Archit Goyal, ag6288

Project goal

Our project goal is to analyze Yelp data and visualize the data using map views and other graphs to answer various questions to determine the best restaurant in the proximity with a type of cuisine which the user wants to eat. Users can locate their location and choose the type of cuisine they want and filter them based on their ratings. This will help them eat their favorite cuisine in the proximity at a place which has good ratings and reviews in the comments. We have tried to make this an interactive webpage for the user so that they can add filters and choose a place to eat with best of their needs.

Main objectives:

1. Build interactive application to answer questions about restaurants in NYC
2. Show which cuisines are most common on different boroughs.
3. Show how cuisine preferences change according to borough in NYC.
4. Show where restaurants cluster more in NYC.
5. Show most high rated restaurants in five boroughs and within boroughs too.

Data Set

URL for the DataSet: https://www.yelp.com/dataset/challenge

Dataset files used:
1. Business.json
2. Review.json
3. Checkin.json
4. user.json

Attributes used from the data set are :
1. Business.json - business_id, name, city, address, postal code, longitude, latitude, stars, review_count, category, outDoorSitting
2. Review.json - review_id, business_id, stars
3. Checkin.json - business_id, time.
4. User.json - user_id, user_count, reviews.

Analytical Questions and Proxy Tasks

1. Where are the restaurants clustered in New York City Area?
   Proxy Tasks:
   1) How is area divided in New York City? Area in New york city can be divided among the boroughs or the zip codes.
   Proxy Values:
   1) Lattitude, Longitude and city: From these three we can find the borough they belong to and where on map we can plot and identify them.

2. Which Borough in New York City has top average rated restaurants?
   Proxy Tasks:
   1) What is the significance of "top"? Listing the restaurant with the most number of 5 stars because there may be restaurants with a single user review but a 5-star rating.
   Proxy Values:
   1) We will be using the Review.json file, in this, we will be using stars, and in the business.json file, we will use stars and review_count.

3. Which cuisine is most common in different boroughs of New York City?
   Proxy Tasks:
   1) Based on the current location, there may be many different types of cuisines which the person can have as options to eat. They can select the nearest one and eat there.
   Proxy Values:
   1) business.json file provide us with the categories which will help us to determine the different cuisines and then the latitude, longitude, and postal code will provide the location of the restaurant.

4. Which are the top rated restaurants in different boroughs of New York City?
   Proxy Tasks:
   1) What is the significance of "top"? Listing the restaurant with the most number of 5 stars because there may be restaurants with a single user review but a 5-star rating.
   Proxy Values:
   2) We will be using the Review.json file, in this, we will be using stars, and in the business.json file, we will use stars and review_count.

5. How cuisine preference change from overall New York to different boroughs?
   Proxy Task:
   1) What does change mean here? Change means how the count of restaurants of different cuisines change from whole new york city to individual boroughs.
   Proxy Value: Count of Restaurants for different cuisines
   2) What is preference?
   Proxy Value: Preference is determined by the highest count of restaurants for the cuisine.

Story Design

Data Analysis

The answer to the above-mentioned questions are the following:

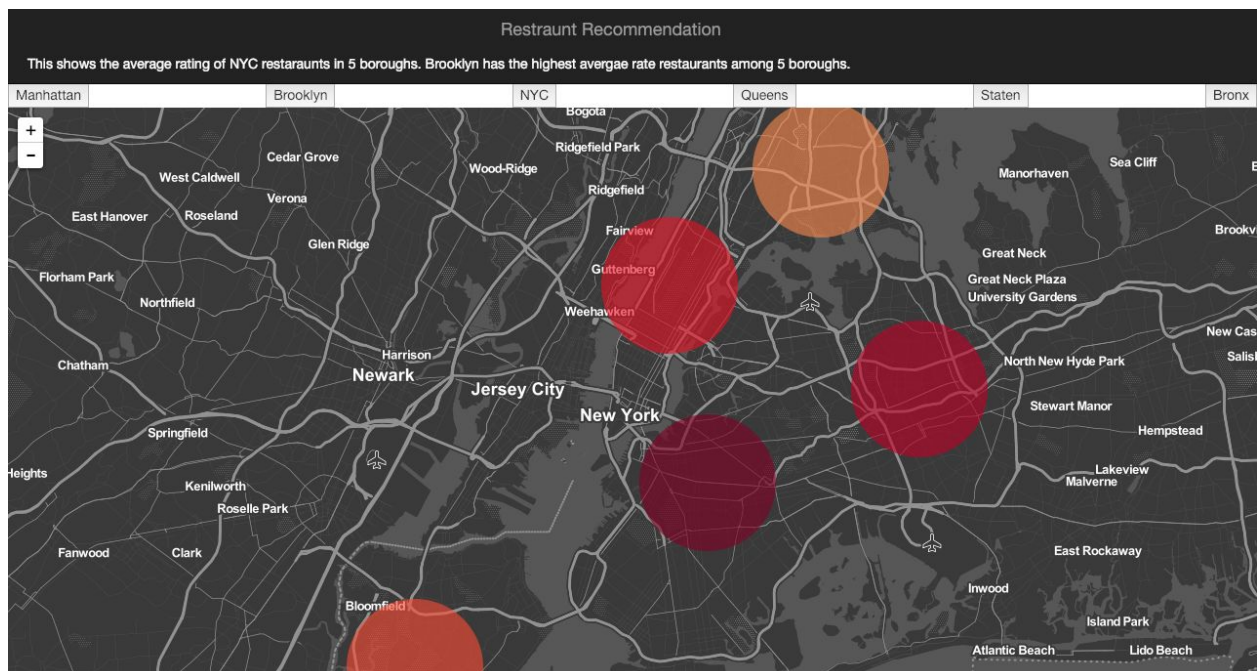1. Restaurants in New York City.

