COURSERA CAPSTONE

IBM APPLIED DATA SCIENCE CAPSTONE

INSIGHGTS INTO NEW YORK CITY'S RESTAURANT BUSINESS

March 2020

Hanieh Daliri

Introduction/ Business Understanding

Restaurants in New York City emerge and disappear very quickly. For every 460 people in New York City, there is only one restaurant. With more than eight million population and around nineteen hundreds of restaurants in this city, the need for opening new restaurants and eateries is increasing more than ever. However, with this blooming food service market, one of the essential factors in the long-lasting success of this businesses is the location of restaurants. This project aims to provide valuable insights into the New York City foodservice market, such as the most common type of restaurants in the city, the distribution of restaurant types by neighborhoods and finding the neighborhoods in the need of new restaurants.

Interest in the project

The audience of this project is potential restaurant owners and investors who are looking to find the best neighborhood to start a specific type of restaurant or to identify the type of restaurant that fits the neighborhood of interest.

Data Acquisition

The data required for this project include the list of:

1. boroughs and neighborhoods of the city,
2. types of restaurants in each location, and
3. the geographical coordinates of the neighborhoods to visualize the data on a map.

The Neighborhood, Borough, and Geographical coordinates data for New York City are scraped from the following page: https://geo.nyu.edu/catalog/nyu_2451_34572

Restaurant data, including the type of restaurants and their geographical coordination, are extracted from Four Square's Places API. Foursquare data is classified into various categories and sub-categories. Categories are identified by the tag "Category ID". The category of interest here is Food and the category ID for food is 4d4b7105d754a06374d81259. The sub-categories of the food category are various types of restaurants located in the venue.

It should be noted that for illustration purposes, the data for segmentation and clustering is simplified the to the neighborhoods only in Brooklyn. The figure below shows the Brooklyn map with the neighborhoods.

Data Cleaning

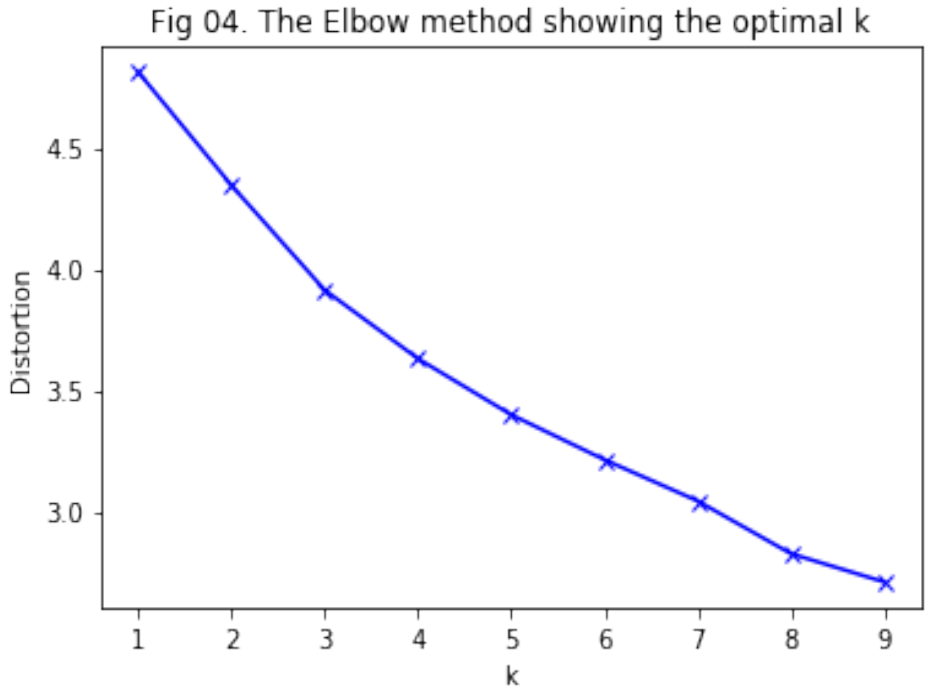
After downloading the data from the link mentioned above, the next step is transforming this data into a Pandas data frame. An empty data frame with four columns including, Borough, Neighborhood, Latitude, and Longitude is created. I Used for-loop, to fill the data into the data frame one row at a time. There are 5 borough and 306 neighborhoods. The Geolocator library is used to find the latitude and longitude of New York City. Then the Folium library is used to visualize the map of New York City along with the neighborhood's marker. However, for illustration purposes, the above map is simplified to only show Brooklyn's neighborhoods. Therefore, the original data frame is sliced to and a new data frame of the Brooklyn data is created. Next, the Foursquare API is used to explore the neighborhoods and segment them. The top 100 restaurant types that are in Brooklyn within a radius of 500 meters of each neighborhood are identified. There is a total of **2195** restaurants in Brooklyn with a **95** unique restaurant type. Then used the grouped by function to group the data by neighborhoods and to show the number of restaurants in each neighborhood.

