Battle of the Neighborhoods:
A City Analysis Report: A comparison of Brooklyn and Queens, NY

Dr. Ahmet TARTAR

I. Introduction

New York is one of the largest metropolitan cities in the world. New York City (NYC), also known as the City of New York or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km^2), New York is also the most densely populated major city in the United States [1]. It is one of the cities where people from all over the world want to visit or live.

New York City encompasses five county-level administrative divisions called boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each borough is coterminous with a respective county of New York State. The boroughs of Queens and the Bronx are concurrent with the counties of the same name, while the boroughs of Manhattan, Brooklyn, and Staten Island correspond to New York, Kings, and Richmond counties [2].

Brooklyn: Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough. Brooklyn is known for its cultural, social, and ethnic diversity, an independent art scene, distinct neighborhoods, and a distinctive architectural heritage. Downtown Brooklyn is the largest central core neighborhood in the outer boroughs [3].

Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough, the most ethnically diverse county in the United States, and the most ethnically diverse urban area in the world [4].

Problem and Purpose

In the light of these information, it is very important to have information about these boroughs for those who want to move from another country of the world, or to relocate to a new city, or to start a new business.

In general, when we think about the best place to live, to relocate to a new business, a lot of parameters should be considered when trying to make a comparison between cities, towns, or neighborhoods. Here is some reasons:

- *Overall Comparison*: This is a comparison of the same factors for each city. Some of the popular factors include population, cost of living, average rent, crime rate, tax rates, and air quality.
- *Crime Rates*: The statistics are two cities, then measures them against national statistics.
- *Comparison of Costs of Living and Salary:* Some factors for this comparison include statistics on food, housing, utilities, transportation and more. This is a useful way to find out if your salary will measure up in the new city.
- *Compare Schools*: This is the best school in the area. It mostly takes into consideration the test scores and the student ratios, including the teacher's experience.
- *Neighborhood Comparison*: This is the best place to live in any given city. These sites allow you to see some interesting facts about the various communities.

In this study, we will prepare a city analysis report on the two major borough, Brooklyn and Queens in NY. So we will learn about the similarities and differences about these cities and the characteristics of the cities themselves.

Of course, for a complete evaluation, many of the above parameters should be evaluated. While all of these analyses are useful for comparing the neighborhoods, there is nothing like visiting the actual city, seeing the neighborhoods, and speaking with residents. If it's possible, an in-person visit is highly recommended before making a big move or relocating decision.

II. DATA

We will convert addresses into their equivalent latitude and longitude values. Also we will use the Foursquare API to explore neighborhoods, **Brooklyn and Queens** in New York City. You will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. You will use the *k*-means clustering algorithm to complete this task. Finally, you will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters.

In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

III. METHODOLOGY

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the data. This was used to:

- Segment the data into Brooklyn and Queens Cities, NY.
- Handling the missing values, verify the quality of data
- Determining likely approaches to modelling which might best yield to good clustering.

Segmenting and Visualizing

In cluster modelling, the selection of the variables of the data is very important part. A prerequisite of the work is that the FourSquare API is used to collect the venues information. So, It is very important that the dataset includes the coordinates of the cities. The subjects in the data for analysis includes Neighborhood name and the values of latitude and longitudes of both neighborhoods and venues for each city.

To view the sliced data for both cities, folium was used. Folium is a powerful python library that builds on the data wrangling strengths of the python ecosystem and the mapping strengths of Leaflet.js library. Generally, data is manipulating in Python, and then visualize it on Folium. In

order to visualize the data on Folium, the coordinate of a location in Brooklyn was obtained and then plotted to view the location on a map. That was also provided for the data of Queens City.

