Restaurant Recommendation System for the Neighborhoods of New York

Prashant H Krishnan

March 07, 2019

1. Introduction

The aim of this document is to explore the idea of a recommendation system that would recommend restaurants to people in the New York neighborhood. This report describes the work to learn to predict and make recommendations to people in New York based on location data related to the restaurants in the neighborhood of New York and the Ratings given to these restaurants by others.

1.1 Business Problem

New York is one of the most populous cities in the United States and in the World. This city is described as the cultural, financial and media capital of the world. This brings a lot of travelers to this city.

People visiting new places would love to get recommendations about things to do, about places to see and about places to dine in. This has resulted in the increased popularity of several recommendation websites and apps like Yelp that provide recommendations to people. In addition, in a populous place like New York where life is fast paced, it would be of great help if people are able to find useful recommendations that would save them a lot of time.

Being able to recommend restaurants for people to dine in when they are travelling or visiting a new place like New York is a very valuable feature as far as the recommendation system goes. In this document we will try and build a model that would recommend restaurants to users in the neighborhoods of New York based on what is popular in a certain place based on ratings and reviews provided by other people and also based on proximity to the user.

In this document we will try to explore and optimize the algorithms to provide a recommendation of the top restaurants to the user and try and validate the accuracy of the model. In our model we will try to take into account the location of the user and also the popularity of the restaurants in the neighborhoods of New York.

2. Data

2.1 Data Requirements

To be able to make recommendations on Restaurants in the Neighborhoods of New York, we will need some data that will form the basis for our recommendation.

How can we make recommendations regarding restaurants to a user? For being able to make recommendations we will need to understand the users preferences in terms of what kind of food they like to eat and their current location to be able to find restaurants near them.

Is it enough if we understand the users' preferences and location for being able to make a recommendation? No, in addition to understanding the preferences and the location of the user for whom we are making the recommendation, we will also need to identify the neighborhoods of New York and then get the list of restaurants that are present in the neighborhoods of New York and the also understand their popularity to know which ones are good and which ones are not.

2.2 Data Collection

There is a variety of data that we need for building our recommendation system and this will need to be collected from various sources as identified below.

2.2.1 Neighborhood data for New York:

The Neighborhood has a total of 5 boroughs and 306 neighborhoods. We will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This dataset is provided by New York University and is available for free on the web. We will download and use this data for the Neighborhood data.

2.2.2 Restaurant Data for New York:

We will use the Places API provided by Foursquare and gather location data regarding restaurants using their API. For using their API to get the location data pertaining to the restaurants in the neighborhoods of New York, we will need the latitude and longitude coordinates for the neighborhoods. The coordinates are available to us as part of the Neighborhood data that we collected in 1).

2.2.3 Ratings Data for the Restaurants:

We will again use the Places API provided by Foursquare to gather the ratings data for the list of restaurants that we collected in 2).

2.3 Data Cleaning

The New York Neighborhood data was combined and stored along with the Restaurants Data set that was retrieved from Foursquare. This restaurant data set had a lot of duplicate restaurant entries. This data set was processed to remove the duplicate entries. This reduced the size of the restaurant dataset by ~50%.

There were several other problems that were identified with this restaurant dataset. The problems identified are as follows:

- The dataset included data for closed restaurants. Closed restaurants are not required for us for the purpose of making restaurant recommendations for people to eat.
- The dataset included venues / businesses that were not restaurants. Again this is also not required for our purpose since we are only interested in restaurants.
- The data set included restaurants that were not categorized. We need the restaurants to be categorized appropriately for our analysis.
- The data set included restaurants that we incorrectly categorized. Again, we need the restaurants to be categorized appropriately for our analysis.

Based on the above observations, the data had to be cleaned up before we could continue with our analysis. We performed the following cleanup on our dataset to fix the problems observed.

- Closed restaurants were removed from the dataset.
- Venues / businesses that were not restaurants were removed from the dataset.
- The category was corrected manually for the restaurants that were not categorized or were incorrectly categorized. The appropriate category for these restaurants were looked up in the web and then the dataset was updated accordingly.

