

Patterns in Types Of Restaurants in New York

VARSHA GADEKAR

NOV 11, 2020

1. INTRODUCTION

1.1 Background

New York City (NYC), often called simply New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km²), it is one of the world's most populous megacities. Situated on one of the world's largest natural harbors, New York City is composed of five boroughs, each of which is a county of the State of New York.

The restaurant industry is essential to New York City's social and economic fabric. From small, family-owned restaurants and food carts to four-star world-famous establishments, the City abounds with restaurants that offer cuisines from every corner of the globe. Restaurants lend vitality to each neighborhood and are integral to Manhattan's central business districts and the City's vibrant tourism industry, attracting millions of visitors each year who spend a total of \$46 billion annually.

1.2 Problem

The project aims to find out if there is any pattern in the type of restaurants located in every borough and neighborhood of New York.

- Are some cuisines / types of restaurants more represented in certain areas than in others?
- Which cuisines / types of restaurants have delivery options?
- Is it possible to group neighborhoods together based on the sole criteria of food?
- Which borough has the most restaurants?

The project will also focus on the borough with most restaurants in New York and provide a list of top restaurants based on user ratings and likes and dine-out and delivery services considering current covid scenarios.

1.3 Interest

Imagine you plan to move to New York or travel for vacation or for work or you are just passing by, your top priorities will always include food. As we all know eating out and sipping on a cup of coffee is a major aspect of people's lives. So, wouldn't it be great to know beforehand the boroughs and neighborhoods of New York based on what we can eat there? What are the major cuisines or types of food famous there? or Which area to visit as per our taste of eating.

This project helps every person travelling or moving to New York by providing details of restaurants in the neighborhoods or borough they would be travelling. And with current covid situations we might need to know restaurants who have dine out options or which provide delivery service. Also it will

help if someone only wants to eat a specific type of food, for example a person only likes to eat korean food so this will help in locating korean restaurants in the neighborhoods the person is. If the neighborhood does not have any korean restaurants then you can get a list of nearby neighborhoods which have it.

2. Data

2.1 Data Sources

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572. Geopy library was used to get the latitude and longitude values of New York City. Next, Foursquare API was utilized to explore all the neighborhoods and segment them. The API was also used to get details of venue

2.2 Data Cleaning

Data downloaded was cleaned and saved in a dataframe. The file contained json data so it was converted into a dataframe. Only required features from data were saved in the data frame. Then I checked if the data has all 5 boroughs and 306 neighborhoods. The data was further divided into two different data frames consisting data of boroughs for further analysis.

2.3 Feature Selection

Each record in json file contain following features:

```
{'type': 'Feature',  
  'id': 'nyu_2451_34572.1',  
  'geometry': {'type': 'Point',  
    'coordinates': [-73.84720052054902, 40.89470517661]},  
  'geometry_name': 'geom',  
  'properties': {'name': 'Wakefield',  
    'stacked': 1,  
    'annoline1': 'Wakefield',  
    'annoline2': None,  
    'annoline3': None,  
    'annoangle': 0.0,  
    'borough': 'Bronx',  
    'bbox': [-73.84720052054902,  
      40.89470517661,  
      -73.84720052054902,  
      40.89470517661]}}
```

For the project only the following features were selected :

- *borough* -> Borough
- *Name* -> Neighborhood
- *Coordinates* -> Latitude, Longitude

3. Methodology

In this project we will direct our efforts on detecting patterns in the type of restaurants located in every borough and neighborhood of New York. We will limit our analysis to area 500m radius.

In the first step we have collected the required data: location and type (category) of every restaurant within 500m radius.

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of New York. We will also explore the borough with most restaurants in New York and provide a list of top restaurants based on user ratings and likes and dine-out and delivery services considering current covid scenarios.

In the third and final step we will focus clustering neighborhoods and boroughs based on types of restaurants. We will present a map of all such locations but also create clusters (using k-means clustering) to derive conclusions discussed in the problem section.

4. Exploratory Data Analysis

In the data gathering phase all the restaurant's data has been collected. We also know which restaurants exactly are in every neighborhood and borough. We're now ready to use this data for analysis to produce the report on the density and types of restaurants in New York and detailed venues based on user ratings and likes.

4.1 Density of restaurants in New York

We calculated the density of restaurants for each borough as well as for each neighborhood.