This is the overall map of New York City showing all the restaurants based on the latitude and longitude plotted on the map. The color range is from yellow to red for restaurants on the basis of their ratings. Red being 5 rating and yellow being lowest rated. For the image we analysed that among the five boroughs, manhattan has the most number of restaurants.
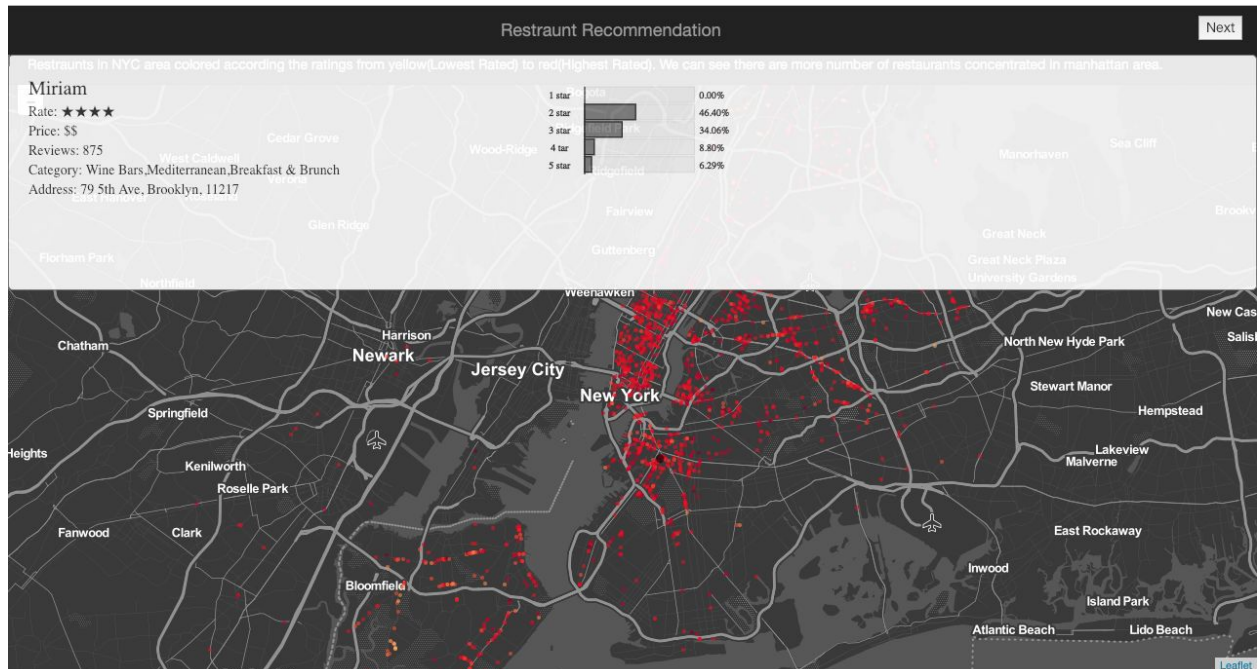
This map shows the number of restaurants when the map is focussed on a particular region and also showing details of the selected restaurant.
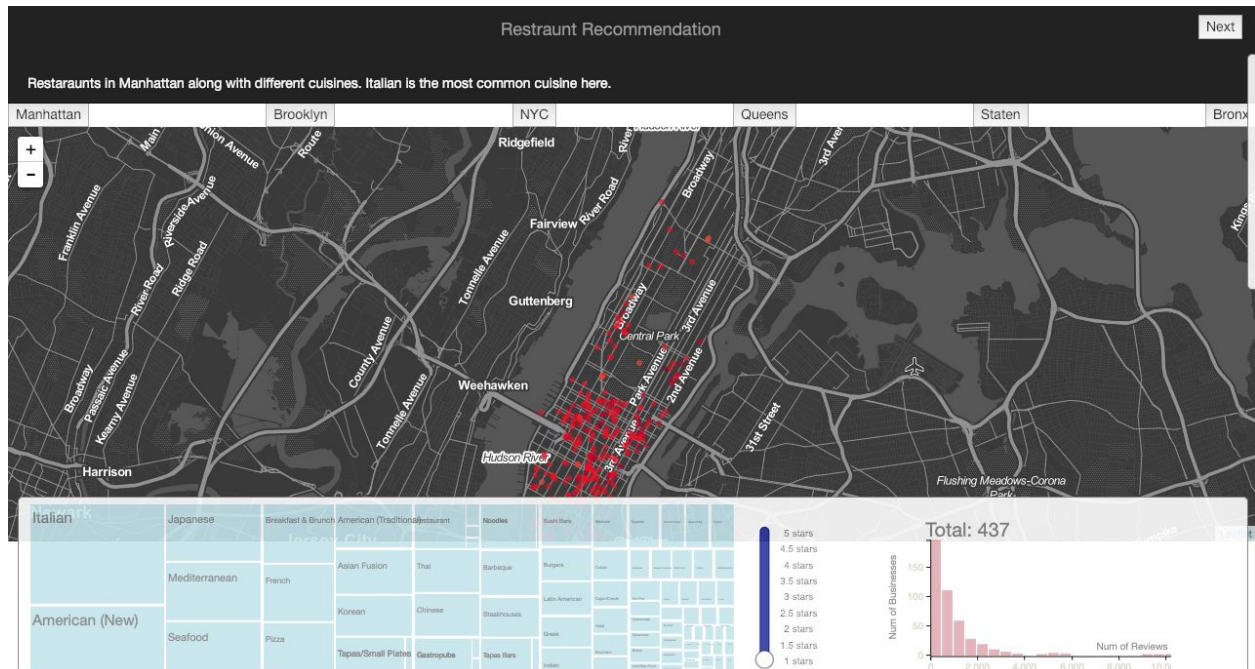
2. Which Borough have the top average rated restaurant in New York City?


