Author

Marko Knight

Course name

IBM Professional Data Science Certificate: Applied Data Science Capstone Project

Title

Battle of the Neighborhoods; what top three business types can an investor invest in or can someone consider opening in neighborhoods of Bronx and Queens (two New York boroughs)

Contents

1 Introduction

The New York City (NYC) is one of the most popular towns, not only in the United States, but worldwide as well. It has several booming business opportunities.

It also holds fame to its Stock Exchange (NYSE). Has quite substantial population that moves around to explore and find survival opportunities. The City has five boroughs [1], namely,

- Bronx
- Manhattan
- Queens
- Brooklyn
- State Island

Therefore whether one seeks to run a startup business or one is a prolific businessman or known investor. Wanting to start a business that will be successful in NYC, requires one to be knowledgeable of the current business dynamics in the boroughs of interest. So having insights about common types of businesses that exist in the boroughs of NYC might be insightful to decide if to what business type to start or invest in. Study about common business types in the 306 neighborhoods will help give the potential investors options as to what types of businesses to consider to be viable in the area. Hence this will help increase the level of confidence in the success of the planned business or investment by starting a business that is more aligned to the needs of the respective neighborhoods.

1.1 Problem statement

What top three common business types can someone (business person/ invester) consider to start in each NYC borough neighborhoods of Bronx and Queens

1.2 Areas of Interest

This study will focus on two boroughs (Bronx and Queens) of NYC. It will survey the existing neighborhoods businesses and advise on what options can someone consider in such areas. Will render possible business types that can successfully run in the neighborhoods. This will hence assist one in choosing a kind of business to start or invest in.

1.3 Target audience

Local or international business people or investors. Local community; those that are interested in starting business in the area but are confused as to what options to consider. Cooperates that want to

franchise in the area. Open to any entrepeneural person interested to start something in neighborhood of Bronx and Queens.

2 Data section

2.1 Data sources

The dataset that will be used is the New York dataset from the New York University, Data link is https://geo.nyu.edu/catalog/nyu_2451_34572. The data essentially contains five boroughs and the neighborhoods data that exist in each borough as well as the latitude and longitude (spatial data) coordinates of each neighborhood. The Boroughs that will be analyzed/of interest for the problem statement are Bronx and Queens and their respective neighborhoods.

New York City spatial coordinates data will be used as input through the Foursquare API, that will be leveraged to provision venues information for each neighborhood. Foursquare API will be used to explore neighborhoods in New York City. Foursquare API credentials are required when exploring the venues.

The typical final data layout when the retrieved Foursquare location data of the venues is merged with the neighborhood spatial data is as shown in Table 1, where the neighborhoods of Manhattan are shown in this case for illustration purposes.

Table 1: Data layout of the neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant

Data analysis from such table above will help generate results for the required outcome.

3 Methodology

- Access the New York dataset source
- Retrieve the boroughs data
- Use geopy library to get spatial data of the neighborhoods of Bronx and Queens
- Utilize Foursquare API to explore neighborhoods venues
- Generate neighborhoods maps through folium library
- Perform data cleaning and preparation i.e. data transformation and exploration
- Explore and analyze the neighborhoods of Bronx and Queens

- List top three most common business types from both boroughs for recommendation
- Use k-Means to cluster the neighborhoods and analyze the resulting clusters
- Generating the expected observations and provide concluding remarks

4 Results and discussion

The New York City (NYC) and its neighborhoods is visualized as shown in Figure 1. It can be observed that it is quite dense. This makes sense as the its boroughs, namely, Bronx, Queens, State Island, Brooklyn and Manhattan hosts quite a large number of people also given that NYC is one of the busiest city in most avenues of business.

The map was generated via the use of the folium library, the spatial details, i.e. latitude and longitude were retrieved via the use of geopy library. The spatial coordinates of NYC are 40.7308619, -73.9871558, latitude and longitude respectively.