BOROUGH:

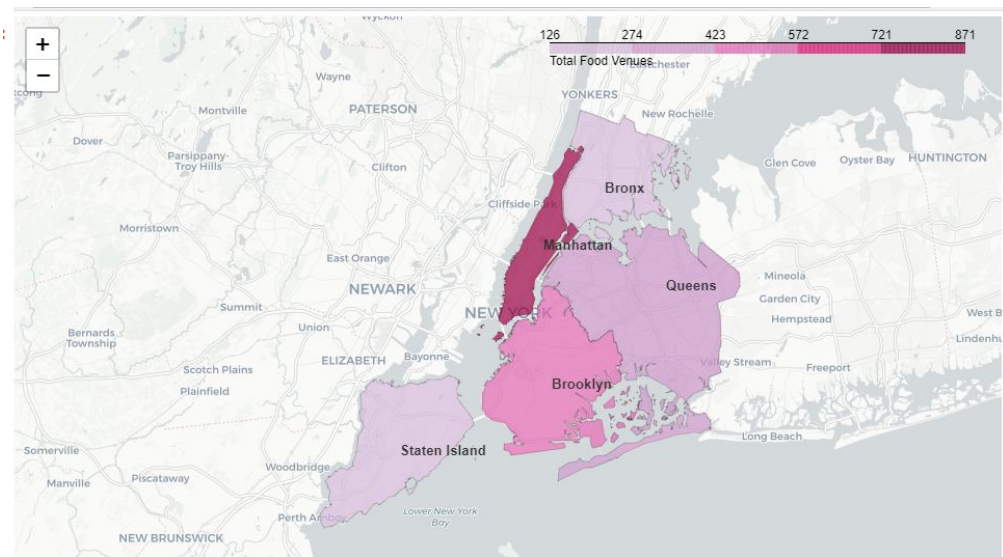
Create a dataframe with total venue count for each borough and then call the function for the geopy library to get latitude and longitude for each borough(that will help in plotting the data). And the final data looks like:

| | Borough | Count | Longitude | Latitude |
|---|---------------|-------|------------|-----------|
| 0 | Bronx | 165 | -73.878594 | 40.846651 |
| 1 | Brooklyn | 488 | -73.949583 | 40.650102 |
| 2 | Manhattan | 870 | -73.959894 | 40.789624 |
| 3 | Queens | 328 | -73.797634 | 40.749824 |
| 4 | Staten Island | 126 | -74.149605 | 40.583456 |

Choropleth map is used to visualize the above data :

A choropleth map is a type of thematic map in which a set of predefined areas is colored or patterned in proportion to a statistical variable that represents an aggregate summary of a geographic characteristic within each area, such as population density or per-capita income. It allows us to study

how a variable evolves along a territory. It is a powerful and widely used data visualization technique.



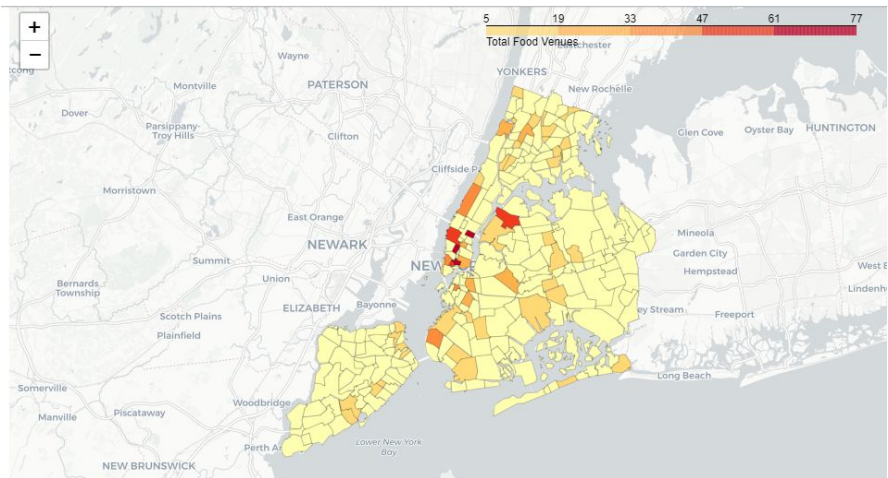
From the above map we can conclude that Manhattan has the highest number of restaurants.