The above map shows that Brooklyn has the highest average of restaurants in New York City.

The Average rating Goes by Brooklyn -> Queens -> Manhattan -> Staten Island -> Bronx.

3. Most common cuisine in different boroughs of New York City.

This shows the restaurants in Manhattan area of New York City and below is the tree map, which shows which cuisine has the highest number of restaurants in the area.
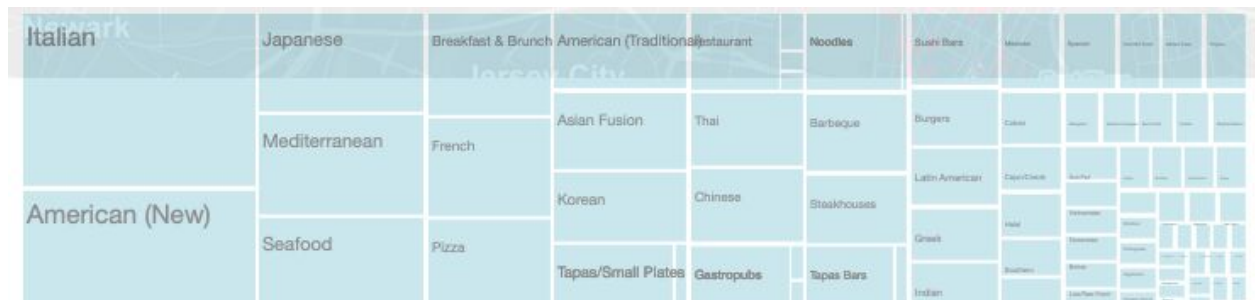
4. Top rated restaurants in different Boroughs of New York City.

For each borough, we can see the color of restaurants in the brough, which signifies their ratings and also we can change the ratings of the restaurants to filter the values.

5. Change in cuisine preference from whole New York City to different boroughs.

The above two graphs show that which cuisine was most common in the New York City area and which cuisine is most common in the given borough.

Storyboard

There have been many mobile applications which help users in finding the restaurants of their choice, but there aren't in web applications. Web applications which we are creating will provide a much clear approach with all necessary question and answers and will allow the user to filter appropriately and show the results on the map, thus the user will have a better user experience. The user along with filtering is able to simultaneously look the result on the map and if the user is not able to get some good results, he can just drag the slider to change the filter value and then get the desired results. Based on our experience, we decided that few key factors are very important while designing our story, like the average rating of the business, the location, number of ratings of the user, stars, review counts. These key factor will help us in better designing our story and also help in a better user experience.

With this web application, we will be taking step by step approach towards helping the user in finding the overview of restaurants in New York City and its different Boroughs. The first step taken would be to provide the user with the number of restaurants within the proximity of the user. Now the user can also have the restaurants with the filter of the borough or even the zip code. This can be visualized on the basis of latitude and longitude in the dataset. The next step which comes is to filter the restaurants on the basis of the desired cuisine. From the dataset, we can get which all cuisines are served by the restaurant and then filter accordingly. Once these are filters, now the user can filter the restaurants on the basis of their star rating, since each restaurant has the star count in the dataset and the user review counts.

Steps:

1. We have made the graph where we use the longitude and latitude to plot the restaurants on the map. We are using the color shade from yellow to red to show the rating of the restaurant where if the restaurant has a red then it has a higher rating and the restaurant with yellow color it has a lower rating.We are plotting the restaurants on the map using the longitude and latitude.