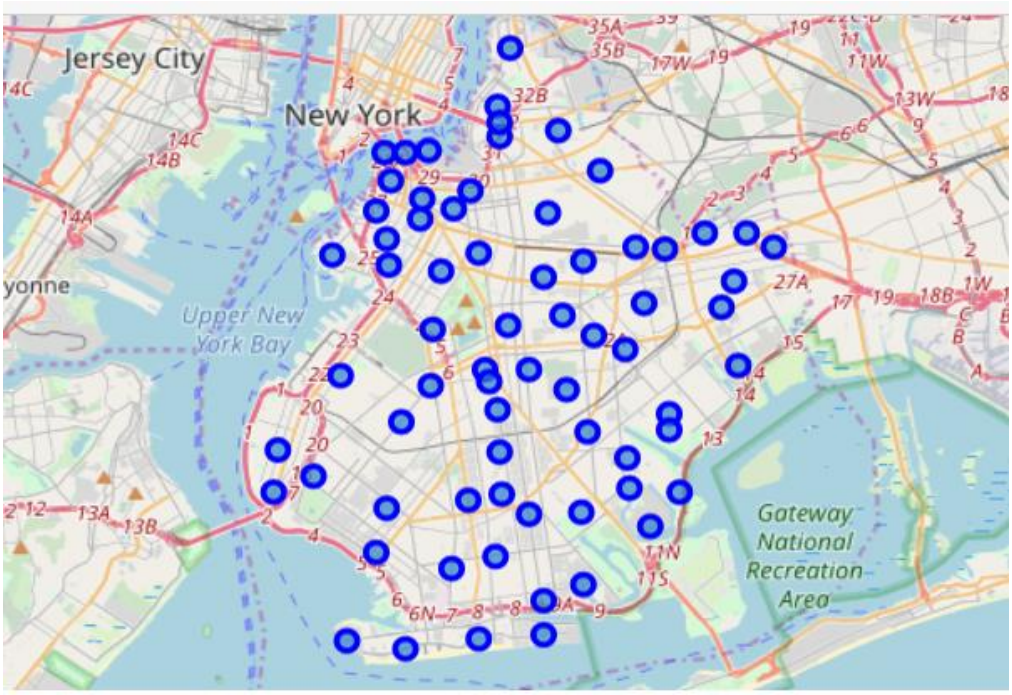

Fig.1: *Brooklyn Map*

Fig.2*: Queens Map*

Neighborhood Exploration and Cluster

The FourSquare API was used for the neighborhood exploration. The get request was deployed on the FourSquare URL to get the category types if venues limiting the number of values to 100 within a 500 radius. In this study, it is going to determine the cluster of venues in the neighborhoods, one-hot encoding was performed on the venue categories to get dummies for each venue. Coding venues into 0s and 1s, the result was grouped by neighborhood by taking the mean of the frequency of occurrence of each category.

Cluster of Neighborhoods

The k-means cluster was performed for the clustering of venues categories in the neighborhoods into 5 clusters. The k-means clustering algorithm is an unsupervised clustering technique searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers in the dataset.

The two assumptions above are presumably the basis of the k-means model.

IV. RESULTS

A. Brooklyn

For Brooklyn city, In Table-1, the variables of the neighborhood, neighborhood latitude, neighborhood longitude, venue, venue latitude, venue longitude and venue category were obtained by using FourSquare API. And 2788 rows and 7 columns were performed. Once we group the unique categories, 289 different categories were found. Similarly, This information is given in Table-2.

Table-1: Neighborhood Information in Brooklyn.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Ho' Brah Taco Joint	40.622960	-74.031371	Taco Place

Table-2: Neighborhood Values by Grouped in Brooklyn

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bath Beach	44	44	44	44	44	44
Bay Ridge	84	84	84	84	84	84
Bedford Stuyvesant	25	25	25	25	25	25
Bensonhurst	32	32	32	32	32	32
Bergen Beach	8	8	8	8	8	8
Boerum Hill	94	94	94	94	94	94
Borough Park	21	21	21	21	21	21
Brighton Beach	45	45	45	45	45	45
Broadway Junction	14	14	14	14	14	14
Brooklyn Heights	100	100	100	100	100	100
Brownsville	19	19	19	19	19	19

For the analysis of the cities, we ranked the most common venues in each neighborhood according to the degree of importance. In this way, it provides detailed information about the place for those who want to move from another country of the world, or to relocate to a new city, or to start a new business. Table-3 gives the top venues for each neighborhood in Brooklyn city. For example, most Italian restaurants, Spa and Pizza Places are observed in Bay Ridge.