NEIGHBORHOOD:

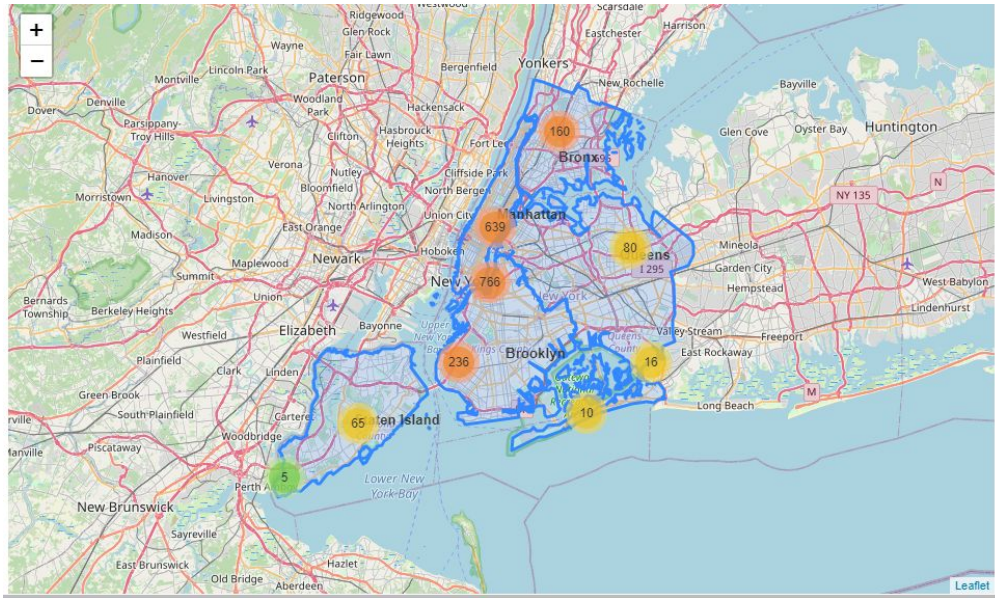
Create a dataframe with total venue count for each neighborhood in new york. I discarded neighborhoods with less than 5 restaurants. And the final data looks like:

| | Neighborhood | Count |
|---|-------------------|-------|
| 0 | Allerton | 5 |
| 1 | Astoria | 49 |
| 2 | Astoria Heights | 5 |
| 3 | Battery Park City | 8 |
| 4 | Bay Ridge | 42 |

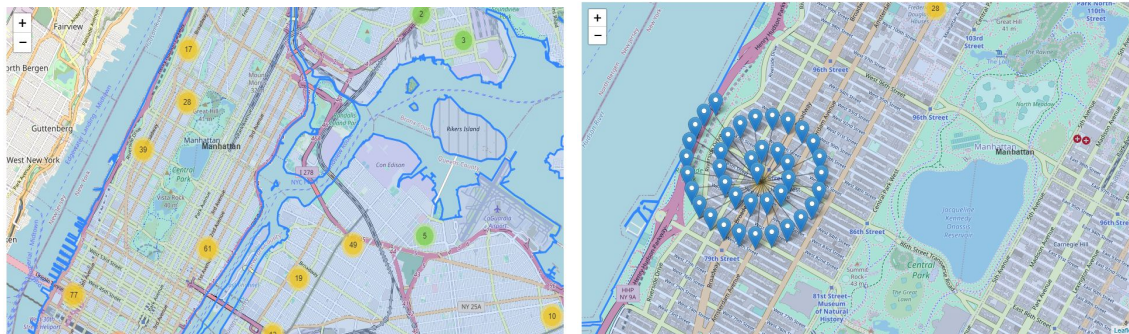
Folium map is used to visualize the above data:



To visualize restaurants in numerical form we used folium map for the same:



As we zoom in the above map the count will be distributed among neighborhoods until the exact location is reached.