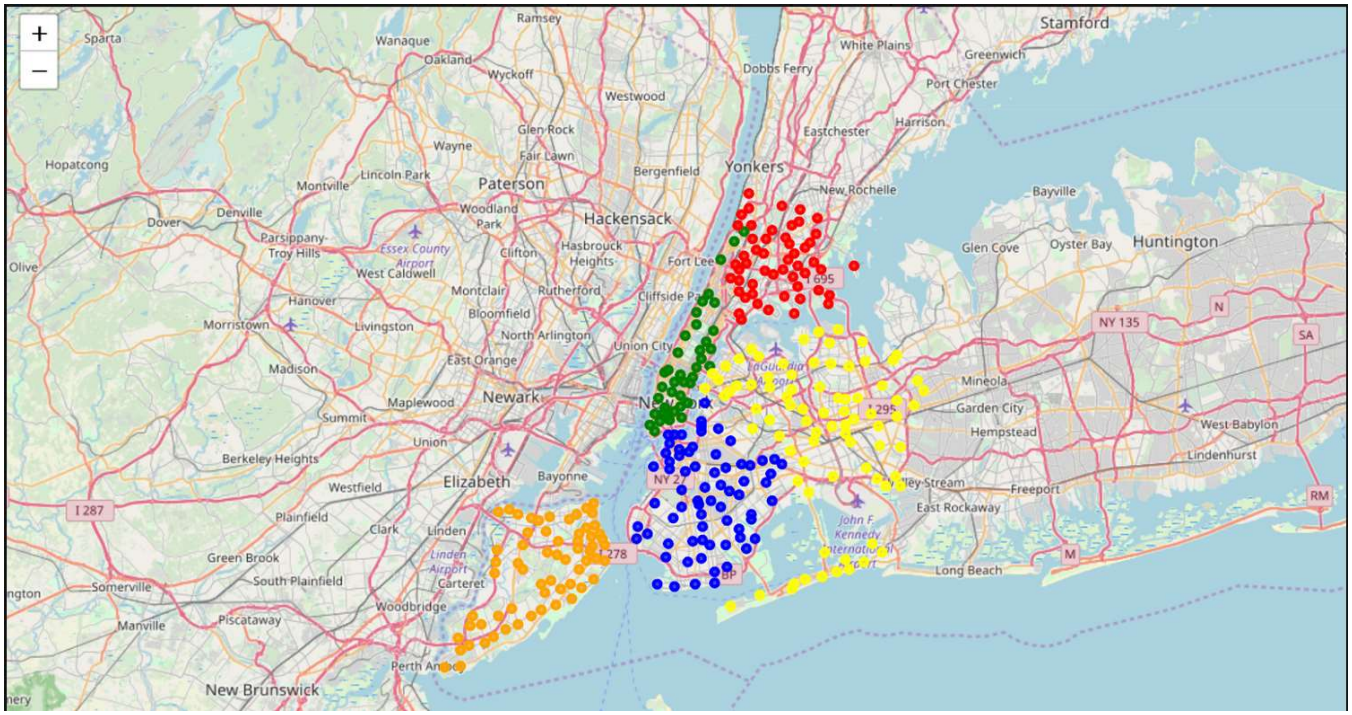
After performing the above mentioned cleanup we now have a good dataset to continue our analysis.

3. Methodology

We will be focusing on finding the restaurants in the neighborhoods of New York in order of their ratings and also in order of distance from where the user is located. The Neighborhood data comprises of the Borough, Neighborhood, Latitude and Longitude information.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Using python's folium library, let us try to visualize these neighborhoods. In the map, each neighborhood is represented as a circle and each borough is represented using a different color.



We then use the Places API from Foursquare to explore these Neighborhoods and search for Restaurants in these Neighborhoods. The Restaurant data retrieved from Foursquare contains information pertaining to the Restaurants, their categories and their location. The data retrieved is as follows.

Restaurant_Name	Restaurant_City	Restaurant_Address	Restaurant_Latitude	Restaurant_Longitude	Restaurant_Category
Big Daddy's Caribbean Taste Restaurant	Bronx	[4406 White Plains Rd (Nereid Avenue), Bronx, ...	40.899767	-73.857135	Caribbean Restaurant
Kaieteur Restaurant & Bakery	Bronx	[4379 White Plains Rd (at E 237th St), Bronx, ...	40.899768	-73.857184	Caribbean Restaurant

We received the above information for 4984 unique restaurants from the Foursquare API. The count of restaurants in each borough is as follows.

Borough	Restaurant Count
Bronx	915
Brooklyn	1516
Manhattan	1001
Queens	1340
Staten Island	212

After receiving this information for the restaurants in the New York neighborhood, the Foursquare Venue Details API was used for getting information about Ratings corresponding to these Restaurants. The Summary of the Ratings received is as follows.