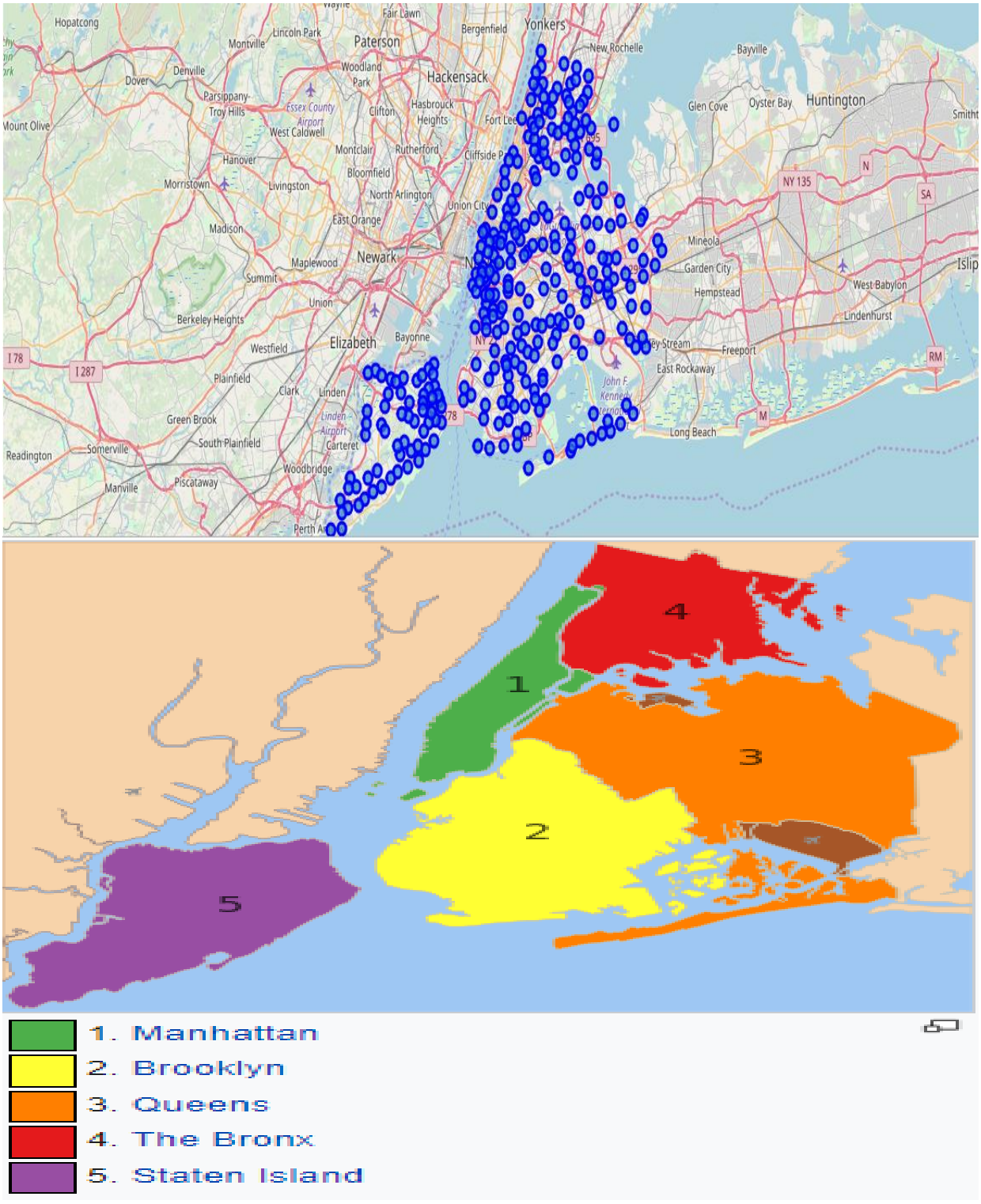

Figure 1: The NYC map and the surrounding places (boroughs).

Upon getting acquainted with the NYC map, the boroughs (Queens and Bronx) to be examined were also plotted as well. The first borough to be looked at was the Queens. Its geographical layout is shown in Figure 2.

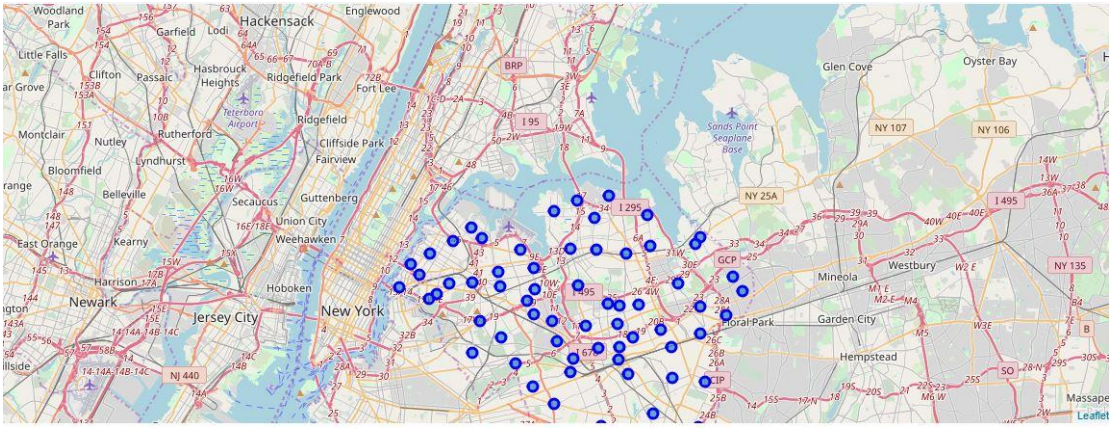

Figure 2: The Queens map.

The spatial coordinates of some of the neighborhoods in the Queens are given in Table 2. As expected the spatial coordinates values are the same as the NYC values which makes sense.

