



Data Science Capstone Project: The Battle of Neighborhoods

Edith Alvarado

1) Introduction/ Business Problem

Problem Statement: Prospects and Competitive Landscape for a Mexican Restaurant in New York City

New York has an estimated 18,804,000 people, a very densely populated and major city in the United States. It consists of five boroughs: Brooklyn, Queens, Manhattan, Bronx and Staten Island, and more than 300 neighborhoods. Our goal is to find a good neighborhood for opening a Mexican Restaurant in New York and what the competitive landscape looks like.

Target Audience:

- Individuals looking into opening a Mexican restaurant. This will provide an analysis of whether the venture is feasible and what the competitive landscape is.
- Investors that seek to invest in the opening of the company. This analysis will highlight potential areas to start or expand a business.
- Residents or visitors looking for Mexican restaurants.

A good location is a safe place with good amount of consumers' footfall who can afford the product and where the cost of doing business is optimum. This means data required for this purpose are: population density, demographics, purchasing power, and competitors. In this

project, data on population density, purchasing power (i.e, per capita income) and competitors (foursquare location data) has been used.

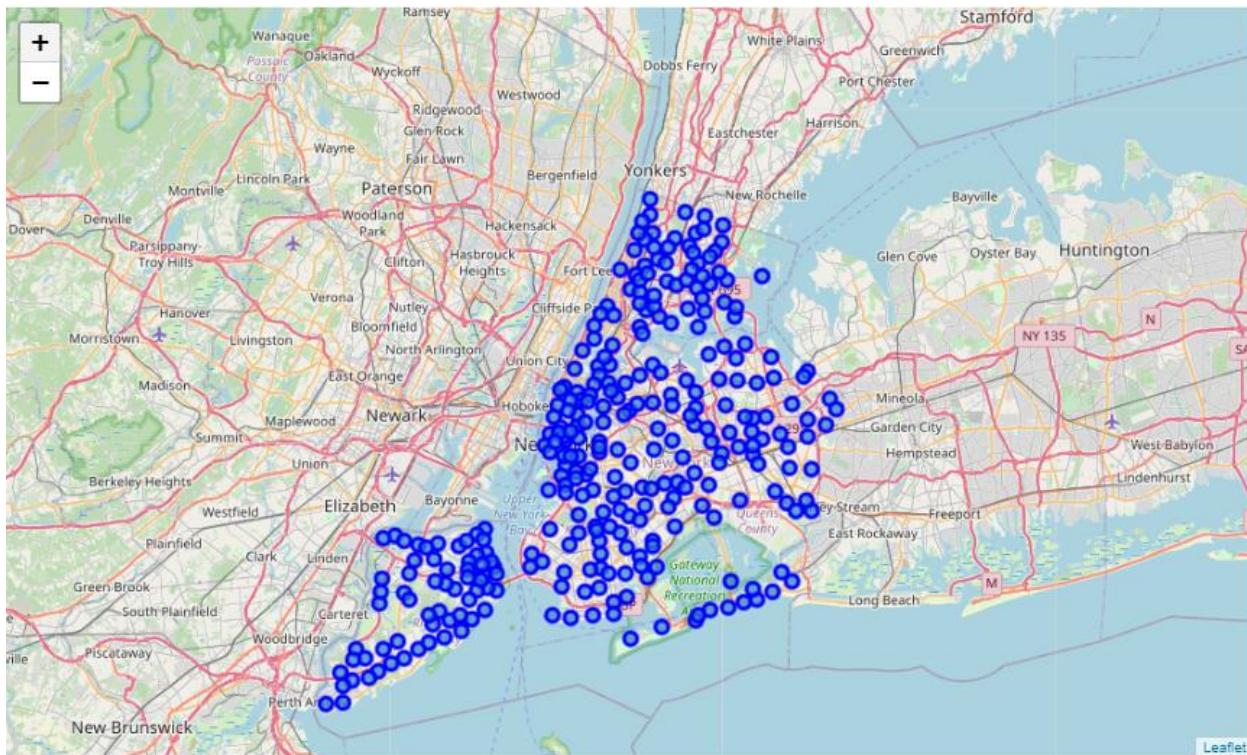
2) Downloading and Prepping Data

In this project, first, I obtained the data that contains the boroughs and the neighborhoods of New York City. It also contains the latitude and longitude of each neighborhood of every borough in the city. Later, using them, I obtained the data of food outlets/ eateries (along with their type) inside each Borough.