	Borough	Not Rated	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
0	Bronx	824	0	0	0	0	0	9	28	33	26	1
1	Manhattan	1002	0	0	0	0	0	0	0	3	0	0
2	Brooklyn	1510	0	0	0	0	0	0	0	0	0	0
3	Queens	1333	0	0	0	0	0	0	0	0	0	0
4	Staten Island	212	0	0	0	0	0	0	0	0	0	0

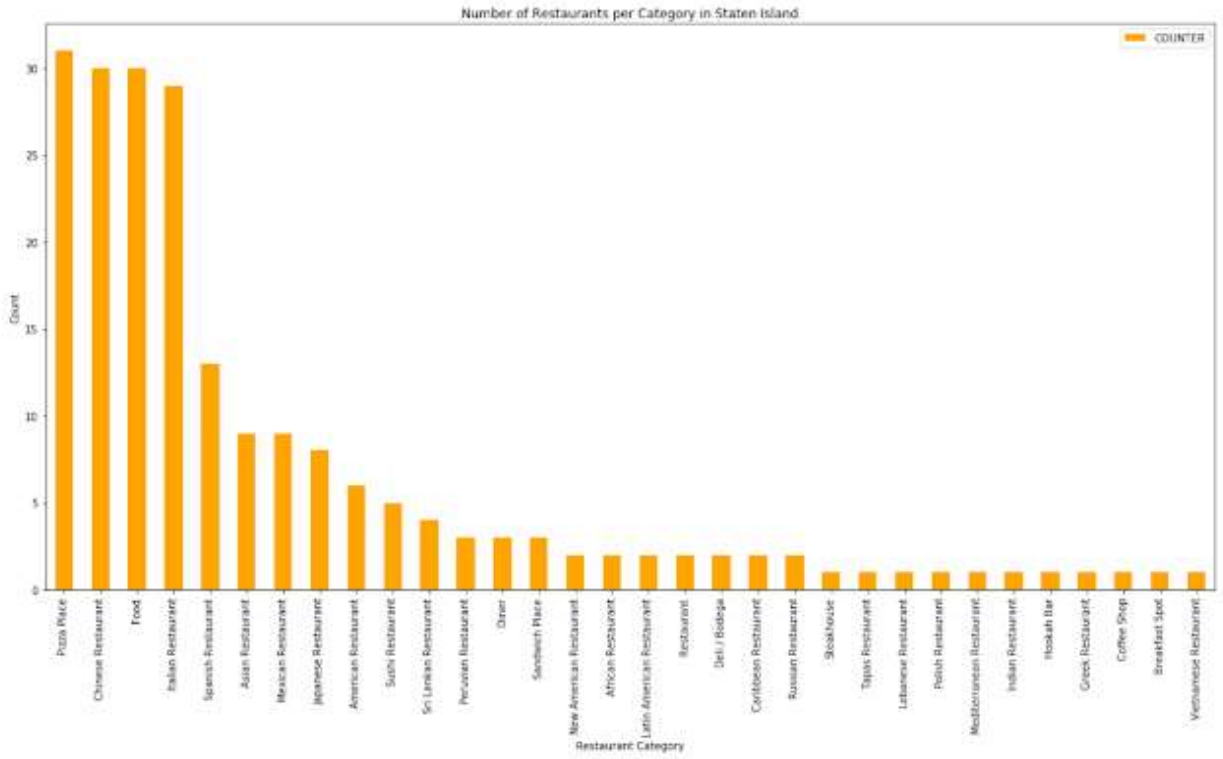
In the above table the columns show the rating ranges and values are the number of restaurants rated in that range for each borough. We have a total of 4984 restaurants out of which only 100 restaurants have been rated. From this it is evident that we do not have enough rating data for these restaurants.

We will now explore the restaurant categories. This information was received from our initial call to the Foursquare API where we had received all the other details for these restaurants. We had also performed cleanup of our categories data as part of our data clean up activity earlier. The count of restaurants post data cleanup is 4787. This is the data that we will be continuing our analysis on.

We will now look at the number of restaurant categories and how they are spread across the neighborhood of New York. The following bar graphs show us the spread of these categories across the boroughs of New York.