Table-3: Top Venues in each neighborhood in Brooklyn.

	Neighborhood	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	Bath Beach	Chinese Restaurant	Pharmacy	Fast Food Restaurant	Donut Shop	Bubble Tea Shop	Italian Restaurant	Pizza Place	Sushi Restaurant	Rental Car Location	Kebab Restaurant
1	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Bar	Chinese Restaurant	Thai Restaurant	American Restaurant	Greek Restaurant	Ice Cream Shop	Bagel Shop
2	Bedford Stuyvesant	Pizza Place	Coffee Shop	Café	Bar	BBQ Joint	Playground	Discount Store	Deli / Bodega	Park	Cocktail Bar
3	Bensonhurst	Chinese Restaurant	Sushi Restaurant	Donut Shop	Italian Restaurant	Ice Cream Shop	Bakery	Factory	Noodle House	Cha Chaan Teng	Spa
4	Bergen Beach	Harbor / Marina	Donut Shop	Hockey Field	Other Repair Shop	Athletics & Sports	Baseball Field	Playground	Flower Shop	Flea Market	Fish Market

Using the data we obtained, we create 5 different clusters. These cluster structures were given in Fig.3. The same colors show the same feature/cluster structure.

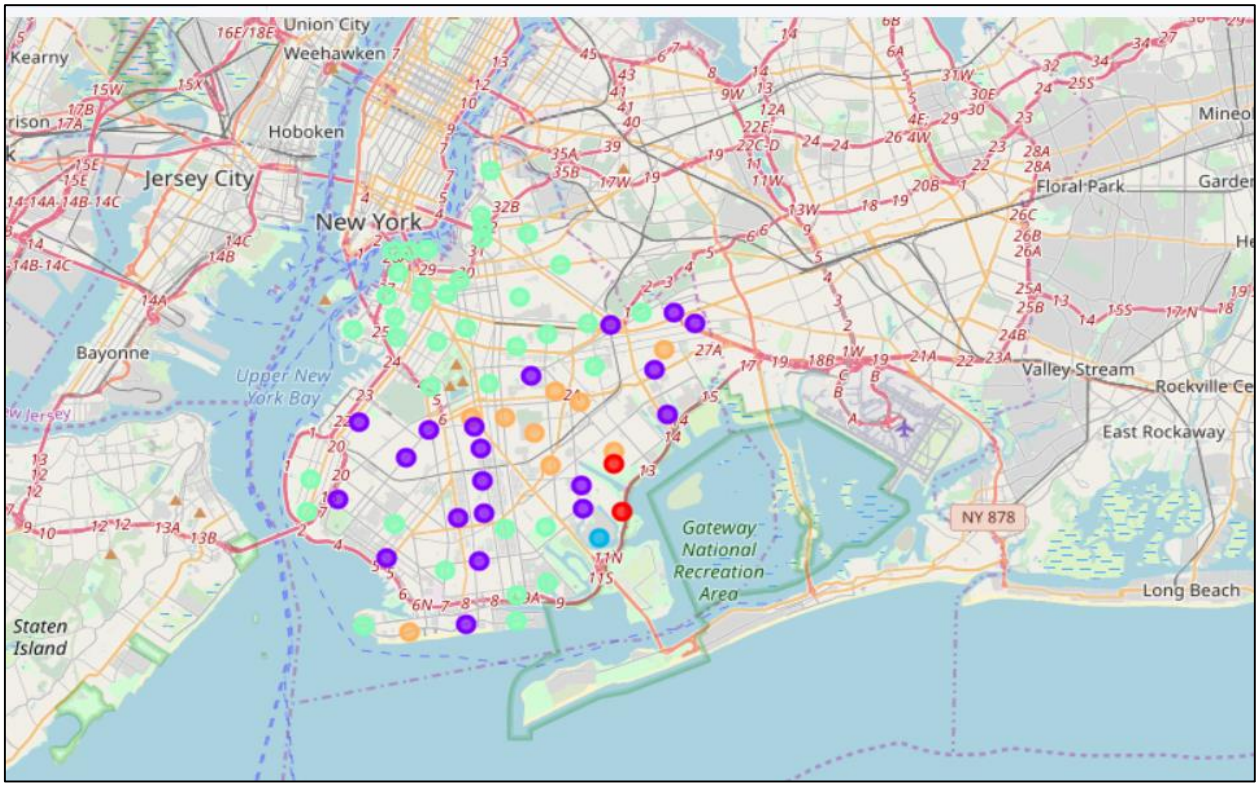

Fig.3: Neighborhood Clusters in Brooklyn.