2A. First I visualized New York City neighborhoods, boroughs, latitudes, and longitudes. New York city neighborhoods, boroughs, latitudes, and longitudes. This JSON file is what I used to extract the 'features' data which has four columns 'Borough', 'Neighborhood', 'Latitude', 'Longitude'. This dataset exists for free on the web. Here is the link

- Links to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572, and to its downloadable json format file: https://cocl.us/new_york_dataset/newyork_data.json

With this extracted data, I make a map of all the 306 New York City Neighborhoods:

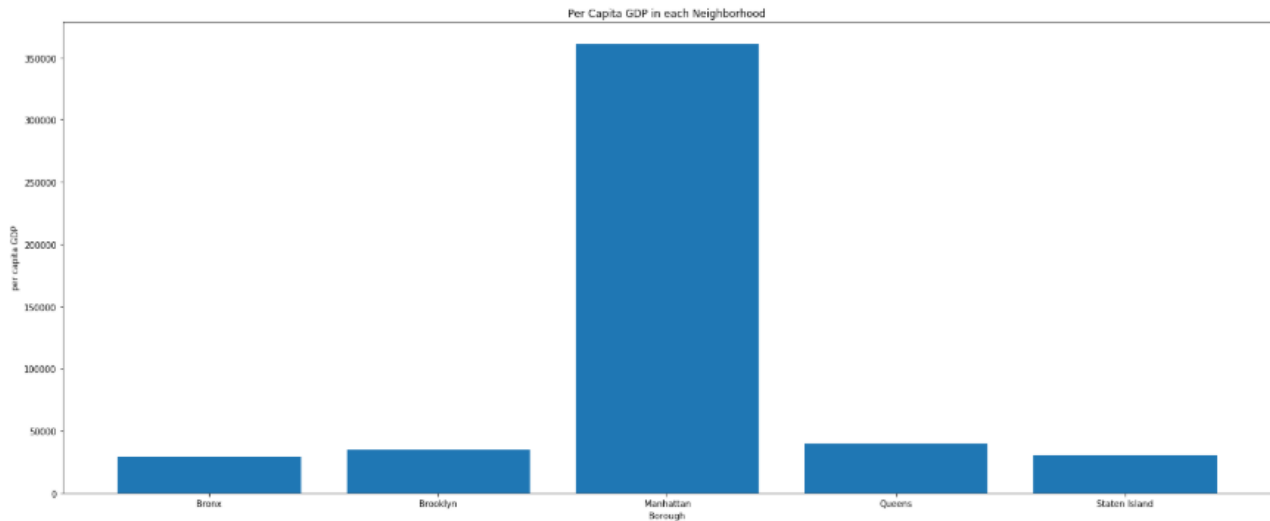


2B. Next, I scraped the Wikipedia webpage to obtain population density and per capita GDP of the five borough boroughs.

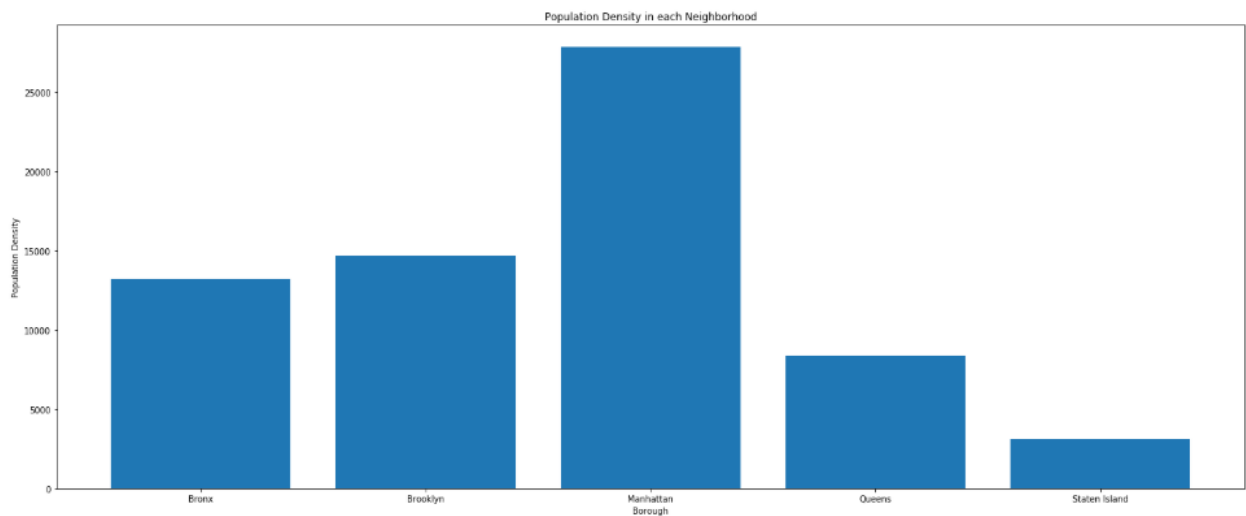
- Link: https://en.wikipedia.org/wiki/Demographics_of_New_York_City

