Coursera Capstone Project
IBM Applied Data Science
Opening a New Filipino Restaurant in New York City

By: Greg Martin Teo
June 2020

Business Problem

- New York City is most densely populated major city in the United States. With an estimated population of 8.3 million in 2019. It is composed of five boroughs, Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.
- Find an optimal location for stakeholders who are interested in opening a Filipino Restaurant in one of the Boroughs of New York City.
- Project will take into account several aspects including restaurant density and diversity within the area and proximity to city centers.
- Business Question: Where in New York City is the optimal location to open a new Filipino Restaurant?

Data

Selection Criteria for our Business Problem:

- Restaurant Density within an area
- Restaurant Diversity within an area
- Distance from city center

Data sources that will be needed for the project:

- Neighbourhoods data from Wikipidea or other Web Pages by web scraping using **BeautifulSoup**
- Latitude and longitude of neighbourhoods using **Google Maps Geodcoding API**
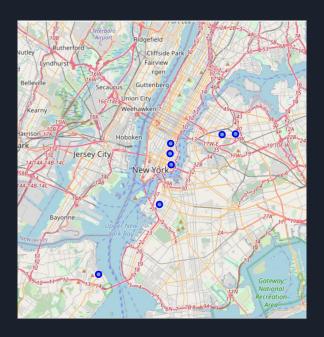- Number and type of restaurant in an area using **Foursquare API**

Methodology

- Collected the data of every establishment within the 10 miles of New York City center and extracted its' location and type and name.
- Identified the number of restaurants within that radius and focused our analysis on Filipino restaurants with the use of Foursquare.
- Analysis will be calculation and exploration of 'restaurant density' across different neighborhoods of New York City.
- Determine the frequency of each type of restaurant in the neighborhood and rank them as to which is more common in the area.
- Use k-means clustering on those neighborhoods we have identified and identify the most promising areas and within those created clusters.
- Present a map of all such locations but also create clusters (using k-means clustering) of those locations to identify clustered neighborhoods to better visualize

Results

Categorized the neighbourhoods into 5 clusters:

- Cluster 0 : Neighborhood with high diversity of Filipino Restaurants
- Cluster 1: Neighborhood with no Filipino Restaurants
- Cluster 2: Neighborhood with high density of Filipino Restaurants
- Cluster 3: Neighborhood with no Filipino Restaurants
- Cluster 4: Neighborhood with no Filipino Restaurants

Discussion

- Only a handful of Filipino Restaurants a 10 mile radius of the City Center.
- Highest concentration of Filipinos restaurants is in Queens Borough, Woodside Neighborhood.
- Identifying that out of 1903 restaurants and only having 12 Filipino Restaurants, we have established that only 0.63% of the total restaurants are Filipino Restaurants.
- Filipino Restaurants are within clusters 0 and 2.

Conclusion

Answer to business question:

It is evident that within a 10 mile radius there is an abundance of restaurants, but the project outlines that there is only a handful of Filipino Restaurants. In order for stakeholders to find the optimal location, they must further look into the listed neighborhoods above based on specific characteristics of neighborhoods and locations in every recommended neighborhood. Stakeholders must also take into account additional factors like attractiveness of each location foot traffic, proximity to public transport, real estate prices, socioeconomic factors and culture of neighborhood.