

Appropriate Neighbourhood for Immigrants

Introduction

Background:

New York is often considered one of the greatest cultural hubs in the country, having welcomed immigrants long before the days of Ellis Island. Immigrants now account for one-fifth of the state's total population and make up a staggering 25 percent of its labor force. As workers, business owners, taxpayers, and neighbors, immigrants are an integral part of New York's diverse and thriving communities and make extensive contributions that benefit all.

One in five New Yorkers is an immigrant, while one in six is a native-born U.S. citizen with at least one immigrant parent.

Business Problem:

It is often imperative that the immigrants need help in trying to find an appropriate neighborhood at least during the initial years of settlement. Current means of selection is solely based on contact(s) they may have in New York or based on employer provided information, which in most cases may not suffice.

There could be some real estate agents or immigration helping services that could provide such information and would charge a good amount of fees for it.

Business Solution proposal:

Idea is to provide an online service which would help the would-be immigrants to explore and choose their neighborhood much before their arrival in New York. Immigrants include people from Asia, Latin America, Europe e.t.c

Plan of action:

To validate the stated problem and solution, we could prototype a solution focused on New York and based on its success, expand it to other cities in the USA.

Data Description

Data Source:

- Information about Neighborhoods in New York would be fetched from NYU Spatial Data Repository available at <https://geo.nyu.edu> that includes Latitude and Longitude along with Borough's of the City.
- Foursquare API (<https://forusquare.com>) would be used to gather most commonly available venues in and around the Borough/Neighborhood based on the Latitude and Longitude obtained from the NYU Spatial Data Repository.

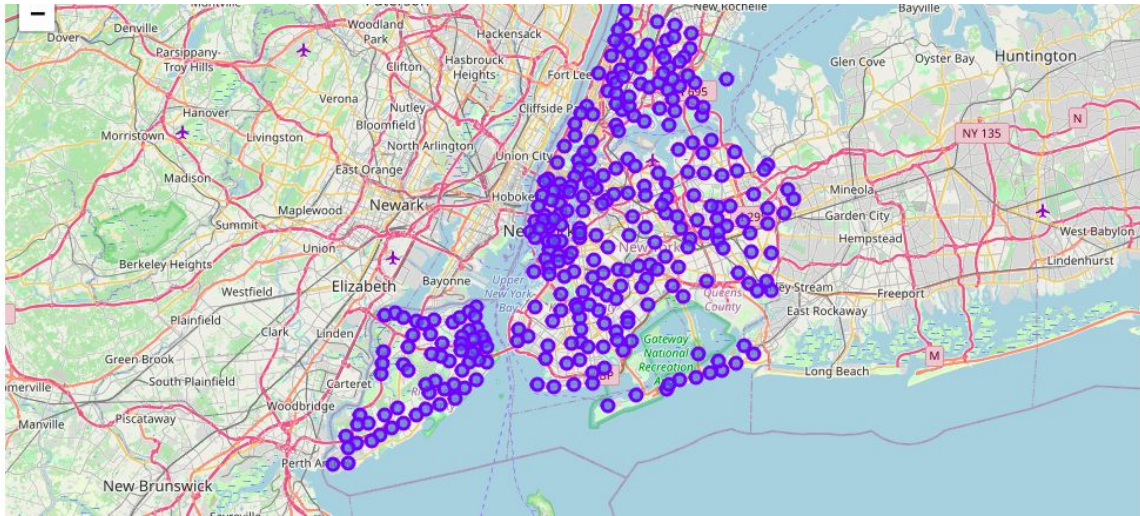
Data Usage:

- The data thus fetched would be analysed, explored and grouped based on multiple criteria and the results would be presented to the user both in maps format and in information data tables format to help choose the right Neighborhood according to their Nativity and/or Preference.
- For example, areas surrounded by Chinese Restaurants and Buddhist temples would be of consideration to Immigrants from China, Indian Restaurants, Temples of both Hindu and Buddhist would be of interest to people from India, SriLanka and Bangladesh.
- Furthermore, Irrespective of the origin Country, Students coming to pursue Higher Education may find Neighborhoods with Universities, Libraries, Parks and Entertainment places to be of Interest.