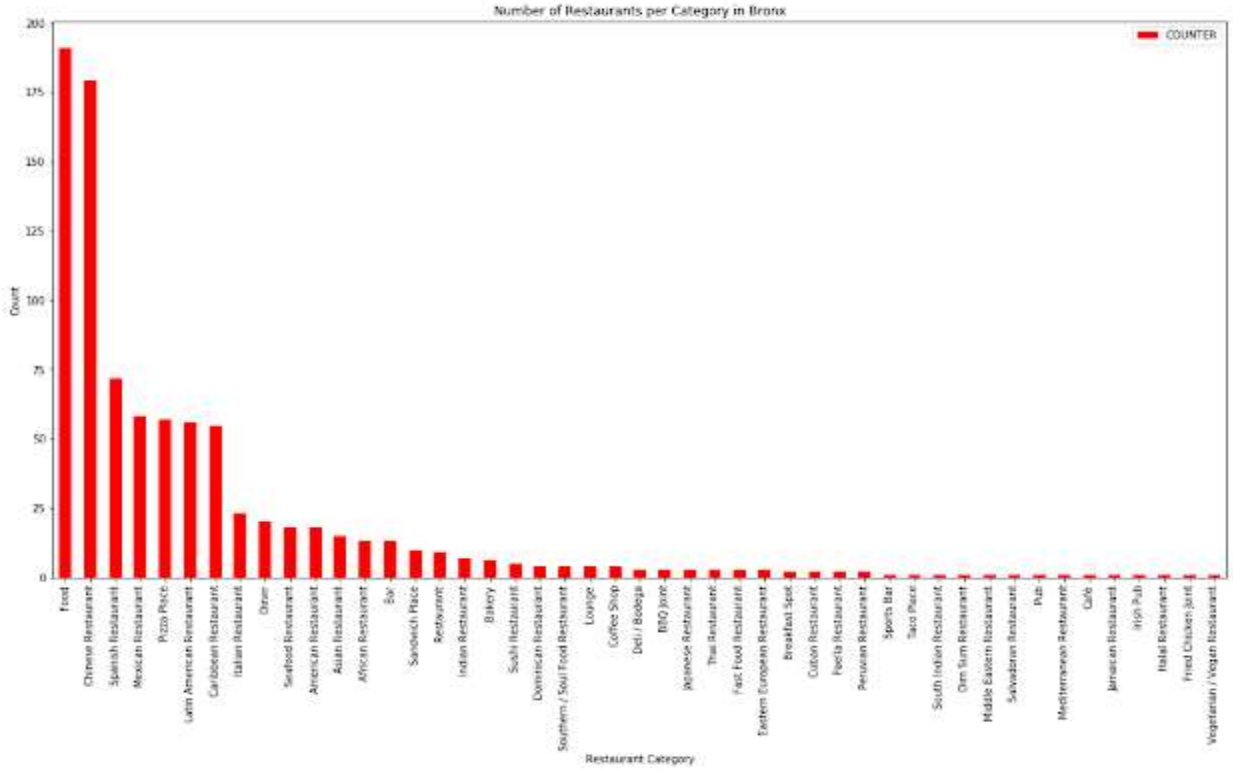
For Staten Island

```
Number of Restaurants in Staten Island is  208
```