**Map with Circles showing the location of restaurant**

2. After this we will answer the question to show average of ratings of restaurants in different boroughs. We can clearly see that brooklyn has highest average restaurant rating over all other boroughs.
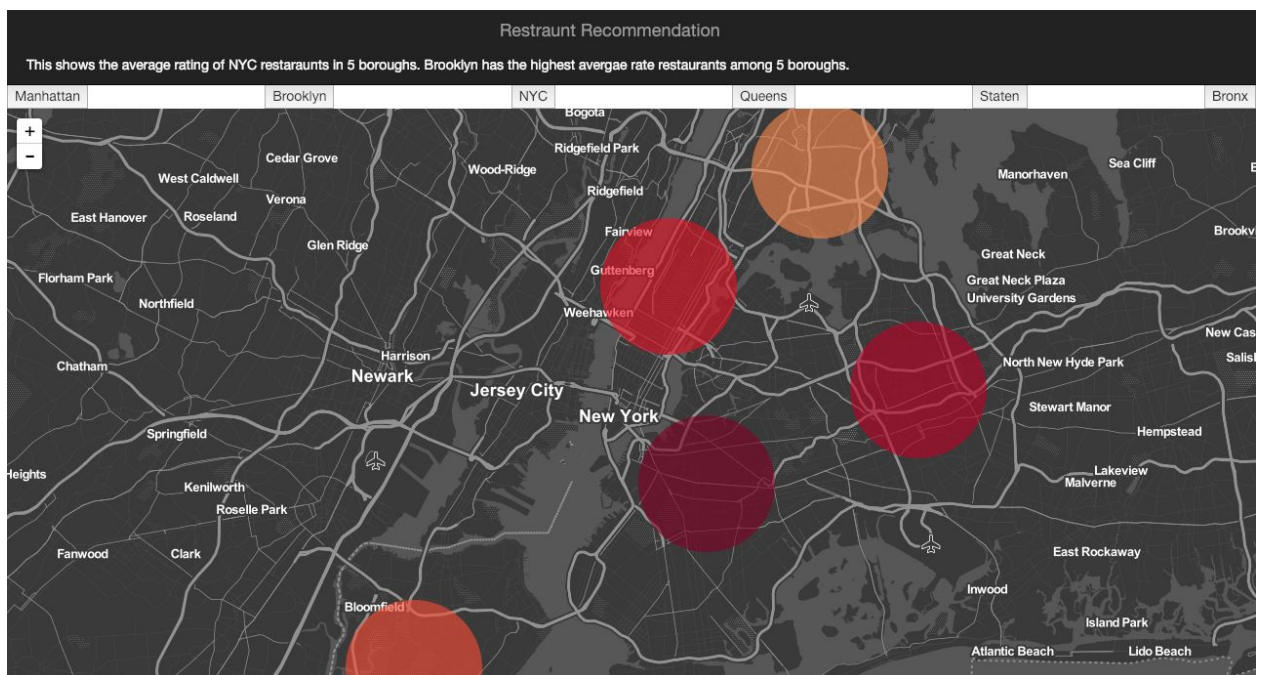
2. After the graph is made, we will add the treemap, which will show the various cuisines present in our dataset. Size of each cuisines depend on the star rating present corresponding to the restaurants. We figured the Italian is the most common cuisne present in manhattan.
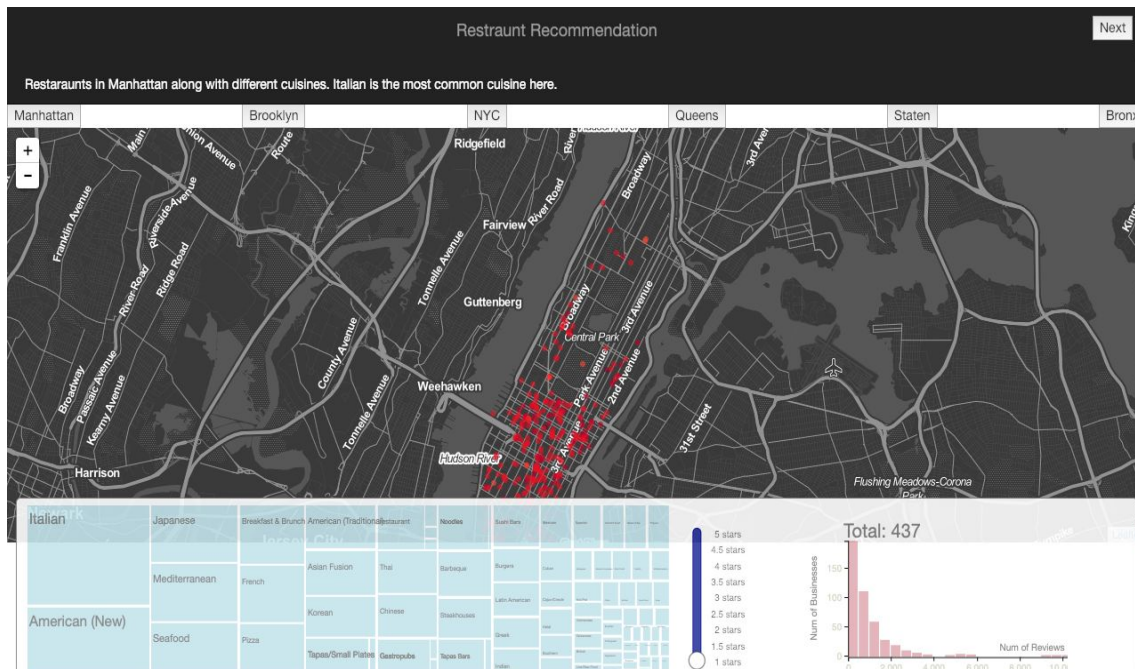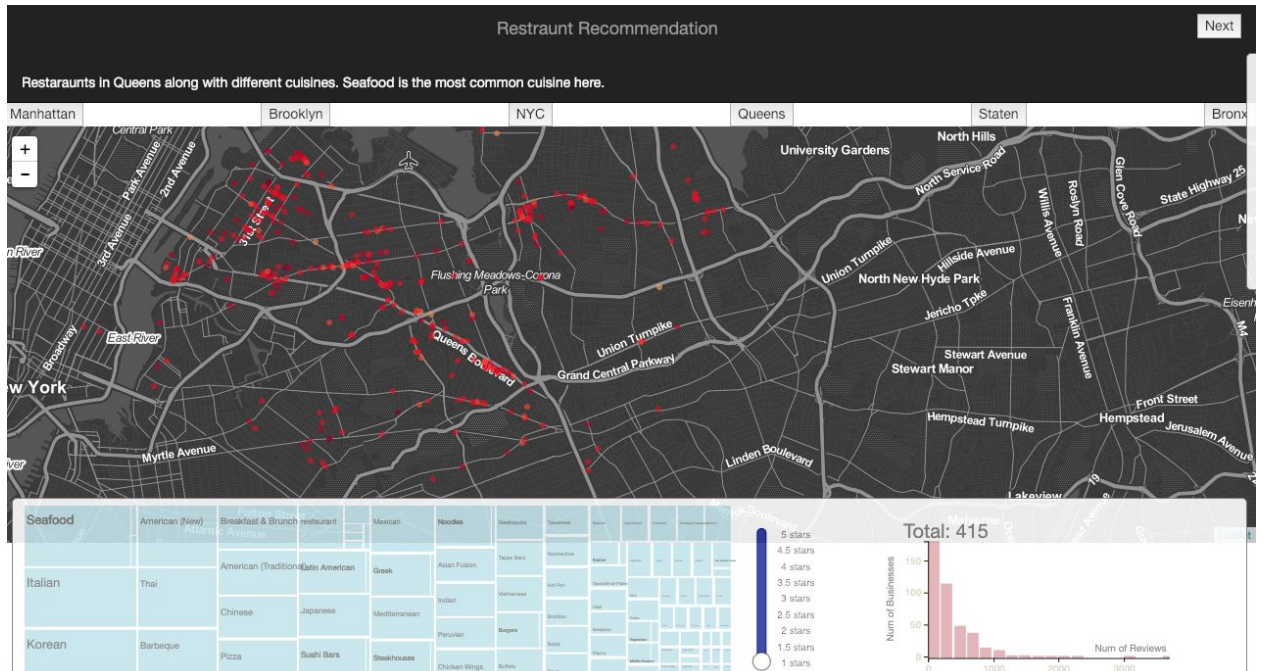