Table 2: Queens neighborhoods spatial coordinates (head of dataframe).

	name	categories	lat	lng
0	Favela Grill	Brazilian Restaurant	40.767348	-73.917897
1	Orange Blossom	Gourmet Shop	40.769856	-73.917012
2	Titan Foods Inc.	Gourmet Shop	40.769198	-73.919253
3	CrossFit Queens	Gym	40.769404	-73.918977
4	Simply Fit Astoria	Gym	40.769114	-73.912403

The next step was to explore all the neighborhoods of the Queens. This is such that the categories of the places are known. To also help classify the leading business types.

The Foursquare API was used in this case to retrieve the venues categories and the location data in the neighborhoods. So the use of the API helped in exploring the neighborhoods, such that occurrences of venues in each neighborhood were retrieved and placed in a dataframe. The next step was to perform one-hot encoding such that the venues could be grouped per the type of the business in each neighborhood.

At the end the final result was as shown in Table 3, where the head of the dataframe of Queens neighborhoods are given along the ranking of the common venues that occur in the respective neighborhood. The selected scale is from 1-10.

Table 3: The Queens venues ranking.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Queens	Astoria	40.768509	-73.915654	0	Middle Eastern Restaurant	Bar	Hookah Bar	Greek Restaurant	Bakery	Seafood Restaurant	Coffee Shop	Indian Restaurant	Latin American Restaurant	Food Truck
1	Queens	Woodside	40.746349	-73.901842	0	Grocery Store	Pub	Filipino Restaurant	Bakery	Thai Restaurant	Donut Shop	American Restaurant	Pizza Place	Bar	Latin American Restaurant
2	Queens	Jackson Heights	40.751981	-73.882821	0	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Mexican Restaurant	Bakery	Mobile Phone Shop	Thai Restaurant	Spanish Restaurant	Diner	Empanada Restaurant
3	Queens	Elmhurst	40.744049	-73.881656	0	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Indonesian Restaurant	South American Restaurant	Park	Colombian Restaurant	Salon / Barbershop
4	Queens	Howard Beach	40.654225	-73.838138	0	Italian Restaurant	Clothing Store	Bagel Shop	Pharmacy	Fast Food Restaurant	Sandwich Place	Chinese Restaurant	Breakfast Spot	Shipping Store	Sushi Restaurant

From the results of Table 3, the series of the 1st, 2nd and 3rd most common venues were chosen upon from which the most common business types will be examined. The results of these series were counted based on the occurrence of each type of business in the respective neighborhood.

The graphical perspective of each series was made so that the most occurring business type can be seen. From the 1st most common venues, the representation is as shown in Figure 3. It is observed that the top three most venues in this series are Deli/Bodega, Pizza place and the Pharmacy.