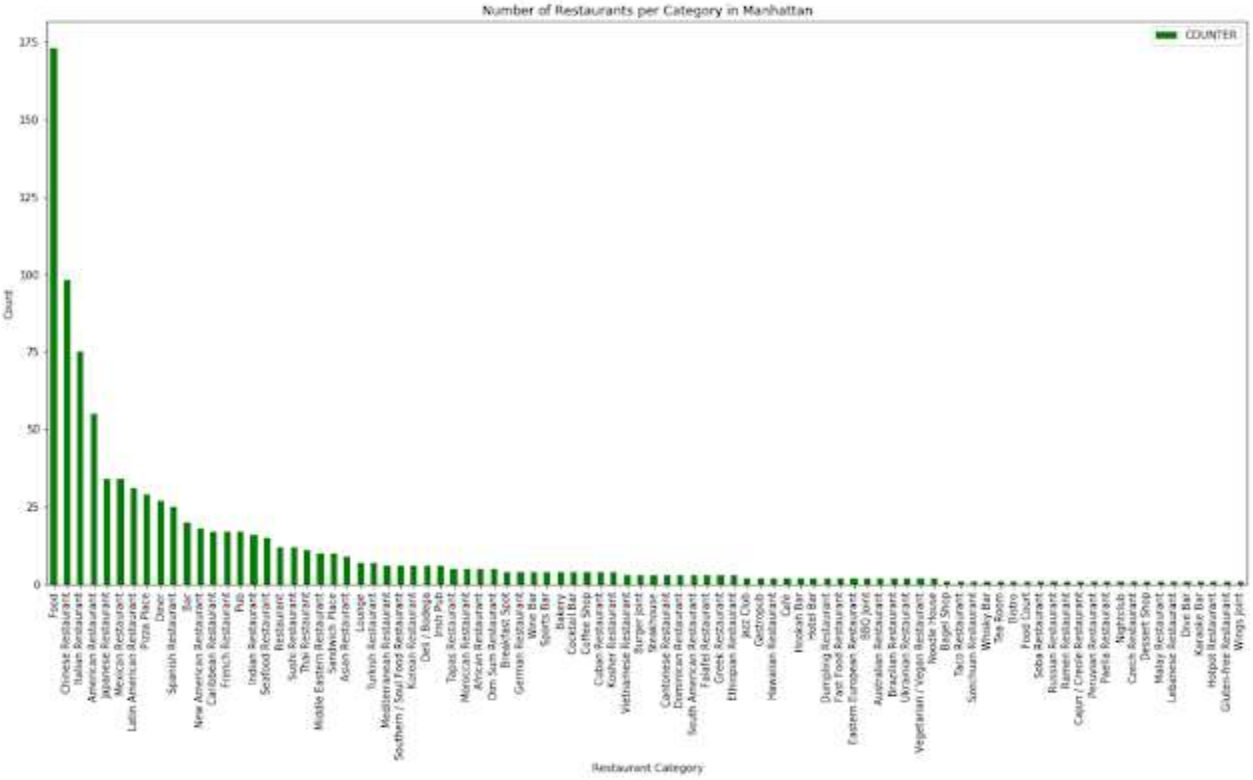
For Bronx

```
Number of Restaurants in Bronx is  881
```



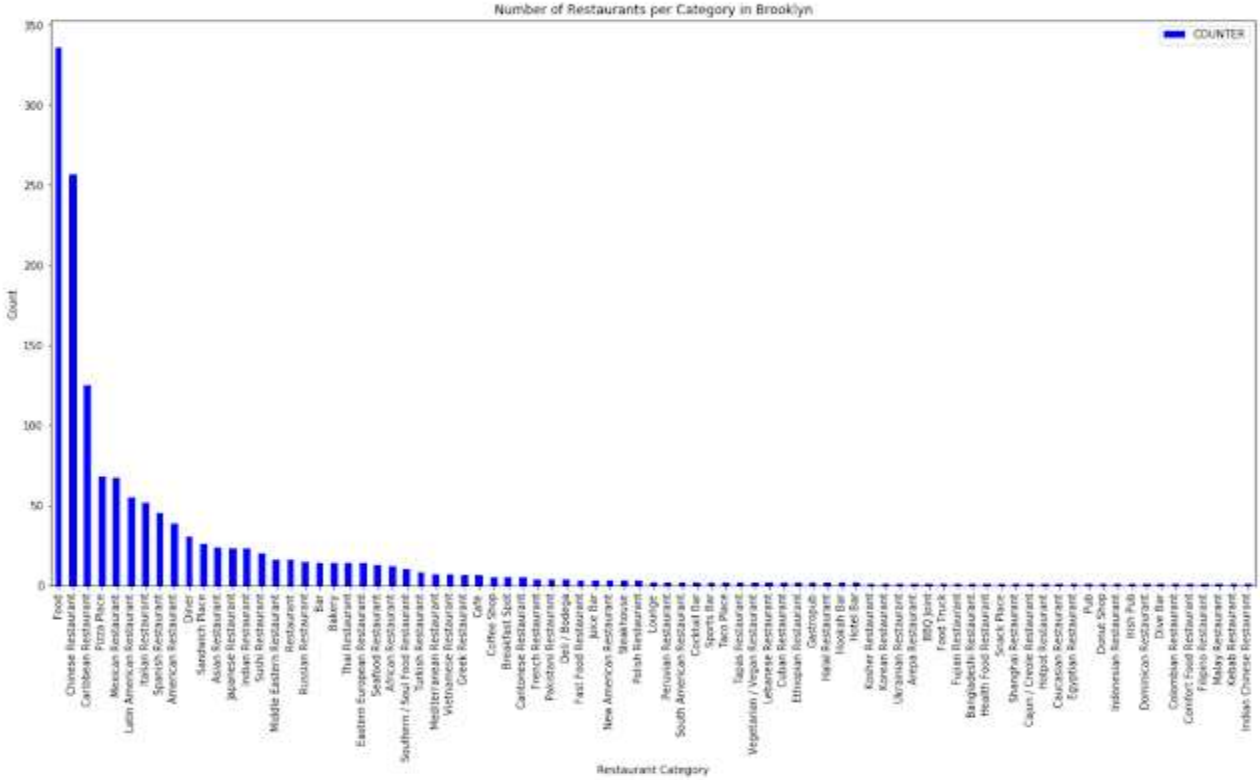
For Manhattan

```
Number of Restaurants in Manhattan is  945
```