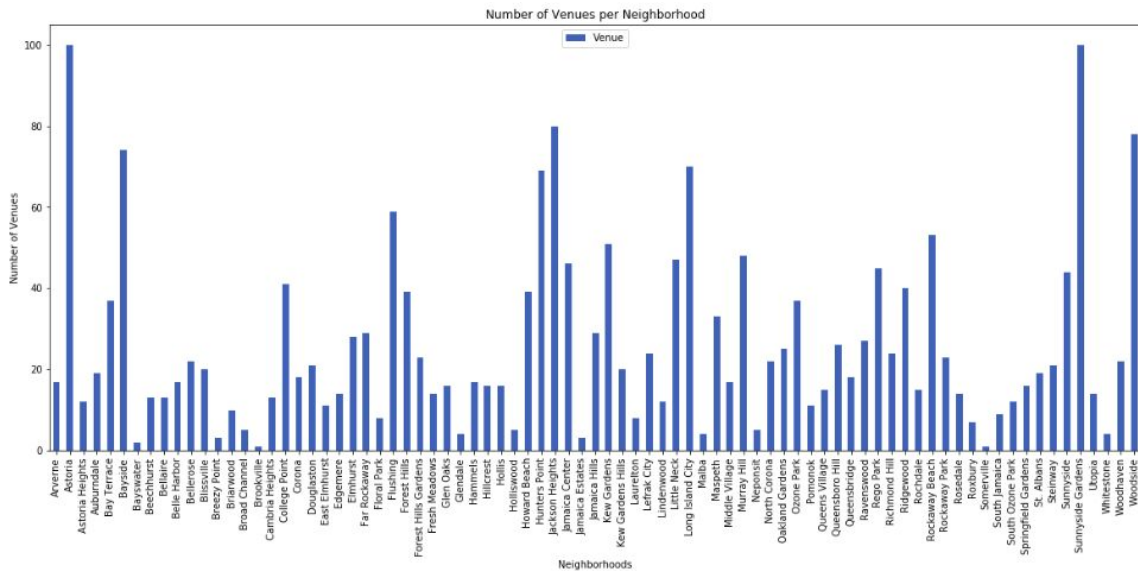
Methodology

The Data that we thus gathered from <https://geo.nyu.edu> has 5 boroughs and 306 Neighbourhoods, out of which "queens" borough have 80 neighbourhoods. We shall therefore consider "queens" for our exploratory analysis.

Visualizing all those neighbourhoods in those 5 boroughs...



Foursquare API is used to gather all the venues available in the “queens” Borough’s Neighbourhoods.
 Plotting the Neighborhoods in “Queens” and their venues we get to see the below visual



We shall reduce our Neighbourhoods to consider from the 80 above to a lesser number based on the number of venues available. Considering the fact that most of the people prefer to live in a place that has less activity we shall thus reduce our dataset to contain Neighbourhoods will less than 20 venues in and around them.

The Neighbourhoods thus available to us are narrowed down to

```
In [80]: quiet_nh['Neighborhood'].unique()
```

```
Out[80]: array(['Corona', 'East Elmhurst', 'Glendale', 'South Ozone Park',  
               'Whitestone', 'Auburndale', 'Glen Oaks', 'Fresh Meadows',  
               'Briarwood', 'Queens Village', 'Hollis', 'South Jamaica',  
               'St. Albans', 'Rochdale', 'Springfield Gardens', 'Cambria Heights',  
               'Rosedale', 'Broad Channel', 'Breezy Point', 'Beechhurst',  
               'Edgemere', 'Arverne', 'Neponsit', 'Floral Park', 'Holliswood',  
               'Jamaica Estates', 'Hillcrest', 'Lindenwood', 'Laurelton',  
               'Belle Harbor', 'Somerville', 'Brookville', 'Bellaire', 'Utopia',  
               'Pomonok', 'Astoria Heights', 'Roxbury', 'Middle Village', 'Malba',  
               'Hammels', 'Bayswater', 'Queensbridge'], dtype=object)
```

Looking at the top 10 most Venues in the above Neighbourhoods, we see the following...