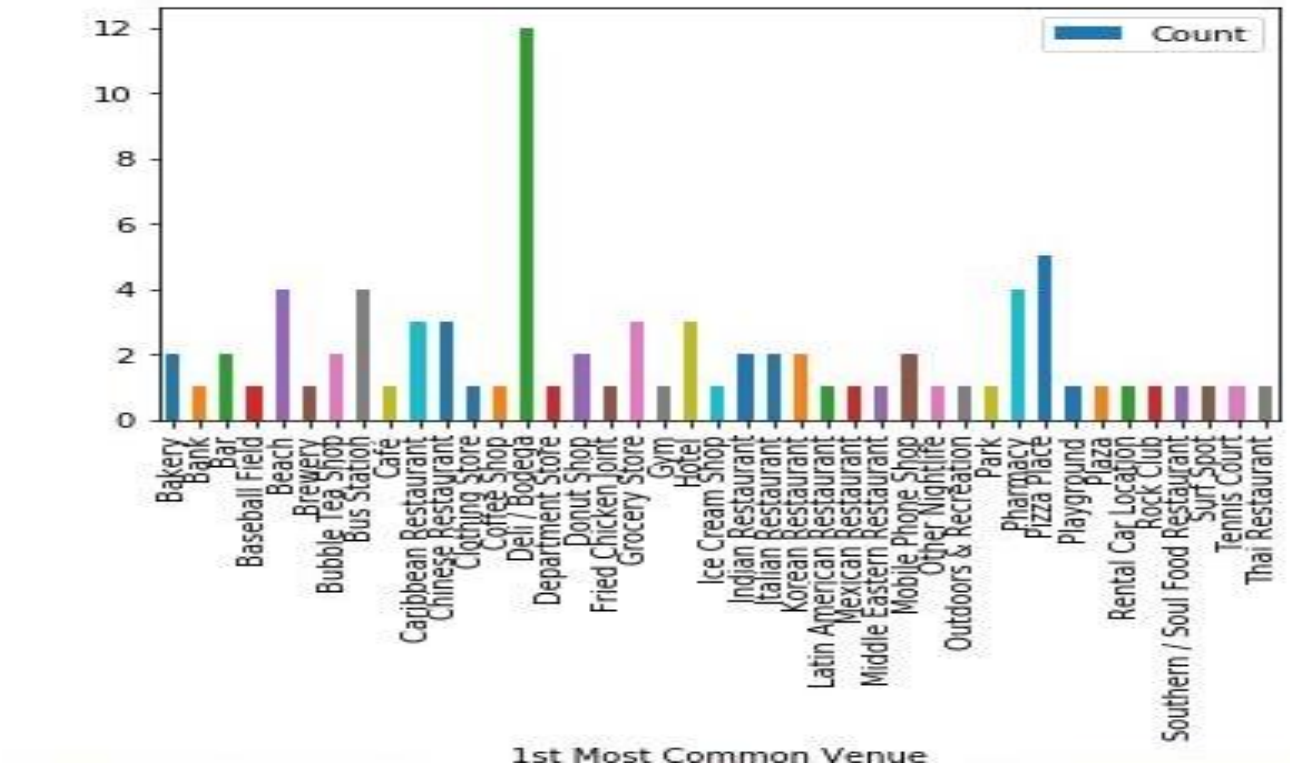

Figure 3: Queens first level venues.

In the 2nd top most common venues series, as shown in Figure 4, the top 3 most common venues are Deli/Bodega, Pizza place and the Grocery store.

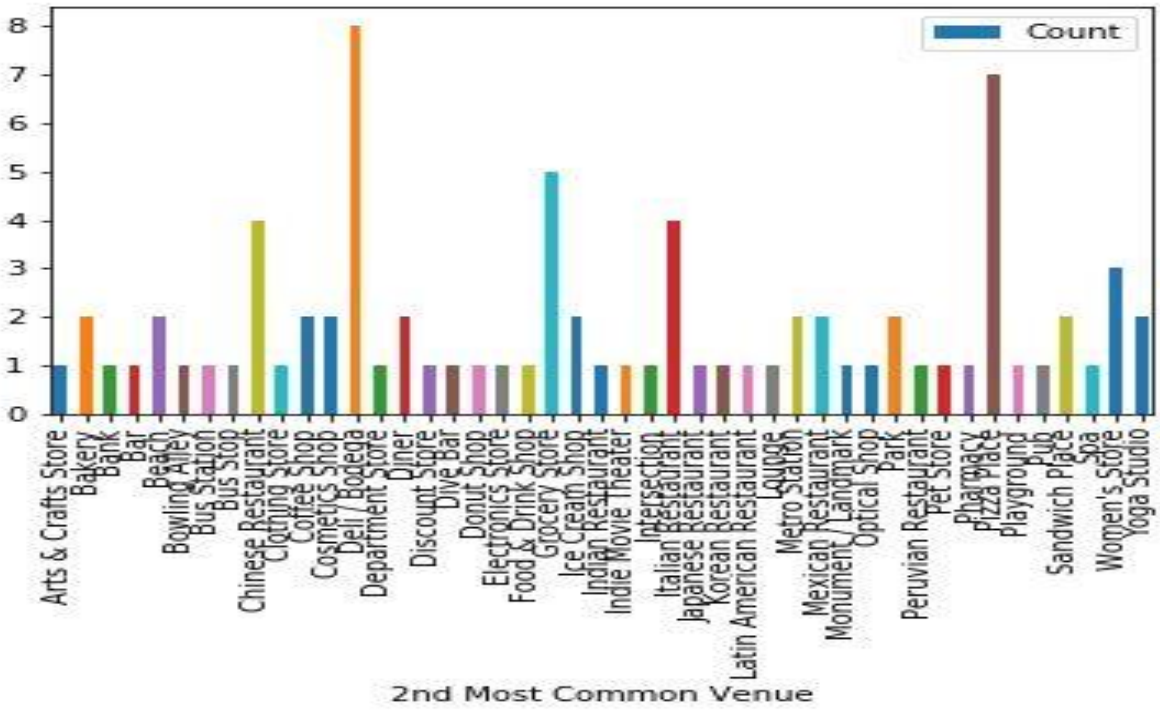

Figure 4: Queens second level venues.

In the 3rd top most common venues series, as shown in Figure 5, the top 3 most common venues are Chinese restaurant(Pizza place), Deli/Bodega and Bakery(Supermarket, Bakery, American restaurant).