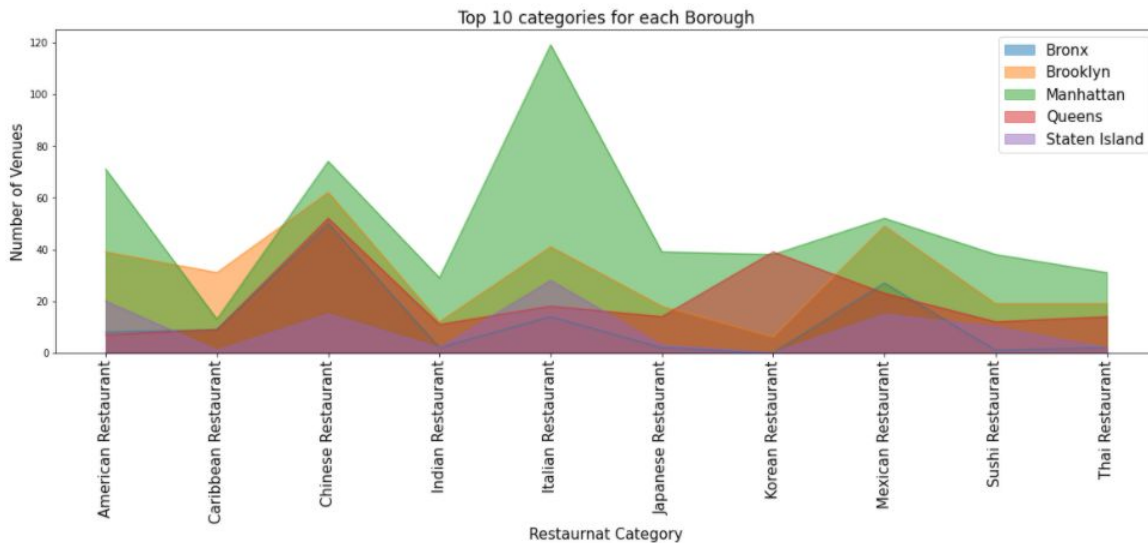


4.2 Top 10 categories of restaurants based on density

I found out that there were 83 unique categories of New York from the data fetched using Foursquare API. Then I performed some calculation to get list of top 10 categories of restaurants in New York and the data looked like:

| | Venue Category | Venue |
|---|----------------------|-------|
| 0 | Chinese Restaurant | 253 |
| 1 | Italian Restaurant | 220 |
| 2 | Mexican Restaurant | 166 |
| 3 | American Restaurant | 145 |
| 4 | Korean Restaurant | 83 |
| 5 | Sushi Restaurant | 80 |
| 6 | Japanese Restaurant | 76 |
| 7 | Thai Restaurant | 68 |
| 8 | Caribbean Restaurant | 63 |
| 9 | Indian Restaurant | 56 |

The above data was used to find top 10 categories for each of the 5 boroughs and the data was plotted on an area chart to find if there is any pattern in types of restaurant in every borough.



From the above map following conclusions were made:

- In Manhattan most popular are Italian restaurant
- ALL the Boroughs has almost equal popularity for Chinese Restaurants
- Brooklyn, Manhattan, Queens have almost similar popularity in types of restaurants

4.3 Restaurant Density based on delivery option

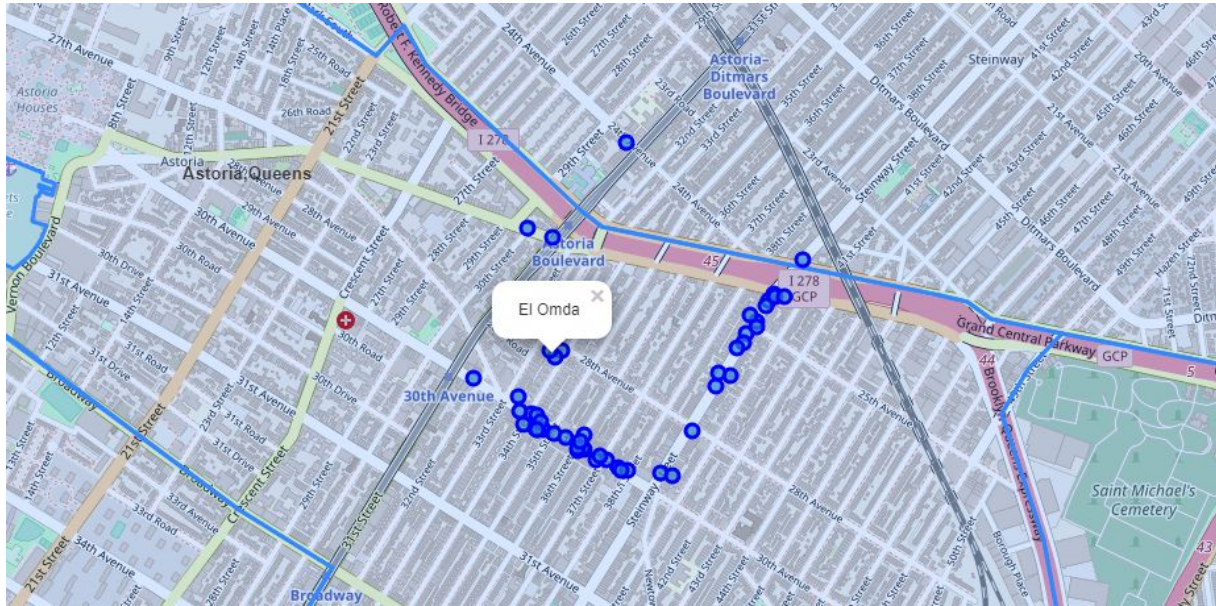
From the data gathered it was calculated that most of the restaurants provide home delivery options in New York.