The most common venues (top three) in each cluster are tabulated in Table-4.

Table 4: Dominant venues in Brooklyn city

Clusters	*1st Common Venues*	*2nd Common Venues*	*3rd Common Venues*
1	Harbor/Marina	Food	Sports Fields
2	Asian Restaurants	Pharmacy	Donut Shop
3	Pool	Women`s Store	Fish&Chips Shop
4	Bar	Coffee Shop	Café
5	Caribbean Restaurant	Grocery Store	Pharmacy

B. Queens

As we did in Brooklyn city, similarly we obtained the data for Queens city. For Queens city, In Table-5, the variables of the neighborhood, neighborhood latitude, neighborhood longitude, venue, venue latitude, venue longitude and venue category were obtained by using FourSquare API. And 2128 rows and 7 columns were executed. Once we group the unique categories, 272 different categories were found. The information is given in Table-6.

Table-5: Neighborhood Information in Queens.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Astoria	40.768509	-73.915654	Favela Grill	40.767348	-73.917897	Brazilian Restaurant
1	Astoria	40.768509	-73.915654	Orange Blossom	40.769856	-73.917012	Gourmet Shop
2	Astoria	40.768509	-73.915654	Titan Foods Inc.	40.769198	-73.919253	Gourmet Shop
3	Astoria	40.768509	-73.915654	CrossFit Queens	40.769404	-73.918977	Gym
4	Astoria	40.768509	-73.915654	Simply Fit Astoria	40.769114	-73.912403	Gym

Table-6: Neighborhood Values by Grouped in Queens.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arverne	17	17	17	17	17	17
Astoria	100	100	100	100	100	100
Astoria Heights	13	13	13	13	13	13
Auburndale	18	18	18	18	18	18
Bay Terrace	37	37	37	37	37	37
Bayside	72	72	72	72	72	72
Bayswater	2	2	2	2	2	2
Beechhurst	13	13	13	13	13	13
Bellaire	13	13	13	13	13	13
Belle Harbor	17	17	17	17	17	17
Bellerose	20	20	20	20	20	20

For the analysis of the cities, we ranked the most common venues in each neighborhood according to the degree of importance. In this way, it provides detailed information about the place for those who want to move from another country of the world, or to relocate to a new city, or to start a new business. Table-7 gives the top venues for each neighborhood in Queens.

Table-7: Top Venues in each neighborhood in Queens.

	Neighborhood	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	Arverne	Surf Spot	Sandwich Place	Metro Station	Wine Shop	Pizza Place	Beach	Playground	Thai Restaurant	Bus Stop	Board Shop
1	Astoria	Bar	Hookah Bar	Middle Eastern Restaurant	Greek Restaurant	Seafood Restaurant	Bakery	Mediterranean Restaurant	Coffee Shop	Café	Italian Restaurant
2	Astoria Heights	Playground	Hostel	Pizza Place	Plaza	Bus Station	Burger Joint	Bowling Alley	Shopping Mall	Business Service	Italian Restaurant
3	Auburndale	Hookah Bar	Furniture / Home Store	Fast Food Restaurant	Athletics & Sports	Toy / Game Store	Noodle House	Train	Discount Store	Mobile Phone Shop	Miscellaneous Shop
4	Bay Terrace	Clothing Store	Women's Store	American Restaurant	Lingerie Store	Donut Shop	Mobile Phone Shop	Cosmetics Shop	Kids Store	Deli / Bodega	Coffee Shop

Using the data we obtained, we create 5 different clusters. These cluster structures were given in Fig.4. The same colors show the same feature/cluster structure.