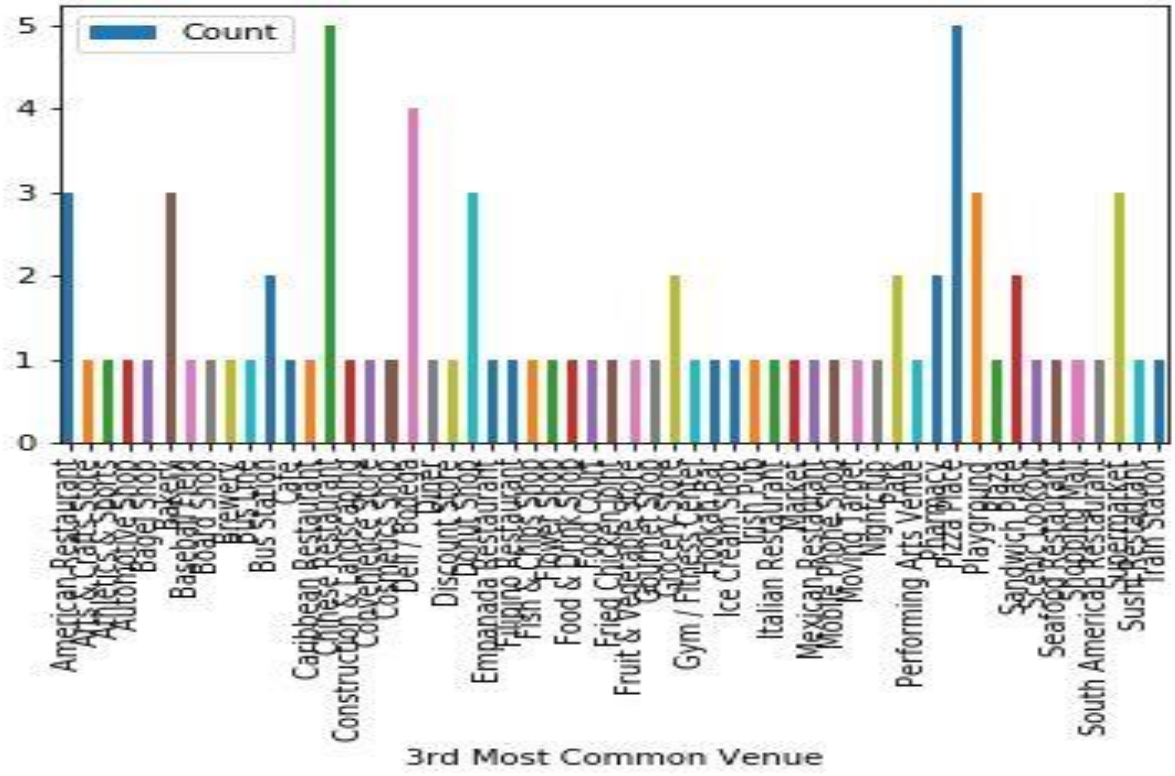

Figure 5: Queens third level venues.

The summary results gathered from the top three most common venues data of Queens is given in Table 4.

Table 4: Summary results of top most venues in the Queens.

Level	1st top most common venues	2nd top most common venues	3rd top most common venues
First	Deli/Bodega	Deli/Bodega	Chinese restaurant and Pizza place
Second	Pizza place	Pizza place	Deli/Bodega
Third	Pharmacy	Grocery store	Bakery, Supermarket, and American restaurant

The last step was to run the k-Means and cluster the neighborhoods into five clusters. The clusters for Queens are shown in Figure 6. The clusters bear similarities based on the types of businesses run in the

neighborhoods of Queens. There is few members that are not of the same category of the high number of fastfoods outlets common in the neighborhoods.

Figure 6: The Queens clusters map.

Since two boroughs are being examined, the next to be looked at was the Bronx. Similar techniques were used again to get similar results but now in the case of the Bronx borough. The Bronx map is as shown in Figure 7.

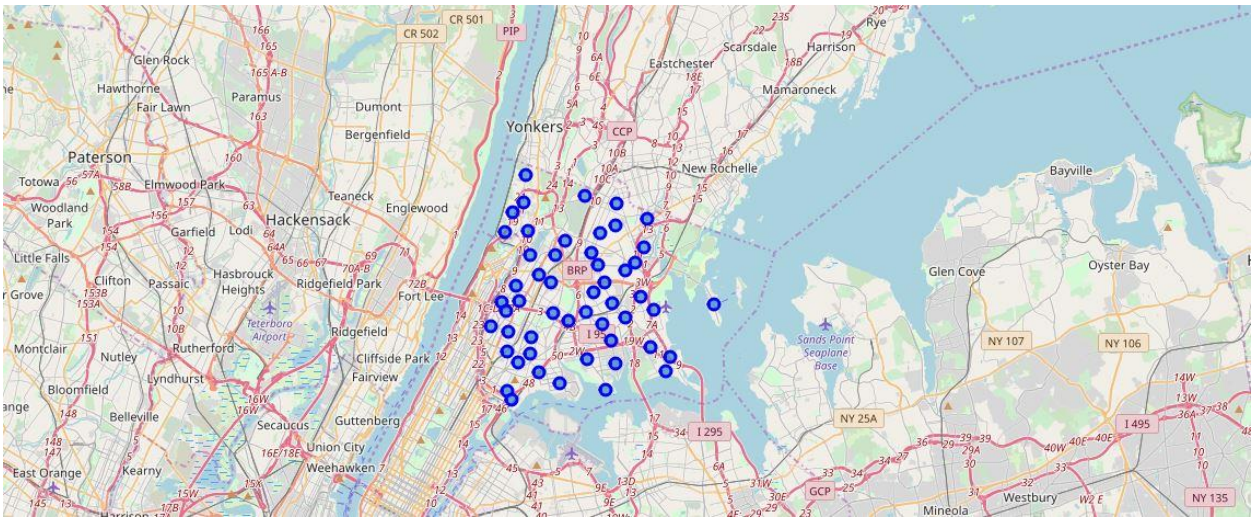

Figure 7: The Bronx map.