Then a function was created which lists and display all the restaurants in a neighborhood which provide home delivery option you just need to pass the name of neighborhood and borough.

```
neighborhood_name='Astoria'
borough_name='Queens'
ny_del_map=get_res_list(neighborhood_name,borough_name)
ny_del_map
```

Output:

```
Restaurant list:
-> Favela Grill -> Off The Hook -> Al-sham Sweets and Pastries -> The Gully -> AbuQir Seafood
-> Duzan -> Sabry's -> Seva Indian Cuisine -> Viva Viva -> Crave Astoria -> Mochiron Izak
aya -> El Cafetal -> Tikka Indian Grill -> Ovelia Psistaria Bar -> Pochana Thai Kitchen -> K
abab Café -> Jujube Tree -> The Grand -> Antika Restaurant & Pizzeria -> King of Falafel and Sh
awarma -> Kal -> Sugar Freak -> Farid Kebab, Steinway -> Gyro Uno -> Mombar -> Gyro World
-> El Omda -> Queens Comfort -> Blend Astoria -> Elias Corner For Fish -> Halal Sandwich Sho
p -> Layali Dubai -> Via Trenta -> Pink Nori -> El-Rawsheh -> Golden House Chinese Restaura
nt -> Sampan -> Aden Restaurant -> Matsu Sushi -> Piatso Souvlaki -> Pita Pan -> Namaste -
-> Golden Kitchen -> Little Morocco -> Mijana Hookah Lounge & Restaurant -> Avenue A -> Aweso
meSauce -> Tikka Roll House -> Hangar Astoria ->
```



4.4 Clustering Neighborhoods to find any similarity

Modeling(k-means clustering):The k-means clustering method is an unsupervised machine learning technique used to identify clusters of data objects in a dataset. There are many different types of clustering methods, but k-means is one of the oldest and most approachable.

Feature Engineering:

Now, each neighborhood is analyzed individually to understand the most common cuisine being served within its 500 meters of the vicinity.

The above process is taken forward by using the 'one hot encoding' function of python 'pandas' library. One hot encoding converts the categorical variables (which are 'Venue Category') into a form that could be provided to ML algorithms to do a better job in prediction.

| Neighborhood | Afghan Restaurant | African Restaurant | American Restaurant | Arepa Restaurant | Argentinian Restaurant | Asian Restaurant | Australian Restaurant | Austrian Restaurant | Brazilian Restaurant | Cajun / Creole Restaurant |
|---------------------|-------------------|--------------------|---------------------|------------------|------------------------|------------------|-----------------------|---------------------|----------------------|---------------------------|
| 0 Allerton | 0.0 | 0.0 | 0.20 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.000000 |
| 1 Annadale | 0.0 | 0.0 | 0.75 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.000000 |
| 2 Astoria | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.020408 | 0.020408 |
| 3 Astoria Heights | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.000000 |
| 4 Battery Park City | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.000000 |