I then transformed data into a data frame to visualize population density and per capita GDP:

Population Density in each Neighborhood:



Per Capita GDP in each Neighborhood:



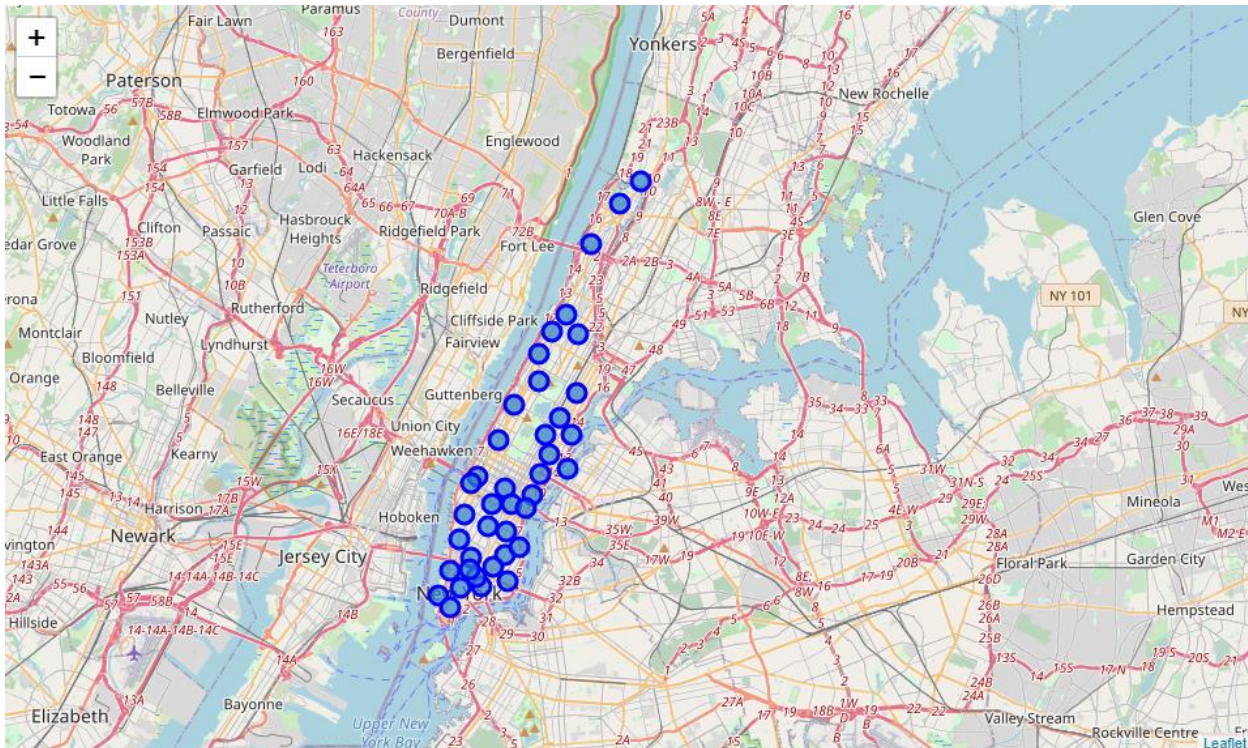
The figures show that **Manhattan** has the highest population density. Manhattan has more than double the population density of the rest of the boroughs.