The spatial coordinates of some of the neighborhoods in the Bronx are given in Table 5.

Table 5: Spatial coordinates.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The ranking of the venues result is shown in Table 6.

Table 6: The Bronx venue levels.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	0	Sandwich Place	Ice Cream Shop	Pharmacy	Food Truck	Dessert Shop	Caribbean Restaurant	Donut Shop	Laundromat	Dive Bar	Fast Food Restaurant
1	Bronx	Co-op City	40.874294	-73.829939	0	Bus Station	Accessories Store	Fried Chicken Joint	Pharmacy	Park	Restaurant	Discount Store	Fast Food Restaurant	Pizza Place	Liquor Store
2	Bronx	Eastchester	40.887556	-73.827806	0	Caribbean Restaurant	Bus Station	Deli / Bodega	Diner	Bus Stop	Metro Station	Bowling Alley	Pizza Place	Platform	Convenience Store
3	Bronx	Fieldston	40.895437	-73.905643	1	Plaza	River	Playground	Distillery	Fast Food Restaurant	Farmers Market	Eye Doctor	Electronics Store	Eastern European Restaurant	Donut Shop
4	Bronx	Riverdale	40.890834	-73.912585	2	Park	Plaza	Bank	Gym	Food Truck	Playground	Bus Station	Home Service	Art Museum	Donut Shop

From the 1st most common venues, the representation is as shown in Figure 8. It is observed that the top three most venues in this series are Pizza place, Deli/Bodega (Italian restaurent), and the Pharmacy.

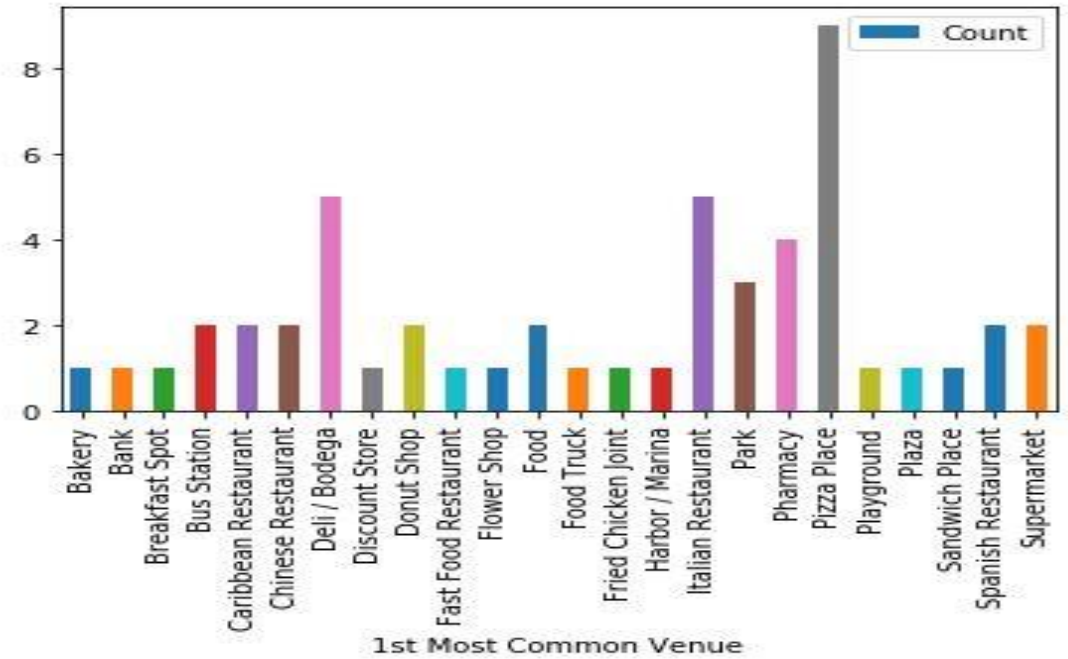

Figure 8: Common places in first level in the Bronx.

In the 2nd top most common venues series, as shown in Figure 9, the top 3 most common venues are Pizza place, Deli/Bodega, and Supermarket.