Next some operations were done to find out top 5 common venues for each neighborhood:

| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---------------------|---------------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| 0 Allerton | Chinese Restaurant | American Restaurant | Mexican Restaurant | Spanish Restaurant | Vietnamese Restaurant |
| 1 Annadale | American Restaurant | Sushi Restaurant | Vietnamese Restaurant | Ethiopian Restaurant | Falafel Restaurant |
| 2 Astoria | Middle Eastern Restaurant | Greek Restaurant | Indian Restaurant | Seafood Restaurant | Mediterranean Restaurant |
| 3 Astoria Heights | Italian Restaurant | Chinese Restaurant | Vietnamese Restaurant | Halal Restaurant | Ethiopian Restaurant |
| 4 Battery Park City | Chinese Restaurant | Steakhouse | Japanese Restaurant | Italian Restaurant | Seafood Restaurant |

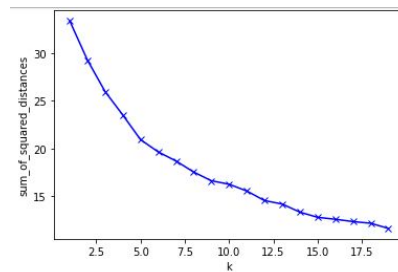
Next we perform clustering using k-means' which creates clusters of data points aggregated together because of certain similarities. This algorithm will be used to count neighborhoods for each cluster label for variable cluster size.

To implement this algorithm, it is very important to determine the optimal number of clusters (i.e. k). We will use The Elbow Method to know the same.

The Elbow Method

The Elbow Method calculates the sum of squared distances of samples to their closest cluster center for different values of 'k'. The optimal number of clusters is the value after which there is no significant decrease in the sum of squared distances.

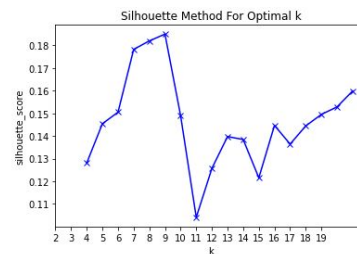
Following is the result from elbow method:



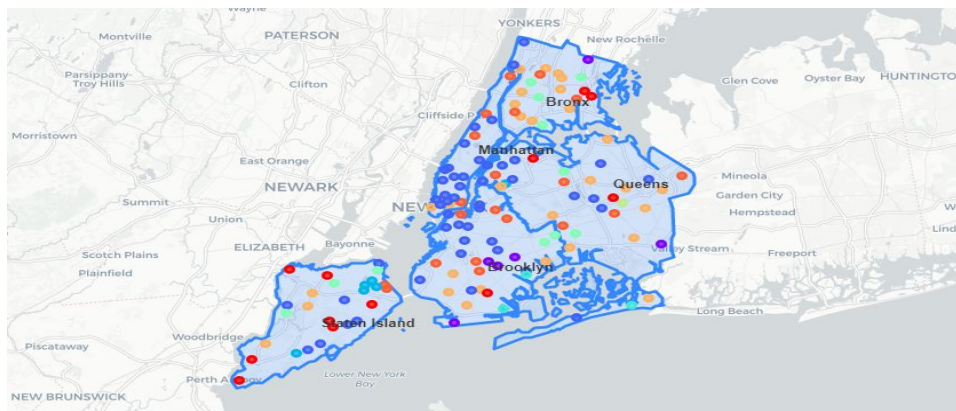
The Elbow method does not seem to help us to determine the optimal number of clusters. Let's use another method.

The Silhouette Method - The silhouette value measures how similar a point is to its own cluster (cohesion) compared to other clusters (separation).

Following is the result from silhouette method:



There is a peak at k=9 so we will cluster our dataset in 9 clusters and below is the output after the clusters were form:



Then some operations were performed on the cluster data to get the result in list form of first and second most common venues.