Tree Map:
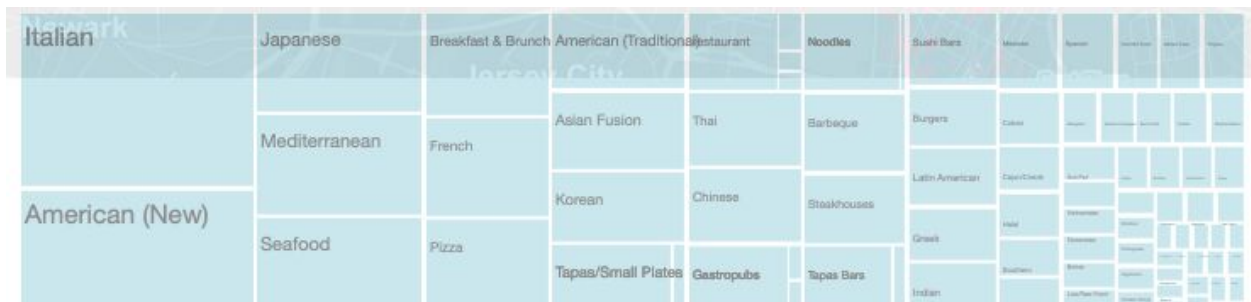
3. We can click on each restaurant and figure out which rating has the most number of reviews. As here, we can see most of the reviews are 62.02%.
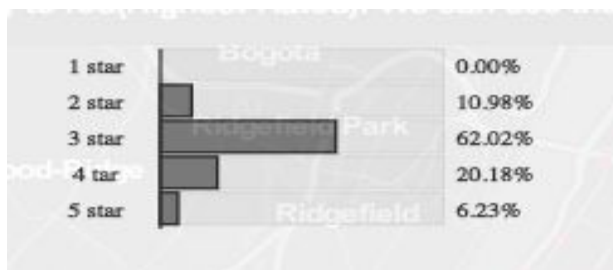
4. After this we can figure out which cuisine and what kind of rating of restaurants are present in other boughs as well.

