IBM Applied Data Science Capstone Project
-- Battle of Neighborhoods(Week 2)

INTRODUCTION

New York, as the largest city in the United State, is composed of people with different races and varied religious and literary backgrounds. According to the 2018 ACS, the racial composition of New York City was:

- White: 42.67%
- Black or African American: 24.27%
- Other race: 15.12%
- Asian: 13.95%
- Two or more races: 3.51%
- Native American: 0.43%
- Native Hawaiian or Pacific Islander: 0.05%

With the very diverse population, all types of cuisines can be found in NYC. American, British, Caribbean, Chinese, French, Greek, Indian, Italian, □□. You name it, you find it! Among all the cuisines, Chinese food is one of the most popular foods. Color, smell and taste are the three traditional aspects used to describe Chinese food, as well as the appearance and nutrition of the food. Not to mention, the big serving amount. You can easily feed yourself to full with a reasonable price.

Therefore it could be a great opportunity for an entrepreneur who is based on NYC. He/She might think of opening a Chinese restaurant. With the purpose in mind, finding a good location to open such a restaurant is one of the most important decisions to make. Meanwhile, with 5 boroughs and 306 neighborhoods in NYC, it is also not an easy decision to make. I am designing this project to not only satisfy my IBM Data Science Professional Certificate program capstone course requirement, but also help the entrepreneur find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the best location for an entrepreneur to open an authentic Chinese restaurant in NYC. By using data science methods and tools with machine learning algorithms such as clustering, the project is aiming to provide solutions to answer the business question: In NYC, if an entrepreneur targets on

opening a new Chinese restaurant, where is the most suitable place should he/she consider?

TARGET AUDIENCE

Someone who wants to find the location to open a new Chinese restaurant in NYC.

DATA

To solve the problem, we need data described as below:

- List of boroughs and neighborhoods in NYC
- Latitude and Longitude of the neighborhoods
- Chinese restaurant venue data

METHODOLOGY

- Loading and exploring NYC boroughs and neighborhoods data from a certain web site
- Using Folium library to visualize the neighborhoods in Queens, NY
- Getting Latitude and Longitude data of the boroughs via Geocoder package
- Using Foursquare API to get Chinese venue data by neighborhoods
- Running *k*-means to cluster the neighborhoods

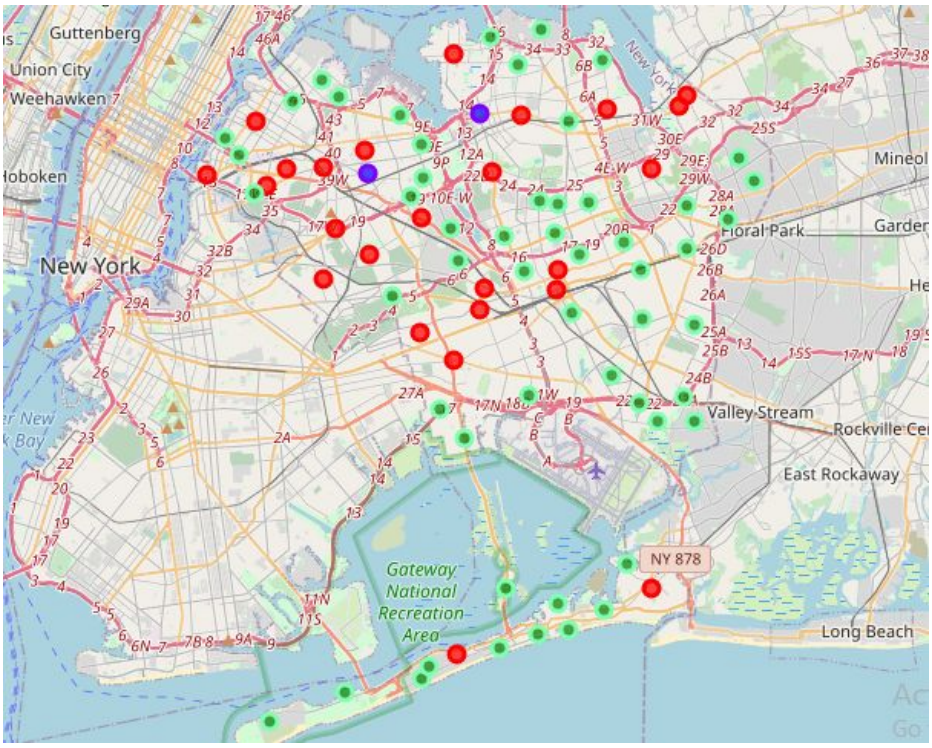
First we need to get new york data from the web site https://cocl.us/new_york_dataset. The data is downloaded as a json file, which contains the 5 boroughs and 306 neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. After exploring the json file, I transform the new york data into a *pandas* dataframe with 4 columns: 'Borough', 'Neighborhood', 'Latitude' and 'Longitude'.

Next, I want to focus on one of the boroughs - Queens right now, so I made a new dataframe with only neighborhoods located in Queens, NY. I ended up with totally 81 neighborhoods in Queens, NY. I visualize the map of Queens using the Folium package. Then I use the Foursquare API to pull the list of top 100 Chinese venues within 500 meters radius of a neighborhood center. I have created a Foursquare developer account in order to obtain an account ID and API key to pull the data. From Foursquare, I am able to get the names, categories, latitude and longitude of the Chinese venues. After going through all the 81 neighborhoods of Queens and combining the obtained data into one dataframe, we find 352 Chinese venues totally for

Queens, NY. Now we clean the dataframe by deleting the non-restaurant venues and end up with 349 Chinese restaurants found in Queens, NY. By using dataframe unique function, I found out that all the Chinese restaurants are located in 61 neighborhoods. It means there are 20 neighborhoods with no Chinese restaurant. Taking this into account, together with the Queens neighborhood dataframe, I made a new dataframe including neighborhood name, latitude, longitude and Chinese restaurants venue count.

Now I am ready to do clustering. I performed the clustering by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is considered as one of the simplest and useful unsupervised machine learning algorithms. I have clustered the neighborhoods in Queens into 3 clusters based on their Chinese restaurants counts. And from the clustering results, the ideal locations to open a Chinese restaurant can be recommended.

RESULTS



The results from K-means clustering show that Queens neighborhoods can be categorized into 3 clusters based on how many Chinese restaurants are counted in each neighborhood:

- Cluster 0: Neighborhoods with median numbers of Chinese restaurants.
- Cluster 1: Neighborhoods with large numbers of Chinese restaurants.

- ❖ Cluster 2: Neighborhoods with zero or almost zero number of Chinese restaurants.

The results are visualized in the above map with Cluster 0 in red, Cluster 1 in blue and Cluster 2 in green.

RECOMMENDATIONS

As one can tell from the map, two neighborhoods in Queens are packed with lots of Chinese restaurants. They are Elmhurst and Flushing. This means, if you want to find authentic Chinese food, you should go visit Elmhurst or Flushing in Queens, NY.

However, other than Elmhurst and Flushing, not so many Chinese restaurants can be found in other neighborhoods in Queens. Furthermost, there is no Chinese restaurant in the neighborhoods as categorized in Cluster 2, such as South Ozone Park, Whitestone, Briarwood,etc. Those are definitely ideal locations to open new Chinese restaurants.

CONCLUSION

We study the neighborhoods in Queens, NY to give suggestions about where to find authentic Chinese food and where to possibly open new Chinese restaurants in Queens. We can extend our study of NYC neighborhoods to other boroughs (Manhattan, the Bronx, Brooklyn and Staten Island) to give useful recommendations.