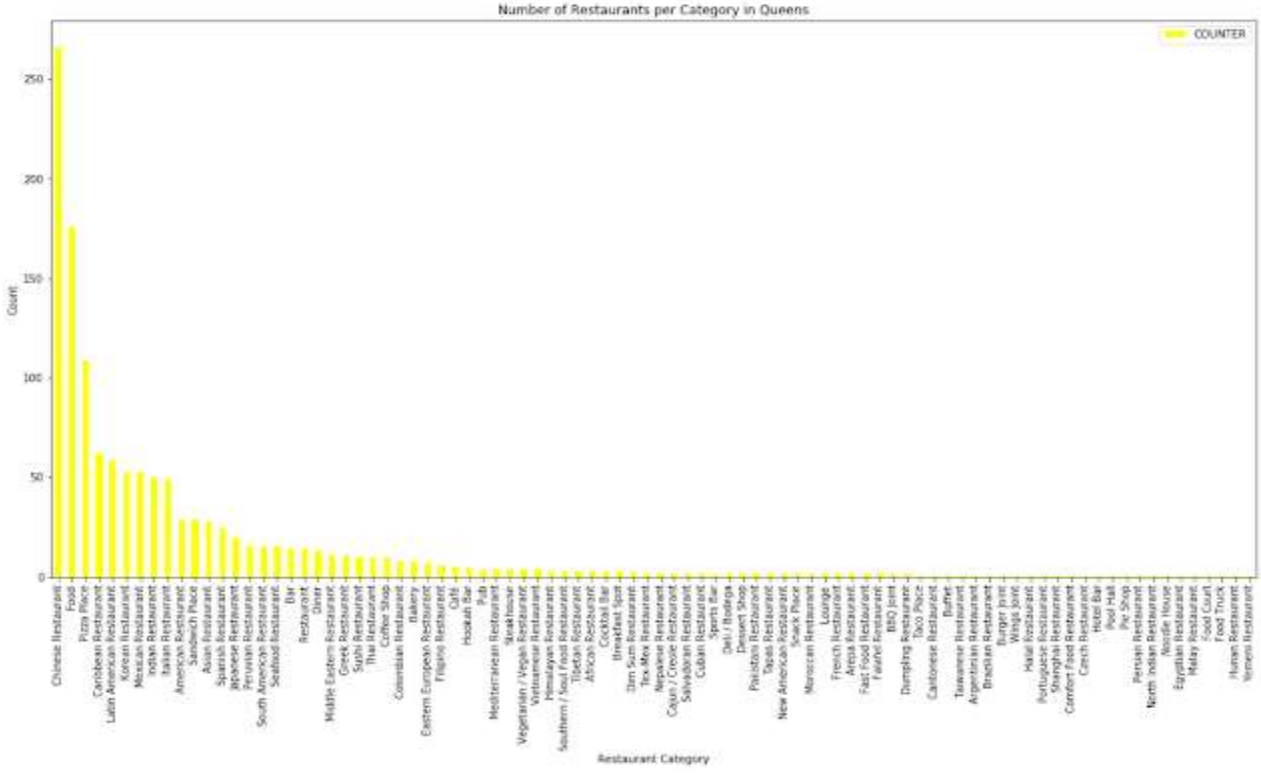
For Brooklyn

```
Number of Restaurants in Brooklyn is  1461
```



For Queens

```
Number of Restaurants in Queens is  1292
```



In summary we have a total of 130 restaurant categories existing in the neighborhoods of New York. To get an idea of the Top 10 restaurant categories in each borough, let us look at the following table.