4. Then Finally we will be able to see the difference in most common cuisine in whole New York vs most common cuisine in a particular borough.

5. Our final solution will look something like this and will be interactive for users to show the results.

Initial Sketches:

1. The sketch for the first page of the project:

Map based on Longitude and Latitude.

2. The tree map to show the cuisine preference:

3. Review count for each business histogram:
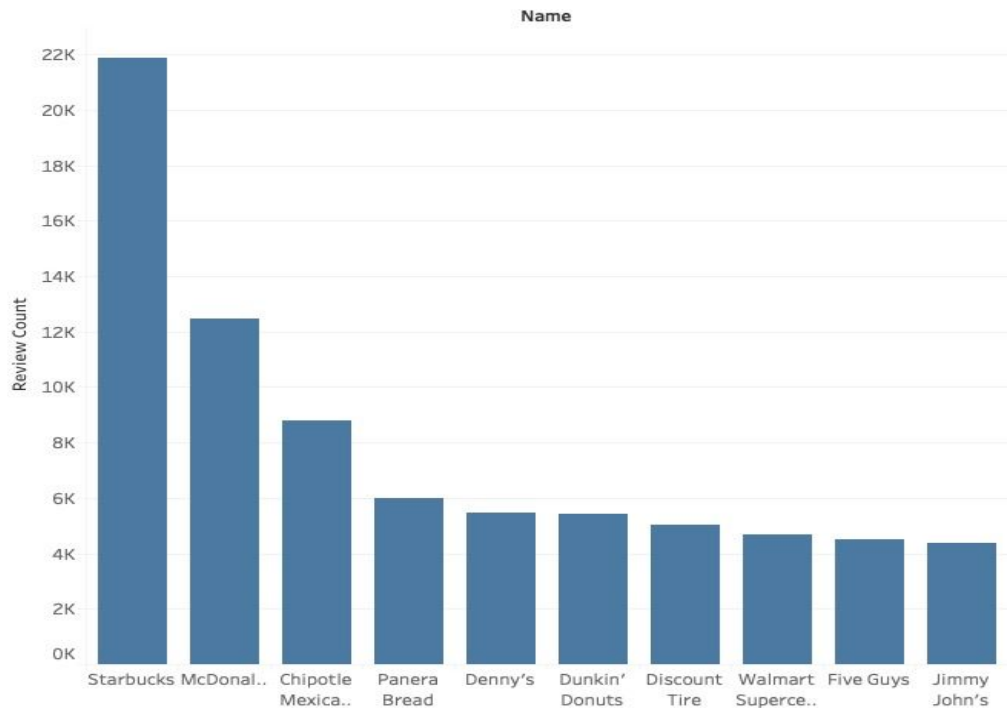
4. Final Solution sketch:

Changelogs:

1. Changed the Analytical Questions as they were very broad. We narrowed it down in a specific direction after the reviews from previous submission
2. Added Proxy Values.
3. Added more specific questions.
4. Made the UI more interactive and intuitive so that user can have a good idea of restaurants in proximity just by looking at map and zoom.
5. Clear description of the objectives based on the comments received in the previous assignment.
6. Added video and actual updated github link.

Implementations:

1. Data processing and transformation :

a. We had multiple data set from which included business data, user data, checkin data, tip data, review data. Out of which we combined the business and user data through business id and user id. And later we plan to join checkin as well.

b. Filtered the dataset on the basis on the required location only as the data contained multiple locations.