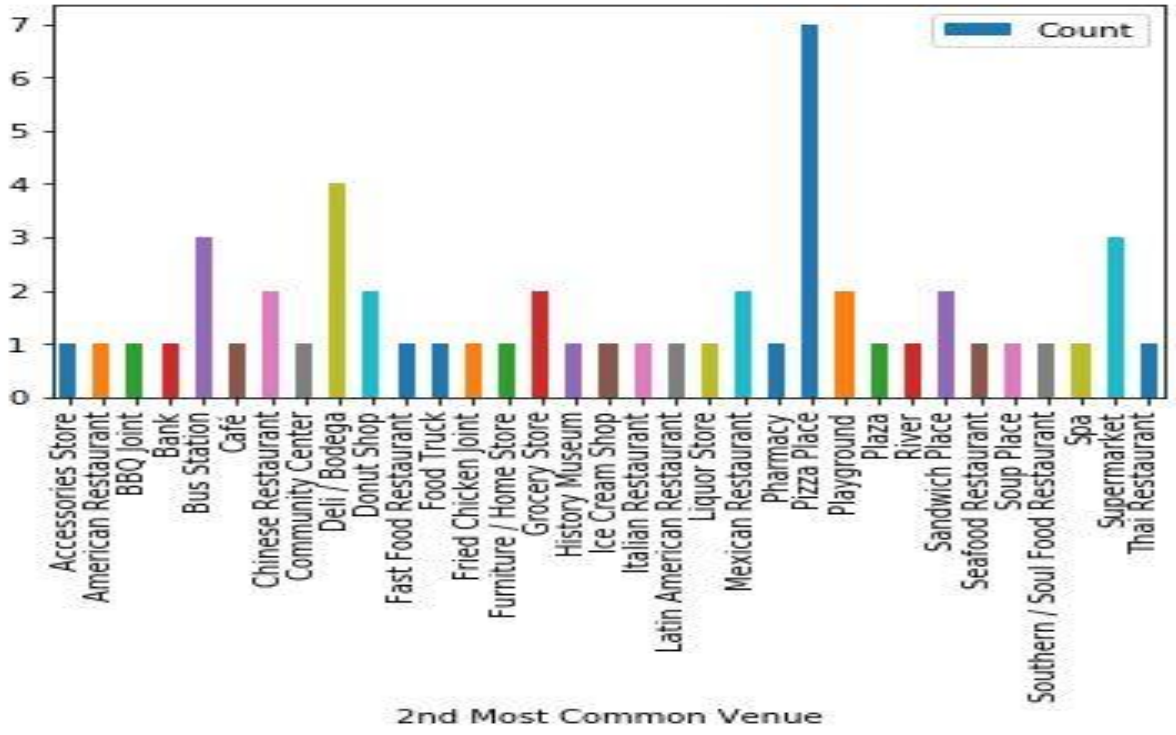

Figure 9: Common places in second level in the Bronx.

In the 3rd top most common venues series, as shown in Figure 10, the top 3 most common venues are Pizza place (Bank and Chinese restaurant), Breakfast spot (Pharmacy) and Deli/Bodega (Diner).

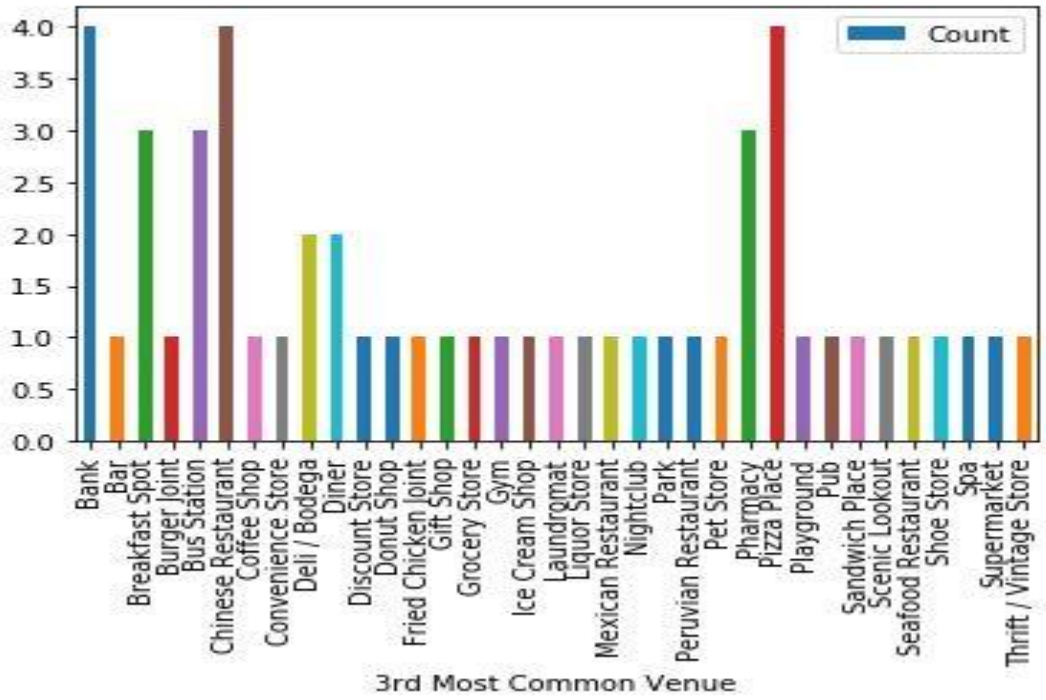

Figure 10: Common places in third level in the Bronx.