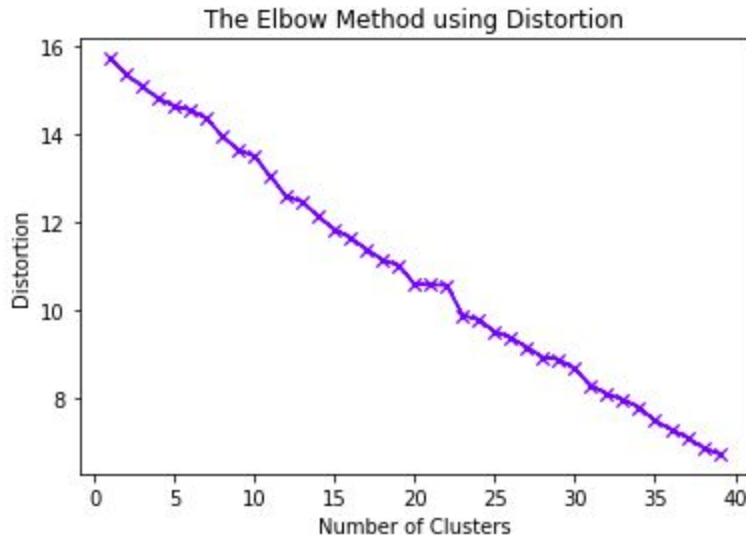
```
OUT[8/]:
```

	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	Metro Station	Sandwich Place	Board Shop	Thai Restaurant	Coffee Shop	Donut Shop	Beach	Bus Stop	Playground
1	Astoria	Bar	Middle Eastern Restaurant	Hookah Bar	Greek Restaurant	Seafood Restaurant	Indian Restaurant	Mediterranean Restaurant	Pizza Place	Food Truck	Bubble Tea Shop
2	Astoria Heights	Deli / Bodega	Playground	Pizza Place	Plaza	Italian Restaurant	Hostel	Bus Station	Burger Joint	Bowling Alley	Bakery
3	Auburndale	Italian Restaurant	Deli / Bodega	Discount Store	Supermarket	Bar	Furniture / Home Store	Korean Restaurant	Fast Food Restaurant	Athletics & Sports	Noodle House
4	Bay Terrace	Clothing Store	Women's Store	Cosmetics Shop	Donut Shop	American Restaurant	Lingerie Store	Shoe Store	Kids Store	Mobile Phone Shop	Furniture / Home Store

We do have some common venue categories in those neighbourhoods. For example Deli/Bodega, Chinese Restaurant e.t.c.

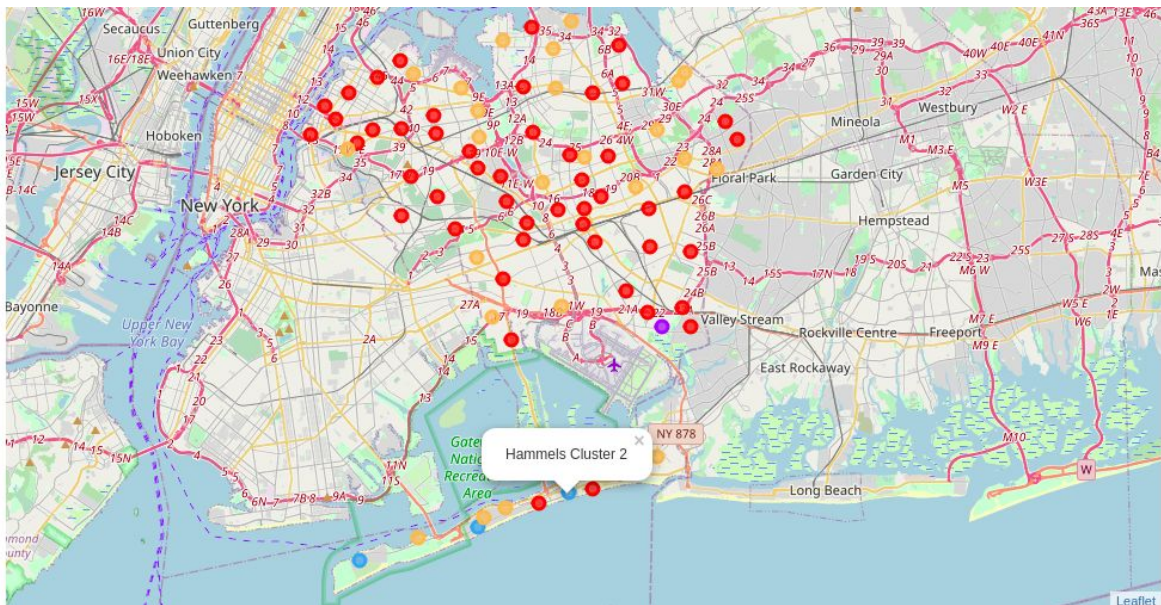
We shall use K-means unsupervised learning to cluster those neighbourhoods. Before we cluster the Neighbourhoods, we need see if we can find an optimum value for the number of clusters that we would want to split our information on.