Manhattan has the highest population density and per capita income, which translates to a higher purchasing power. Since Manhattan has the highest for both, Manhattan will be an ideal location to set up a new Mexican restaurant. I focused on Manhattan to figure out which neighborhood has the highest competition for Mexican restaurants. After I obtained the data needed for further exploratory data analysis.

3) Methodology

In this section, we will conduct exploratory data analysis. We have identified Manhattan as providing the highest population density and purchasing power. We will explore which neighborhood has the highest Mexican Restaurant competition.

3A. We will visualize the 40 Manhattan neighborhoods:



3B. Next, we will leverage Foursquare data to display the current restaurants in each region. Description: Foursquare API, a location data provider, will be used to make RESTful API calls to retrieve data about venues in different neighborhoods.

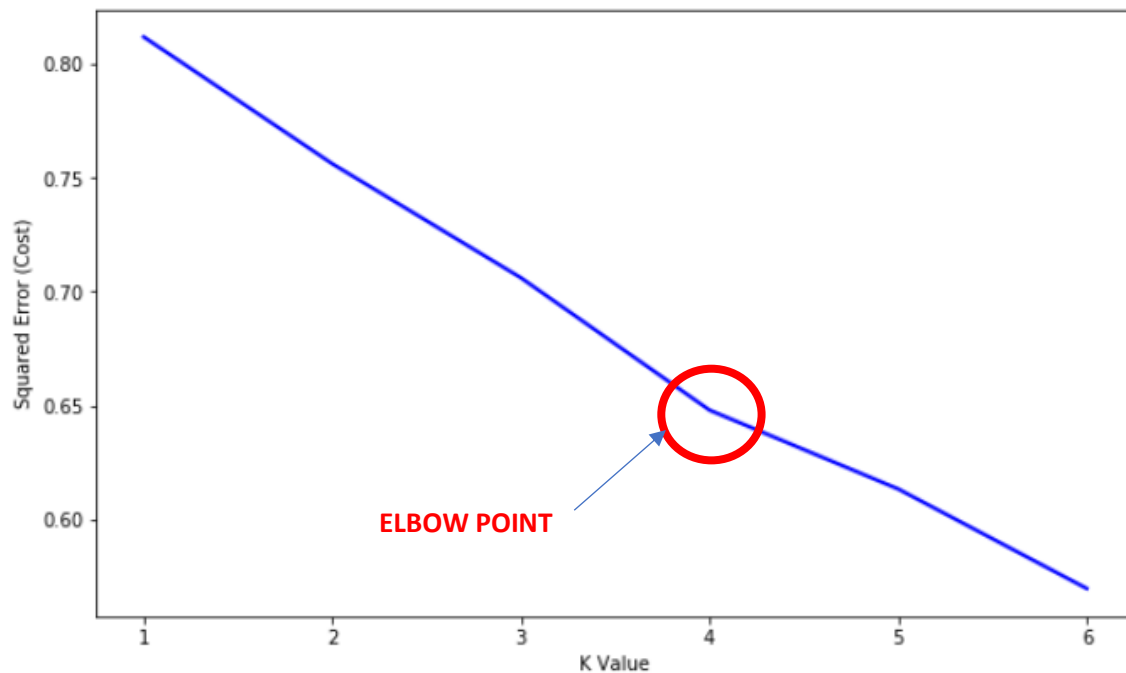
- Link: <https://developer.foursquare.com/docs>

In this section, we will save our data to a CSV file because of the limited API calls. After we group the data frame by neighborhood, we get the most common venues in a neighborhood and create a data-frame of the top 10 most common venues in each Manhattan neighborhood.

3C. After transforming the data, we will be using K-means clustering. In our problem, clustering helps to divide the neighborhoods of a given borough into clusters so that each neighborhood in a given cluster will show the venues with respect to the other neighborhoods within the same

cluster. This will also show the dissimilarity with respect to the neighborhoods present in the different clusters.