The summary results gathered from the top three most common venues data of Bronx is given in Table 7.

Table 7: Summary results of top most venues in the Bronx.

Level	**1st top most common venues**	**2nd top most common venues**	**3rd top most common venues**
First	Pizza place	Pizza place,	Pizza place, Bank and Chinese restaurant
Second	Deli/Bodega and Italian restaurant	Deli/Bodega	Breakfast spot , Pharmacy
Third	Pharmacy	Supermarket	Deli/Bodega and Diner

The last step was to run the k-Means and cluster the neighborhoods into five clusters. The clusters for Bronx are shown in Figure 11. Also even here, the clusters bear similarities based on the types of businesses run in the neighborhoods of Bronx. There is quite few members that are not of the same category of the high number of fastfoods outlets common in the neighborhoods.

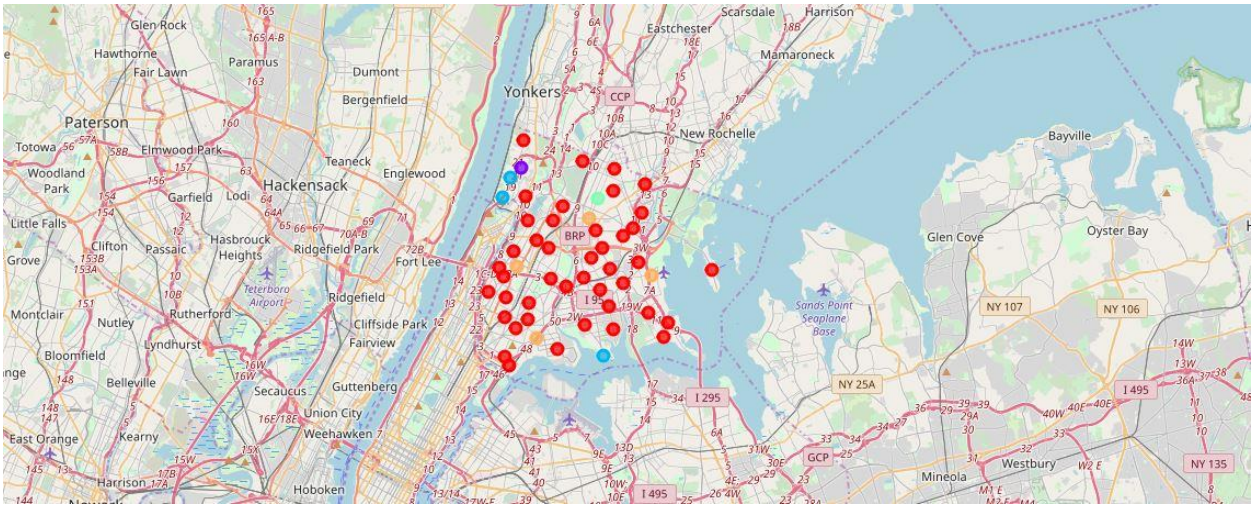

Figure 11: The Bronx clusters map.

Therefore from the observations made on the data of Queen and Bronx one can confidently suggest that any person (investors/franchising companies/entrepreneurs/businessperson) wanting to start business in these areas, they can choose from the ones in Table 8 or consider a similar combination.

Hence the most business types with high likelihood of success in the neighborhoods of Queens and Bronx are exactly as outlined in Table 8. So someone can with high level of certainty consider investing in or starting similar business types of Table 8. One can creatively consider a combination of the business types in Table 8 to suit the market of the neighborhoods.

Table 8: Most common business types in Queens and Bronx worth considering.

Level	**Queens recommendations**	**Bronx recommendations**
First	Deli/Bodega	Pizza place
Second	Pizza place	Deli/Bodega
Third	Pharmacy/Grocery store/Bakery, Supermarket, and/ American restaurant	Pharmacy/Supermarket and/Diner

5 Conclusion

Most three trending businesses in Bronx and Queens offer eatery services. Therefore it is essential for a business person to consider restaurant/fast foods oriented outlets in order to succeed in the neighborhoods of Bronx and Queens. The business types common in the neighborhoods are descriptive of the environment dynamics of NYC boroughs where most people are busy and prefer quick ready to eat food.

6 Recommendations

Additional details on the dataset such as the profit margins of each business can be helpful in selecting the most profitable business of them all. Also it would be great to get the health-status dataset of the people of NYC boroughs to explore which ones are healthiest/least-healthy based on the business services that are easily available in the neighborhoods.

7 References

[1] “Boroughs of New York City - Wikipedia”. [Online]. Available: https://en.wikipedia.org/wiki/Boroughs_of_New_York_City. [Accessed: 27-Mar.-2019].