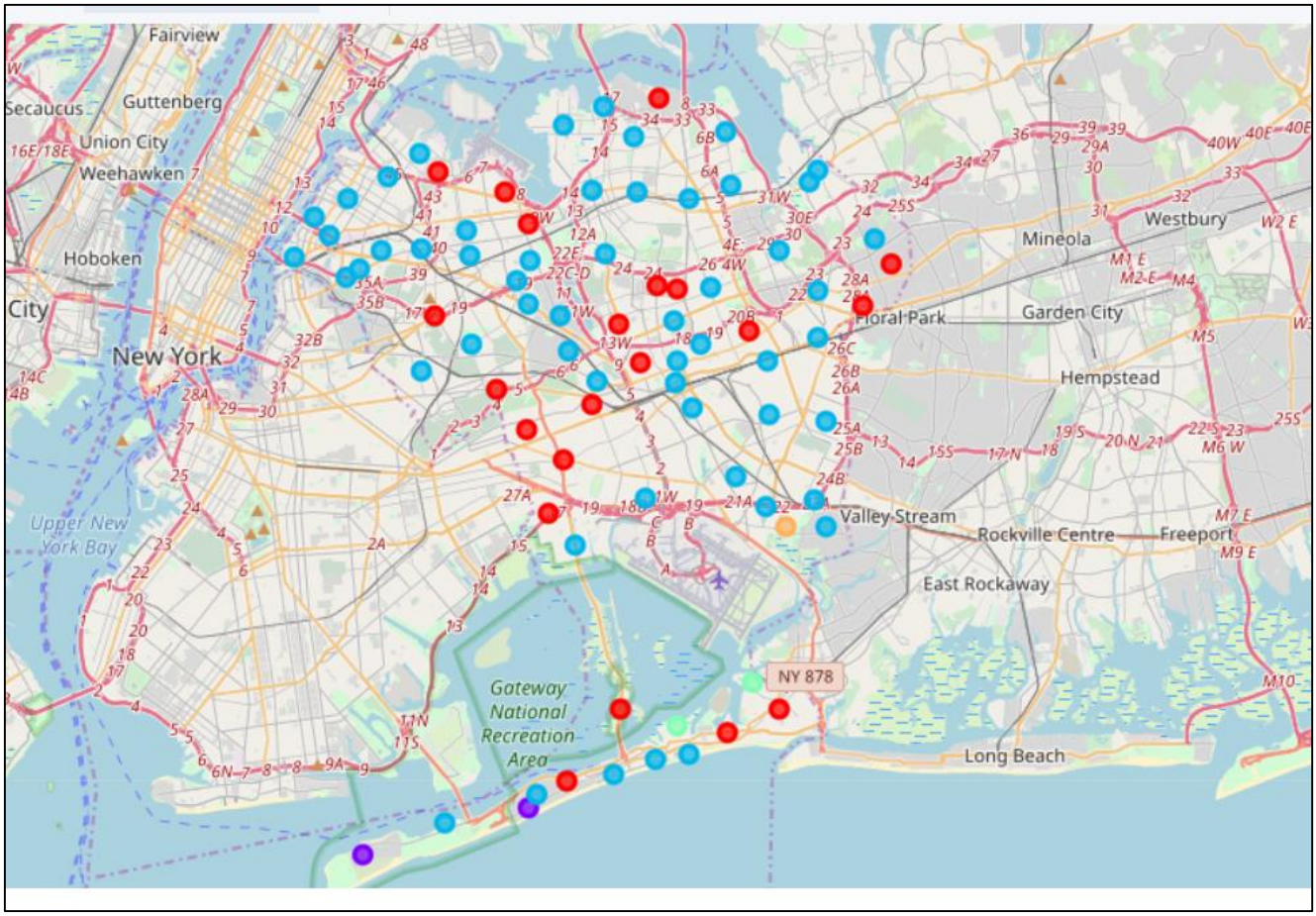

Fig 4: Neighborhood Clusters in Queens.

For Queens city, the most common venues (top three) in each cluster are tabulated in Table-8.

Table 8: Dominant venues in Queens.