| Cluster Labels | 1st Most Common Venue | Count | Neighborhood | Borough |
|----------------|-----------------------|----------------------|--|---|
| 0 | 0 | Italian Restaurant | 10 Country Club,Pelham Bay,Tottenville,Old Town,C... | [Bronx, Staten Island, Queens, Brooklyn] |
| 1 | 1 | Caribbean Restaurant | 4 Eastchester,Laurelton,Remsen Village,Erasmus | [Bronx, Queens, Brooklyn] |
| 2 | 2 | Italian Restaurant | 15 North Riverdale,Bay Ridge,Cobble Hill,Fort Gre... | [Bronx, Brooklyn, Manhattan, Staten Island, Qu... |
| 3 | 3 | American Restaurant | 5 Sunnyside,Sunnyside,Grymes Hill,Annadale,Silve... | [Queens, Staten Island] |
| 4 | 4 | Seafood Restaurant | 2 Gerritsen Beach,Edgemere | [Brooklyn, Queens] |
| 5 | 5 | Spanish Restaurant | 4 Fordham,East Tremont,Travis,Westerleigh | [Bronx, Staten Island] |
| 6 | 6 | Indian Restaurant | 1 Jamaica Estates | [Queens] |
| 7 | 7 | Chinese Restaurant | 22 Marble Hill,Norwood,Melrose,Olinville,Concours... | [Manhattan, Bronx, Brooklyn, Queens, Staten Is... |
| 8 | 8 | Mexican Restaurant | 13 Bedford Park,Westchester Square,Bushwick,East ... | [Bronx, Brooklyn, Manhattan, Queens, Staten Is... |

Above data represents data of each cluster with the first most common venue and the corresponding neighborhoods and boroughs that contain those categories of venue. Like in CLuster 0 the most common category is Italian Restaurant and the neighborhoods that have restaurants in the same category are Country Club, Pelham Bay, Tottenville etc.

| Cluster Labels | 2nd Most Common Venue | Count | Neighborhood | Borough |
|----------------|-----------------------|-----------------------|---|--------------------------------------|
| 0 | 0 | Chinese Restaurant | 3 Country Club,Pelham Bay,Astoria Heights | [Bronx, Queens] |
| 1 | 1 | Caribbean Restaurant | 2 East Flatbush,Coney Island | [Brooklyn] |
| 2 | 2 | American Restaurant | 8 Fort Greene,Downtown,North Side,Clinton,Chelse... | [Brooklyn, Manhattan, Staten Island] |
| 3 | 3 | Mexican Restaurant | 2 Sunnyside,Sunnyside | [Queens, Staten Island] |
| 4 | 4 | Asian Restaurant | 1 Edgemere | [Queens] |
| 5 | 5 | Spanish Restaurant | 3 East New York,Ozone Park,North Corona | [Brooklyn, Queens] |
| 6 | 6 | Vietnamese Restaurant | 1 Jamaica Estates | [Queens] |
| 7 | 7 | Chinese Restaurant | 6 Longwood,Hollis,Rochdale,Lindenwood,Rossville,... | [Bronx, Queens, Staten Island] |
| 8 | 8 | Mexican Restaurant | 6 Sunset Park,Flatbush,Prospect Park South,Lower... | [Brooklyn, Manhattan, Queens] |

Above data represents data of each cluster with the second most common venue and the corresponding neighborhoods and boroughs that contain those categories of venue. Like in CLuster 0 the most common category is Chinese Restaurant and the neighborhoods that have restaurants in the same category are Country club, Pelham Bay, Astoria heights.

5. Results and Discussions

Our analysis shows that there are a great number of restaurants in New York but Manhattan has the largest number of restaurants among all boroughs. And the people of Manhattan love Italian food. We also found out that there are 83 types of food category/cuisines in New York among which the top 10 based on density are : Chinese, Italian, Mexican, American, Korean, Sushi, Japanese, Thai, Caribbean and Indian.

Most of the restaurants provide food delivery options so if you want to grab a bite from the comfort of your own place then you can just ask for home delivery.

Then we clustered the neighborhood data to find out if there is any relation between the neighborhood based on types of food and we found out neighborhoods who have similar food tastes.

6. Conclusion

Purpose of this project was to find if there is any pattern in types of restaurants located in every area of New York. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis. We found out that there were some similarity in most common venues in some neighborhoods. Also it can be concluded from visual analysis that Manhattan's most popular popular food category is Italian. ALL the Boroughs have almost equal popularity for Chinese Restaurants. And Brooklyn, Manhattan, Queens have almost similar popularity in types of restaurants. Most of the restaurants in New York provide home delivery options too. So in current scenarios that is a best option. You can get a list of all the restaurants just by feeding the neighborhood name.

Finally, from all the analysis it is clear that New York has a great diversity of food categories in almost every area. It is easy to locate a restaurant according to your taste in New York.