We used Foursquare data on venues within a 500-meter radius of each neighborhood in Manhattan. In order to run the k-means clustering, we needed the optimum k-value, which was discovered by analyzing the squared error cost against k-values. The elbow depicts the number of clusters needed to run:

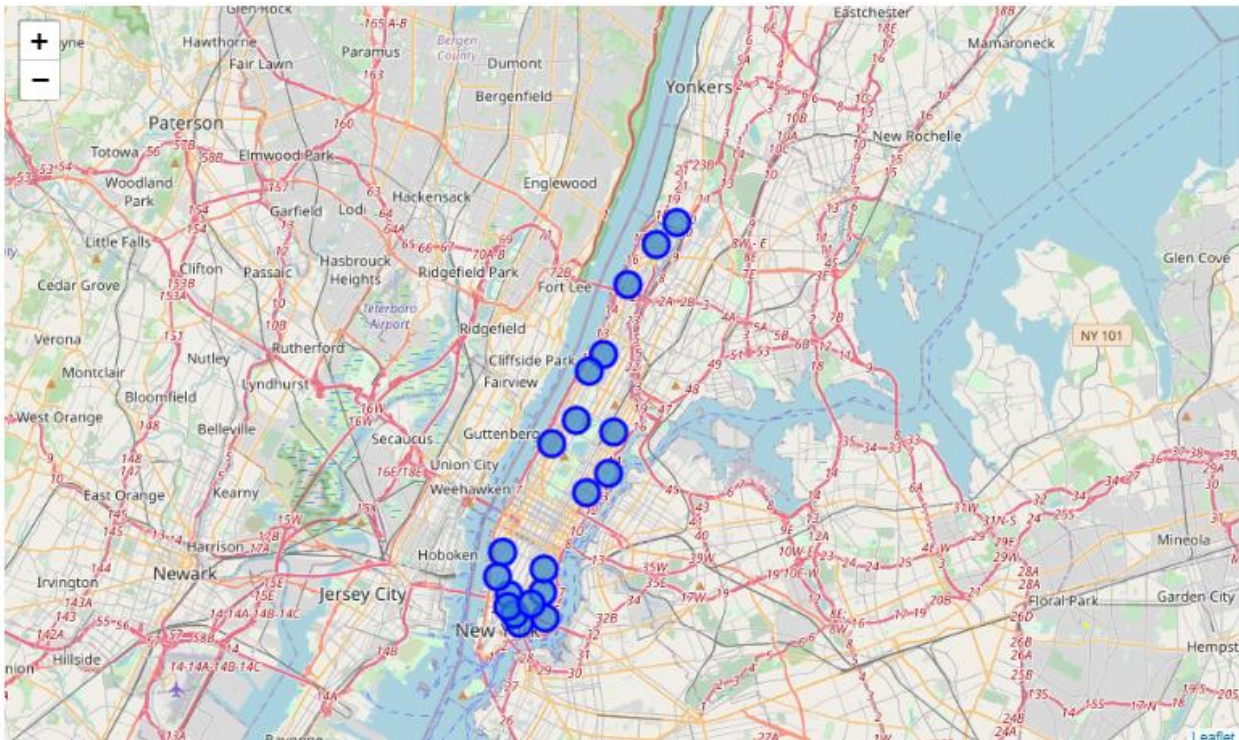


The optimum value of K= 4, visible at the elbow point

4) Results

In this section, I analyzed the five clusters to illuminate which cluster has the most restaurants as the most common venues.