For Analyzing the neighborhoods, the Restaurant type columns are first one-hot encoded and grouped by neighborhood. Next, the rows are grouped by neighborhood and by taking the mean of the frequency of occurrence of each category. in the next step, new is created to display the top 10 restaurant types for each neighborhood. Now the clustering process starts. The optimal value of the number of clusters, k, is determined using the elbow method to be 3. Then, the k-means method is used to cluster the neighborhood into 3 segments.

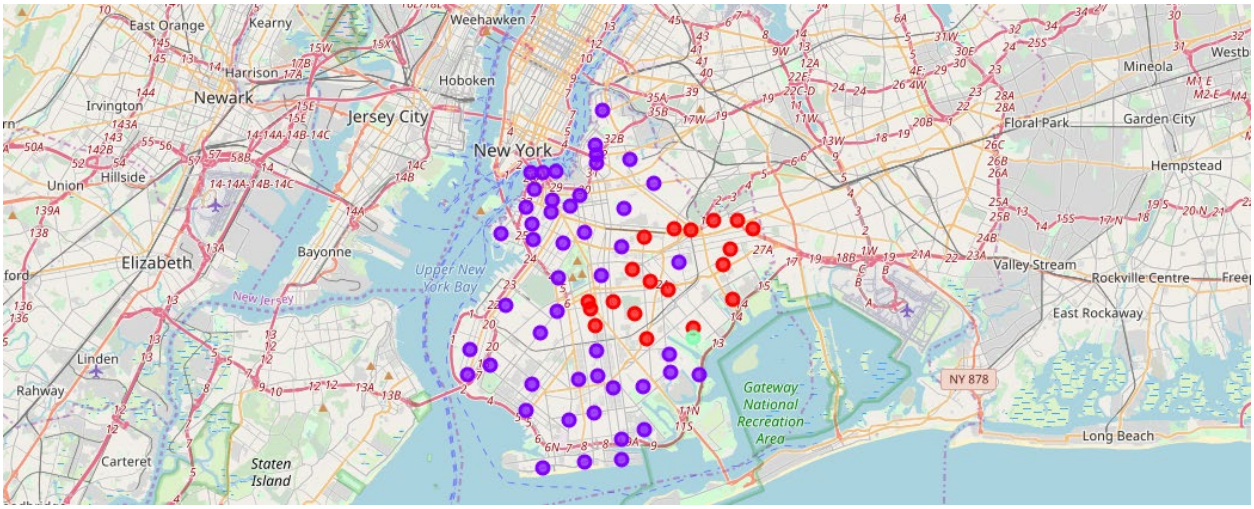
Methodology

K means, an unsupervised clustering algorithm which is commonly used for market segmentation, is used to divide the city's restaurant market into clusters based on the types of restaurants established in the city. Cluster analysis is the process of segregating data into groups

(clusters) in such a way that objects in the same cluster are more like each other than those in other clusters. For this project, the top 10 Restaurant types within a radius of 500 meters of each neighborhood have been identified. As shown in the figure below, the elbow method is used to determine the optimal value of the number of clusters. Then, the neighborhoods are grouped into clusters accordingly.



Finally, the resulting clusters is visualized as below:



Cluster Results

In this section, I can examine each cluster and determine the discriminating restaurant type that distinguishes each cluster. Based on the most common type of restaurant in each cluster, we can identify the neighborhoods that best fit a particular type of restaurant.