Elbow method is the most common one used to find an optimal number for finding clustering number. Implementing it we find that there not much of help that we get from it, partly because the neighbourhoods are so vibrant and probably really cosmopolitan, accommodating variety of Immigrant cultures with no specific pattern.



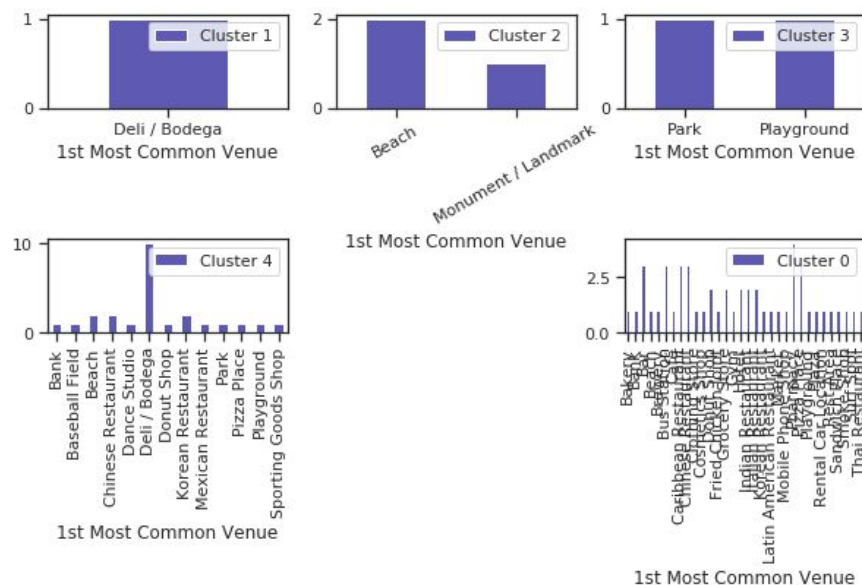
Hence, we shall use an arbitrary number 5 to cluster our neighbourhoods based on the Venue Categories.

Choropleth map visualizations were used to picture the thus obtained Clusters.



Results

In the results section, our goal is to project all the clusters thus obtained to present the type of venues (top 10) in each of those cluster thus helping to find an appropriate Neighborhood for consideration.



As we can see, from the picture above, Cluster 0 and Cluster 4 contain Neighborhoods that have quite a few number of Venues, while Cluster 1,2,3 are very calm and with less number of venues. They must be mostly visiting places and not really action packed.

Also by examining the Venues in all the clusters, we cannot draw any conclusion as to which cluster is good for what kind of immigrants as most of the venues there are non-specific and those clusters seem to host all kinds of restaurants, (i.e. from different continents/countries) and are generic in nature.

However, we could broadly classify Cluster 4 and Cluster 0 for Immigrants.

Cluster 0 seems to have more presence from latin American Countries with Sporadic presence from China, India and Korea.

Cluster 4 seems to have more presence from Asian & Mediterranean Countries with sporadic presence from Latin American countries.

Discussion

As we have been observing, New York is a Big City with 306 Neighbourhoods, high population density and miraid of cultures living together not in a very specific pattern all across.

It seems to be such a complex amalgamation of Natives and Immigrants not a single Clustering & Classification technique can be used to obtain satisfactory results.

We tried to use the Elbow method to find an appropriate cluster number to be used with K-means cluster algorithm and were less successful.

There is lot of information available to us considering the Neighbourhood location co-ordinates and the venues/Venue Categories that we were able to access with Foursquare API. More in depth study should help considering not just the top 10 venues but all of the available venues/Venue Categories, Addresses of them, also by considering the amount of traction in each of those venues.

I could broadly cluster the neighbourhoods into Visiting places and vibrant places in our clustering method adopted and presented the results in both data format and visual format, for now.

Here with I conclude.

Conclusion

This Problem is definitely of potential to be solved which would be of great help to all those Migrants who come to the city either for Business or work.

We could succeed to some extent of classifying the Neighborhoods into Vibrant and quiet places and also generalize the Clustered attributes.

Further studies could narrow down to more cluster with stringent information and additional features to be able to classify at least to the level of Country specific congruence.