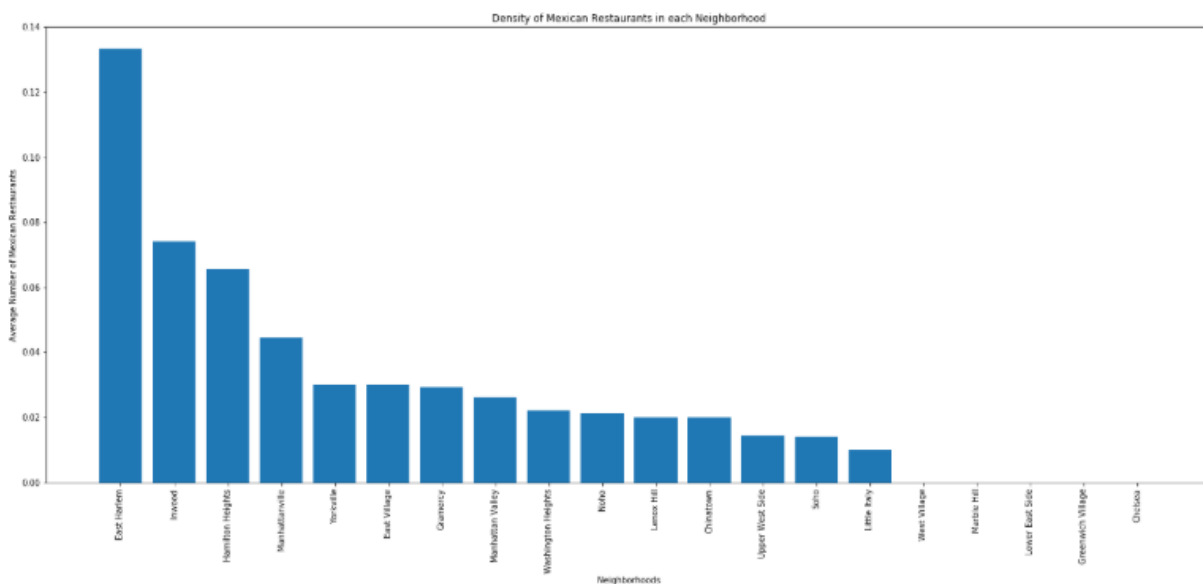
We saw that Cluster 0 is the cluster with the most restaurants and eateries as the 1st most common venue. We will select Cluster 0 as our final cluster to create a map. This will provide us with the number of neighborhoods we will further use to visualize the results. We will plot the mean value of Mexican restaurants for each selected neighborhood in the cluster to study the presence of competition in each neighborhood. After running the four clusters, we observed that cluster 0 provided the cluster with the most amount of food restaurants as the most common venue. With this cluster, we narrowed down the list of Manhattan neighborhoods to 20:



Neighborhoods identified from Cluster 0

5) Discussion

In this section, we build a bar chart of the density of Mexican Restaurants in these 20 neighborhoods for our discussion of the results:



Density of Mexican Restaurants in each Neighborhood

The graph depicts 20 neighborhoods obtained from Cluster 0, which is the cluster with the most restaurants as the most common venue. Based on the graph, you can see that the first neighborhood, East Harlem, has the most competition for Mexican Restaurants in Manhattan, indicating that it has the greatest obstacles in opening a new restaurant. Of these neighborhoods, it is also the most populated. East Harlem has at least double the competition than the other neighborhoods. Inwood is the second neighborhood with the most Mexican restaurants, followed by Inwood. However, the following neighborhoods have moderate competition, which will enable a new business to establish itself easier:

- Manhattanville
- Yorkville
- East Village
- Gramercy
- Manhattan Valley
- Washington Heights
- Noho
- Lenox Hill
- Chinatown

6) Conclusion

This project aimed at identifying a potential neighborhood to open a restaurant in New York City. For which neighborhood in New York City offers the greatest population density and which neighborhood offers the greatest per capita income. In this study, we concluded that Manhattan offers both, most people and also the greatest purchasing power of the population of the other four boroughs in New York. We obtained the location data of different types of venues with Foursquare API. This locational data aided us in identifying which cluster had restaurants as the most common venue in Manhattan. Neighborhoods were partitioned into different clusters using K-means clustering algorithm. Further analysis identified that for a Mexican restaurant, East Harlem provides the most competition, which will provide a harder ground to break. However, East Harlem is the most populated of the 20 neighborhoods analyzed. The 9 neighborhoods with moderate competition also had a lower population density than the top 3 neighborhoods with the most average number of Mexican Restaurants.

- Manhattanville
- Yorkville
- East Village
- Gramercy
- Manhattan Valley
- Washington Heights
- Noho
- Lenox Hill
- Chinatown

This analysis can be helpful for individuals looking into opening a Mexican restaurant. This will provide an analysis of whether the venture is feasible and what the competitive landscape is. This can benefit investors that seek to invest in the opening of the company by highlighting potential areas to start or expand a business. Lastly, this can assist residents or visitors by providing the areas with the most amount of Mexican restaurants per neighborhood.