The figures Below show the first five rows of each clusters:

Cluster 1-Deli/Bodega

	Neighborhood	1st Most Common Restourant type	2nd Most Common Restourant type	3rd Most Common Restourant type	4th Most Restourant type	5th Most Restourant type	6th Most Restourant type	7th Most Restourant type	8th Most Restourant type	9th Most Restourant type	10th Most Restourant type
8	Flatbush	Deli / Bodega	Chinese Restaurant	Caribbean Restaurant	Bakery	Mexican Restaurant	Pizza Place	Diner	Sandwich Place	Fast Food Restaurant	Middle Eastern Restaurant
10	East Flatbush	Deli / Bodega	Fast Food Restaurant	Chinese Restaurant	Caribbean Restaurant	Food	German Restaurant	Greek Restaurant	Dumpling Restaurant	Eastern European Restaurant	Hawaiian Restaurant
25	Cypress Hills	Fried Chicken Joint	Pizza Place	Latin American Restaurant	Spanish Restaurant	Fast Food Restaurant	Donut Shop	Chinese Restaurant	Food	South American Restaurant	Seafood Restaurant
26	East New York	Deli / Bodega	Spanish Restaurant	Chinese Restaurant	Fast Food Restaurant	Fried Chicken Joint	Caribbean Restaurant	Latin American Restaurant	Pizza Place	Asian Restaurant	Dumpling Restaurant
27	Starrett City	American Restaurant	Caribbean Restaurant	Pizza Place	Chinese Restaurant	Donut Shop	German Restaurant	Gastropub	Dumpling Restaurant	Hawaiian Restaurant	Eastern European Restaurant

Cluster 2-Pizza Places-Italian and Mexican Restaurants

	Neighborhood	1st Most Common Restourant type	2nd Most Common Restourant type	3rd Most Common Restourant type	4th Most Restourant type	5th Most Restourant type	6th Most Restourant type	7th Most Restourant type	8th Most Restourant type	9th Most Restourant type	10th Most Restourant type
0	Bay Ridge	Pizza Place	Italian Restaurant	Chinese Restaurant	Bagel Shop	Greek Restaurant	American Restaurant	Bakery	Thai Restaurant	Mediterranean Restaurant	Fast Food Restaurant
1	Bensonhurst	Chinese Restaurant	Pizza Place	Sushi Restaurant	Italian Restaurant	Donut Shop	Bakery	Russian Restaurant	Noodle House	Café	Halal Restaurant
2	Sunset Park	Bakery	Deli / Bodega	Mexican Restaurant	Pizza Place	Latin American Restaurant	Fried Chicken Joint	Café	Donut Shop	Chinese Restaurant	Bagel Shop
3	Greenpoint	Deli / Bodega	Pizza Place	Mexican Restaurant	French Restaurant	Restaurant	Café	Chinese Restaurant	Sushi Restaurant	Sandwich Place	New American Restaurant
4	Gravesend	Italian Restaurant	Chinese Restaurant	Pizza Place	Deli / Bodega	Bakery	Donut Shop	Diner	Eastern European Restaurant	Breakfast Spot	Food

Cluster 3-Asian restaurants

	Neighborhood	1st Most Common Restaurant type	2nd Most Common Restaurant type	3rd Most Common Restaurant type	4th Most Restaurant type	5th Most Restaurant type	6th Most Restaurant type	7th Most Restaurant type	8th Most Restaurant type	9th Most Restaurant type	10th Most Restaurant type
59	Paerdegat Basin	Food	Asian Restaurant	Korean Restaurant	Wings Joint	Food Court	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant

Discussion and Conclusion

The number of neighborhoods within each cluster is summarized in the below table.

Cluster Name	Cluster Label	Number of Neighborhoods
Deli/Bodega	0	19
Pizza place, Italian & Mexican restaurants	1	48
Asian Restaurants	2	1

The cluster 0 has 19 neighborhoods and has the highest number of delis and bodegas followed by the Caribbean and fast food centers. This cluster is a thriving market for those investors who are looking to open a deli or new Caribbean restaurant. Since the number of neighborhoods in t his cluster is moderate, then these neighborhoods are possibly one of the promising locations t o start a new deli or Caribbean restaurant.

Cluster 1 consists of 48 neighborhoods. Its top food services are pizza places, Italian and Mexic an restaurants. With the variety of restaurant types and the high number of neighborhoods, the neighborhoods in this cluster might be a good choice to start up a new restaurant, especially if i t is from the top restaurant categories of this cluster.

Cluster 2 consists of only 1 neighborhood and the most common type of restaurants are Asian and Korean foods. It should be noted that this cluster is a thriving market for the foodservice industry, but startups may also face stiff competition as there is only one neighborhood is in this cluster.