2. Created an interactive webpage with a Map and Bar chart.

3. When we hover over the plotted points, then an additional bar is displayed on the screen which gives us details of the restaurant.
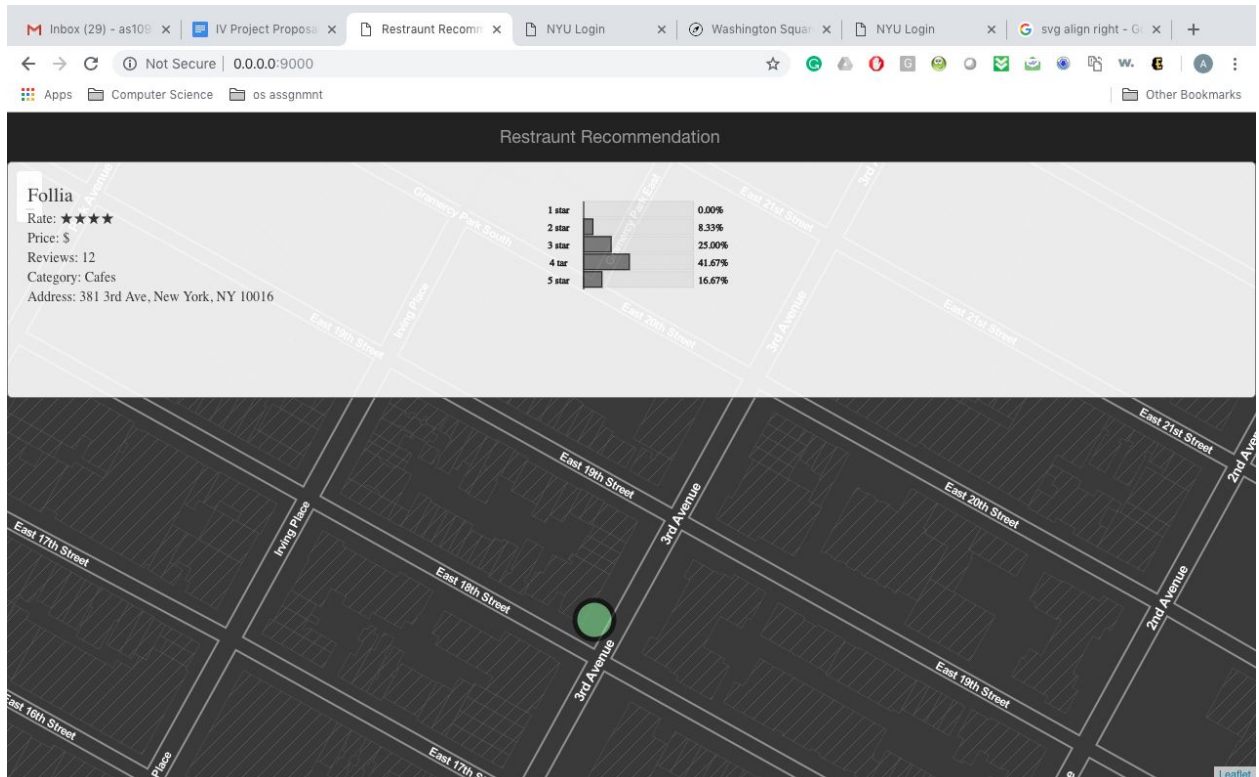

4. Restaurant details include the :
   a. Restaurant name
   b. Rating(number of stars)
   c. Number of Reviews posted
   d. Category(Cafe, Restaurant, Italian, Mediterranean)
   e. Address of the restaurant

5. The color of the plotted points depend on the star ratings. If the rating is less the shade of the point is lighter and vice versa.

Ratings	Shade
5	
4	
3	
2	
1	
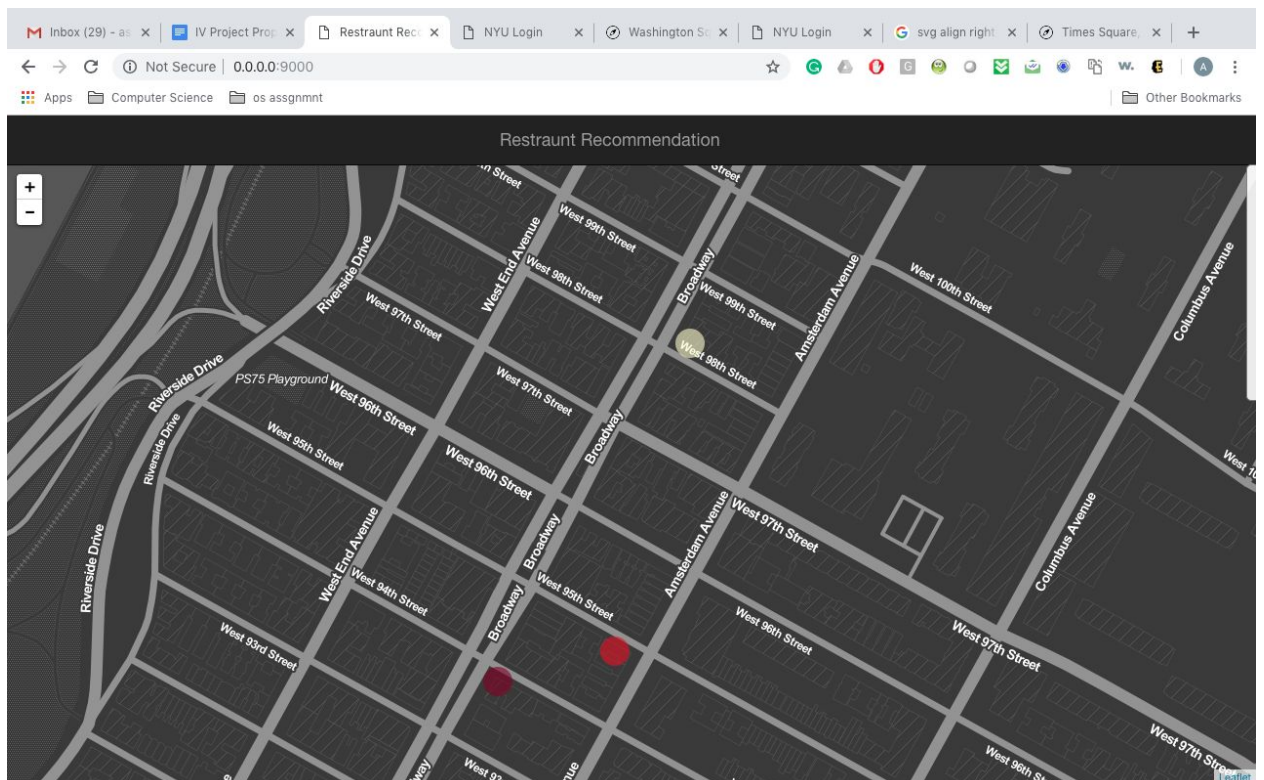
6. When we hover over the point, the details appear and when we move the cursor away, the details go away.
7. When we click on the plotted point, the details appear and are pinned to the screen for the user to view it.
8. The bar graphs shows the percentage of total number of reviews for each rating.
9. Tree map will also change based on the zoom in and zoom out of the map which means that it will only plot tree map for the list of restaurants shown on the screen