No.	Queens	Brooklyn	Manhattan	Bronx	Staten Island
1	Chinese Restaurants	Chinese Restaurants	Chinese Restaurants	Chinese Restaurants	Pizza Places
2	Pizza Places	Caribbean Restaurants	Italian Restaurants	Spanish Restaurants	Chinese Restaurants
3	Caribbean Restaurants	Pizza Places	American Restaurants	Mexican Restaurants	Italian Restaurants
4	Latin American Restaurants	Mexican Restaurants	Japanese Restaurants	Pizza Places	Spanish Restaurants
5	Korean Restaurants	Latin American Restaurants	Mexican Restaurants	Latin American Restaurants	Asian Restaurant
6	Mexican Restaurants	Italian Restaurants	Latin American Restaurants	Caribbean Restaurants	Mexican Restaurants
7	Indian Restaurants	Spanish Restaurants	Pizza Places	Italian Restaurants	Japanese Restaurants
8	Italian Restaurants	American Restaurants	Diners	Diners	American Restaurants
9	American Restaurants	Diners	Spanish Restaurants	Seafood Restaurants	Sushi Restaurants
10	Sandwich Places	Sandwich Places	Bars	American Restaurants	Sri Lankan Restaurants

We have enough data pertaining to the restaurant categories from Foursquare. However we are still lacking information regarding user preferences, without which we will not be make meaningful use of this category data for making recommendations.

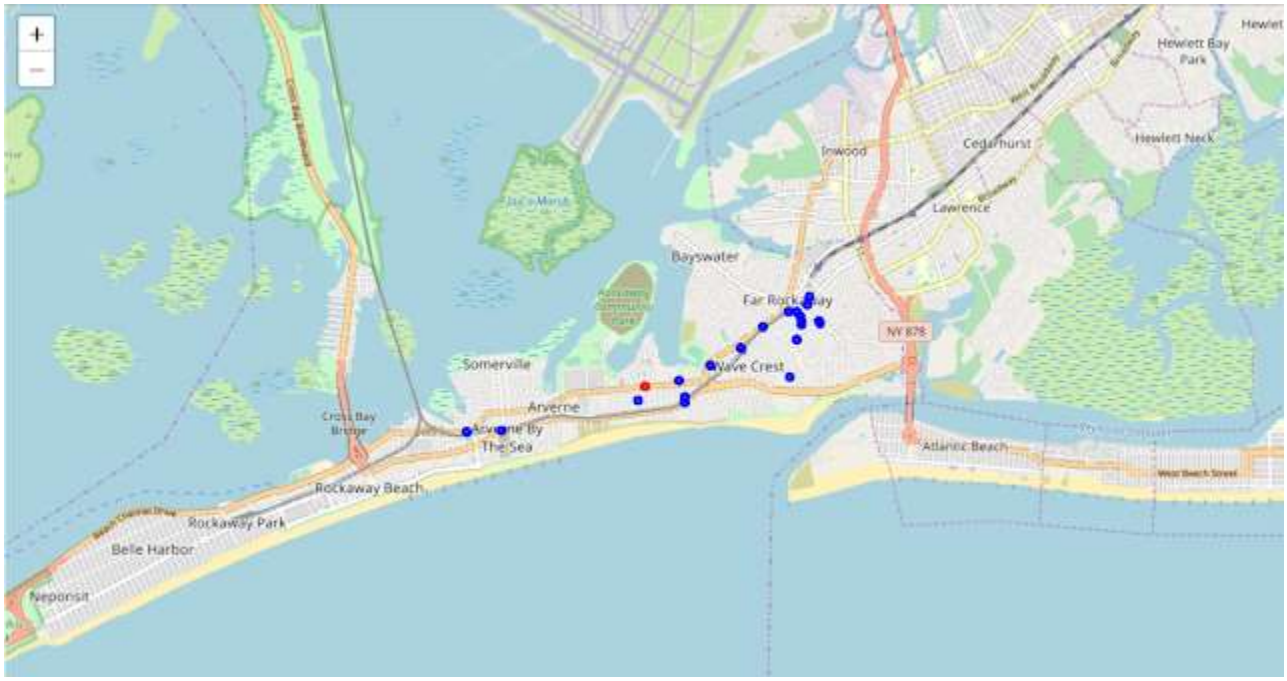
As already stated above, we do not have all the required information for being able to use categories and ratings for providing recommendation. So we will consider location alone for providing restaurant recommendations that are in close proximity to the user's location.

For being able to make recommendations based on location proximity, we will calculate the distance between the user's location and each of the 4787 restaurants that exist in our dataset and then sort the dataset based on the dataset to get the list of restaurants from the closet to the farthest. Below is a list of recommended restaurants that are in close proximity to the following user location.