Clusters	*1st Common Venues*	*2nd Common Venues*	*3rd Common Venues*
1	Pizza Place	Deli/Bodega	Playground
2	Beach	Women`s Store	Farm Markets
3	Asian Restaurants	Bus Stop	Latin Restaurants
4	Park	Electronics Store	Farm Markets
5	Deli/Bodega	Women`s Store	Filipino Restaurants

Finally, from the data we obtained, we also obtained the most dominant venues in each cluster for Brooklyn and Queens cities. Table-9 provides a key information about the cities.

Table 9: A comparison of dominant venues for Brooklyn and Queens.

CLUSTERS	***BROOKLYN***	***QUEENS***
1	Harbor/Marina	Pizza Place
2	Asian Restaurants	Beach/Women`s Store
3	Pool	Asian Restaurants
4	Bar	Park/Electronics Store
5	Caribbean Restaurant	Deli/Bodega

V. DISCUSSION

In this study, a city analysis was made for Brooklyn and Queens cities that are the two of the five boroughs of New York City. It is observed that similar features are together for each city. In Brooklyn, We have generally Asian restaurants (Chinese, Japanese, indian, sushi etc.), Italian restaurants, pizza places, spa, café, coffee shops, donut shops, pharmacy etc. Based on people's demands, as some neighborhoods are expected, there are many Asian restaurant options in same place.

The role of geographical features in determining the structure of the city is extremely important. We also got some information for Queens city, which has a long coast to the ocean. In Queens, we

have Beach, bar, train / bus stop, café, seafood or Mediterranean restaurants, women`s store, pizza place, farm market etc.

In our daily life, it makes our lives easier to present our needs in different ways in one place. In other words, when we go to Mall or Shopping Center, that there is a lot of different shopping places in one place helps us to see our needs quickly. Likewise, having more than one bar in a bar area is extremely important in terms of providing people with many alternatives for fun/entertainment and providing easy access. It is observed that structures of similar characteristics are together in both city structures.

However, the final decision is left to the individual. But In general, though all these analyses are very helpful and extremely useful, there is nothing like visiting the actual city, seeing the neighborhoods, and speaking with residents and businessmen. If it's possible, an in-person visit is highly recommended before making a big move.

VI. CONCLUSISON

The aim of this study is to provide the necessary information to help people who want to move from another country of the world, or to relocate to a new city, or to start a new business in Brooklyn and Queens cities.

In the analysis, folium-python library was used in building a quick interactive data visualization and FourSquare API for neighborhood data collection. And it is feasible to cluster neighborhoods cities data based on known and accepted machine learning clustering techniques, k-means algorithm. These results will be of interest to people or businessmen whose aim to compare different neighborhoods when thinking about relocation or vacationing in a different environment or start a new business in a new city/ neighborhood, considering the case of accessing numerous venues within a clustered area.

In order to make a good evaluation, it will be incomplete to make an evaluation only according to the venues. In this context, as we mentioned in the introduction, it will be more appropriate to examine the population, the crime rates, the life expense, and the characteristics of schools or neighborhoods. For example, in regions where the crime rates are high, nobody will want to move in or invest in any business because of their security of life and property. Similarly, families with kids will not want to send their children to the low-rated schools, of course, any family will want to send their children to good schools. All these factors have a decisive role in relocation or in establishment of a new business.

In the study, a city analysis was made only from venues in Brooklyn and Queens cities. Despite the shortcomings, considering the different priorities of people, it has an analysis structure giving the different meanings to every segment.

The structures in the cities, we consider from this point of view, have an effective role in people's decision to move in or start a new business in terms of the combination of similar structures.

References

[1] Mike Maciag (October 2, 2013). "Mapping the Nation's Most Densely Populated Cities". Governing – The States and Localities. Archived from the original on August 27, 2016. Retrieved August 27, 2016.

[2] https://en.wikipedia.org/wiki/New_York_City

[3] Immerso, Michael (2002). Coney Island: The People's Playground. Rutgers University Press. p. 3. ISBN 978-0-8135-3138-0.

[4] O'Donnell, Michelle (July 4, 2006). "In Queens, It's the Glorious 4th, and 6th, and 16th, and 25th ..." The New York Times. Retrieved July 20, 2014.