10. Created a tree map with all types of cuisine for each borough.
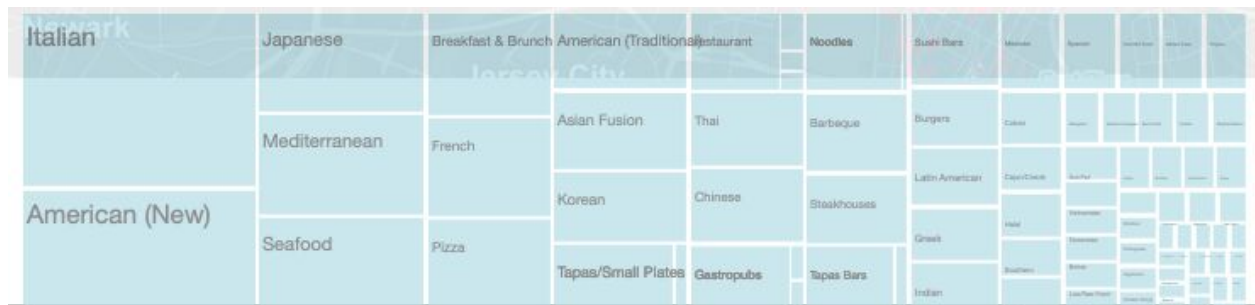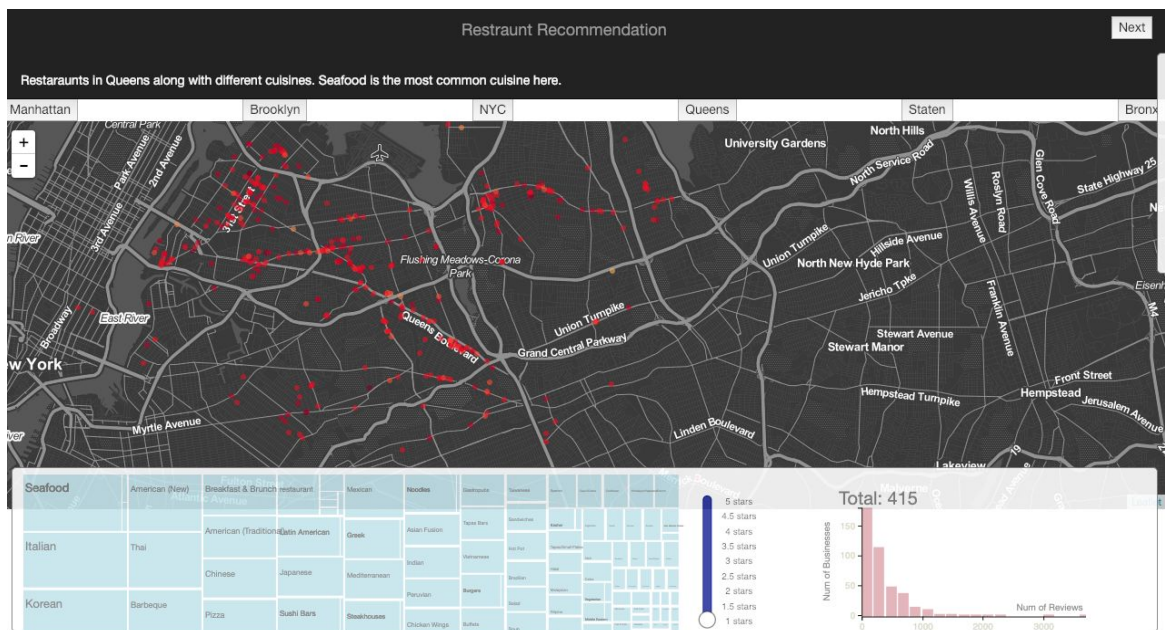
11. Created a filter for the user so that they can filter the restaurants based on their ratings.

12. Updated UI to make the app more interactive and each page answering more questions

13. Created a bar graph to show the cuisine preference of whole New York vs preference in individual boroughs.

Implementation Links:

GitHub Link: https://github.com/NYU-VIS-FALL2018/storytelling-group-5-offline

Demo Link: https://nyu-vis-fall2018.github.io/storytelling-group-5-offline/

Demo Video Link(This is the video reated by us to show how to navigate between pages and what question they answer):
https://drive.google.com/file/d/182rBHppIN1On16LHG9tIRIuzMUf_Jw46/view