User Coordinates are: [40.595641807368494, -73.77613282391705]

	Restaurant_Name	Address	Latitude	Longitude	Category	Distance
4019	New Garden Restaurant	['4317 Beach Channel Dr', 'Far Rockaway, NY 11...	40.594204	-73.777074	Asian Restaurant	0.178426
4020	Rockaway Pizza & Restaurant	['3213 Beach Channel Dr', 'Far Rockaway, NY 11...	40.596339	-73.771097	Pizza Place	0.433259
4022	Golden Lee Chinese Restaurant	['3886 Rockaway Beach Blvd', 'Far Rockaway, NY...	40.594535	-73.770300	Chinese Restaurant	0.508779
4023	Xing Xing Chinese Restaurant	['3712 Rockaway Beach Blvd', 'Far Rockaway, NY...	40.593898	-73.770340	Chinese Restaurant	0.527164
4021	Buona Pizza Restaurant	['32-13 Beach Channel Dr', 'Far Rockaway, NY 1...	40.598049	-73.766563	Pizza Place	0.852982
3981	Six Flags Chinese Restaurant	['558 Beach 25th St', 'Far Rockaway, NY 11691'...	40.599981	-73.762078	Chinese Restaurant	1.283513
3977	Beautiful Chinese Restaurant	['2103 Cornaga Ave', 'Far Rockaway, NY 11691',...	40.602320	-73.758810	Chinese Restaurant	1.643125
3986	Cheung Hing Kitchen	['1729 Seagirt Blvd', 'Far Rockaway, NY', 'Uni...	40.596716	-73.754741	Chinese Restaurant	1.814659

This list of recommended restaurants will be different based on the location of the user seeking the recommendation. Let us also visualize the above in the map of New York generated using python's folium library.



On the map, the red circle is the user's location and the blue circles are the locations of the recommended restaurants that are in close proximity to the user.

4. Results and Discussion

Our analysis above shows that there is a good number of restaurants i.e. 4787 restaurants existing in the Neighborhoods of New York. We can also see that there is a good spread of these restaurants across the neighborhoods of New York. The spread of the restaurants across the neighborhoods of New York is as shown below.

Borough	Restaurant Count
Bronx	881
Manhattan	945
Brooklyn	1461
Queens	1292
Staten Island	208

We had started off with the Neighborhood data for New York. This we downloaded from the web. We then used this information (coordinates of the neighborhoods) to find restaurants in these neighborhoods. We used the Foursquare Places API for finding the restaurants that are in these neighborhoods and also gather the details that we needed pertaining to these restaurants i.e. Name, Address, Coordinates, ratings, restaurant categories etc.

We performed detailed analysis of this restaurant data that we had collected by 1) correcting inaccurate data, 2) identifying and correcting missing data, 3) removing unwanted pieces of data. We then further analyzed the restaurants data to see if we had the following pieces of information for being able to make appropriate recommendations about restaurants:

- Location information for all the restaurants
- Ratings information for all the restaurants
- User information regarding user's rating history / preferences in terms of restaurants
- Category information for restaurants

From our analysis we see that we have the following pieces of information.

What we know	What we do not know
1) Location of Restaurants	1) User Ratings / Preferences
2) Category of Restaurants	2) Restaurant's Ratings

Based on the above information and our analysis, we have enough information to make a restaurant recommendation based on location proximity using: 1) location of the user (which we have assumed for now and will be provided by the user in the actual system) 2) location of the Restaurants in the neighborhoods of New York (we got this information from Foursquare) We used the above location data to calculate the distance of all the restaurants from the users location and then sorted the restaurants based on distance so that we have the list of restaurants from the closest one to the farthest one.

Although we also have information regarding the Categories of the restaurants, we are not able to use this data to make a recommendation in the absence of information regarding User Preferences and User past ratings for various cuisines.

Finally we have built a recommendation system that would recommend the top 20 restaurants that are closest to the location provided by the user. We have validated the results based on the distance of each restaurant from the user provided location. We have also visualized the user location along with the recommended restaurants' locations on the map of New York to validate our recommendation.

There is ample scope for future studies to improve this recommendation engine and expanding it to include other aspects that we have identified above for making a more complete recommendation. This can also be extended further to provide restaurant recommendations for locations other than New York and also for providing recommendations on other venues.

5. Conclusion

Purpose of this project is to model a recommendation system that utilizes location data for recommending restaurants to users in the Neighborhood of New York. There is a whole world of people traveling, who are looking out for recommendations to meet their needs. The goal here is to provide better outcomes to such people that are looking for recommendations.