

The Battle of Neighborhood(New York)

Find optimal location for most common venue which is not already
crowded

Introduction: Business Problem

- Since there are lots of variety of venues in New York, we will try to detect **most popular venue and location that are not already crowded with that venue**.
- We would also prefer locations **as close to city as possible that covers most nearby neighbourhoods**, assuming that first two conditions are met.

Data

Below mentioned factors will influence my decision:

- Variety of venues in the neighbourhood.
- Number of different venue type in each Borough.
- Number of neighbourhoods in each Borough that is not already crowded with most popular venue.
- Location of each uncrowded neighbourhood to cluster them and find out highly dense cluster.

I will be using the following data sources to extract the required information:

- Raw data for New York City will be downloaded from external source.
- Borough, neighbourhood and their latitude and longitude data will be extracted from downloaded data.
- Number of venues, venue category and corresponding latitude and longitude of each venue in every neighbourhood will be obtained using **Foursquare API**.

Map of New York with each Borough & neighbourhoods

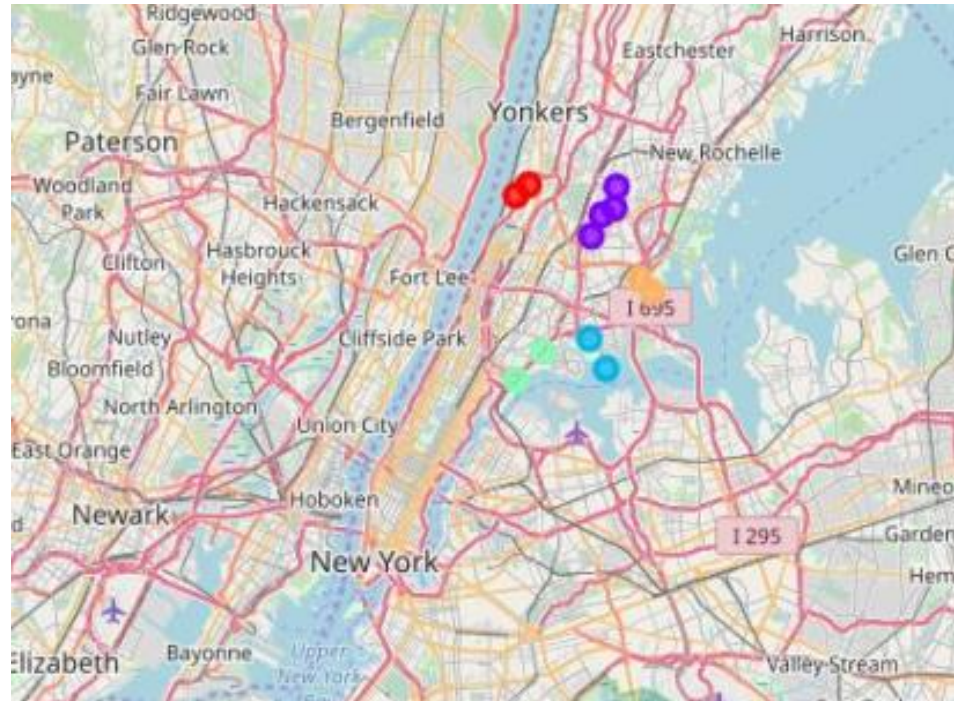


Analysis of each Borough

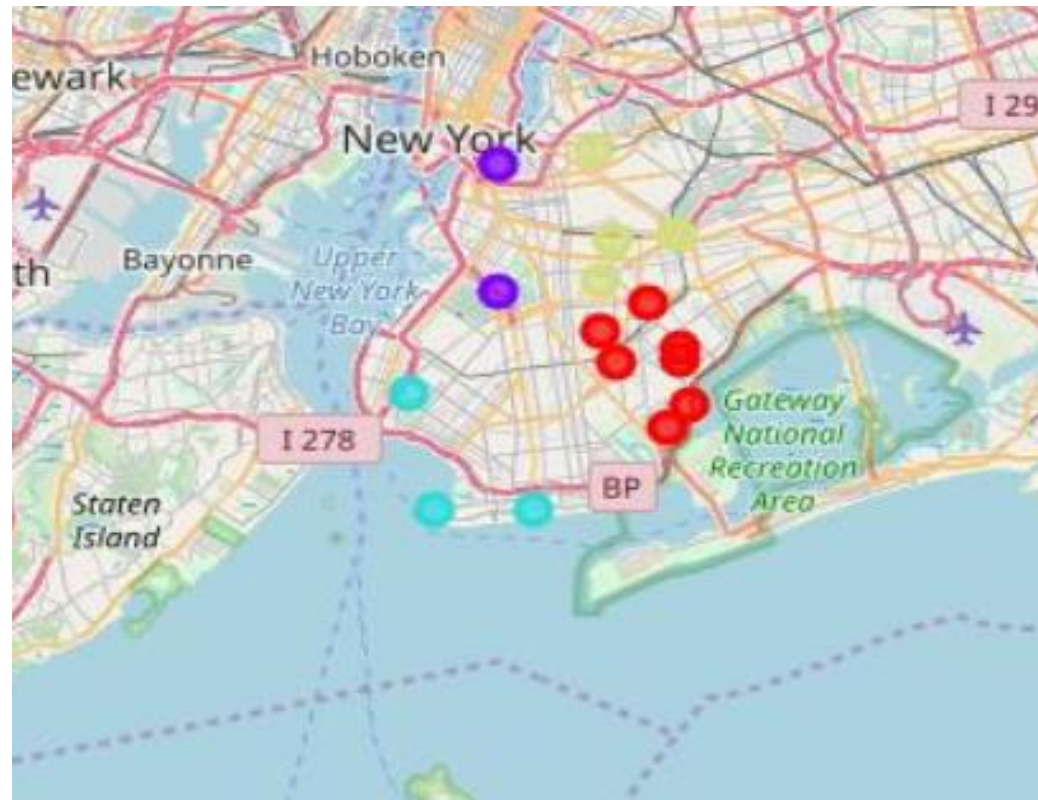
	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronx	Pizza Place	Deli / Bodega	Donut Shop	Pharmacy	Supermarket
1	Brooklyn	Pizza Place	Coffee Shop	Bar	Bakery	Deli / Bodega
2	Manhattan	Coffee Shop	Italian Restaurant	American Restaurant	Café	Pizza Place
3	Queens	Pizza Place	Deli / Bodega	Chinese Restaurant	Bakery	Donut Shop
4	Staten Island	Bus Stop	Pizza Place	Italian Restaurant	Deli / Bodega	Bagel Shop

Clustering each Borough

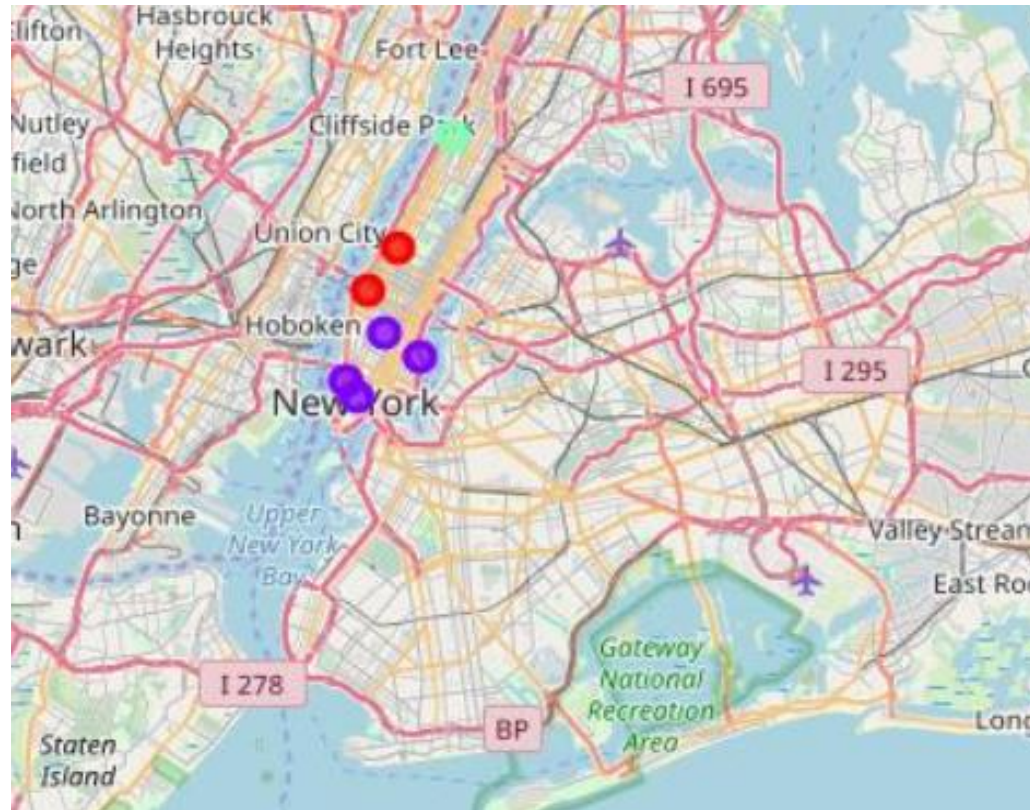
- Cluster Analysis for Bronx



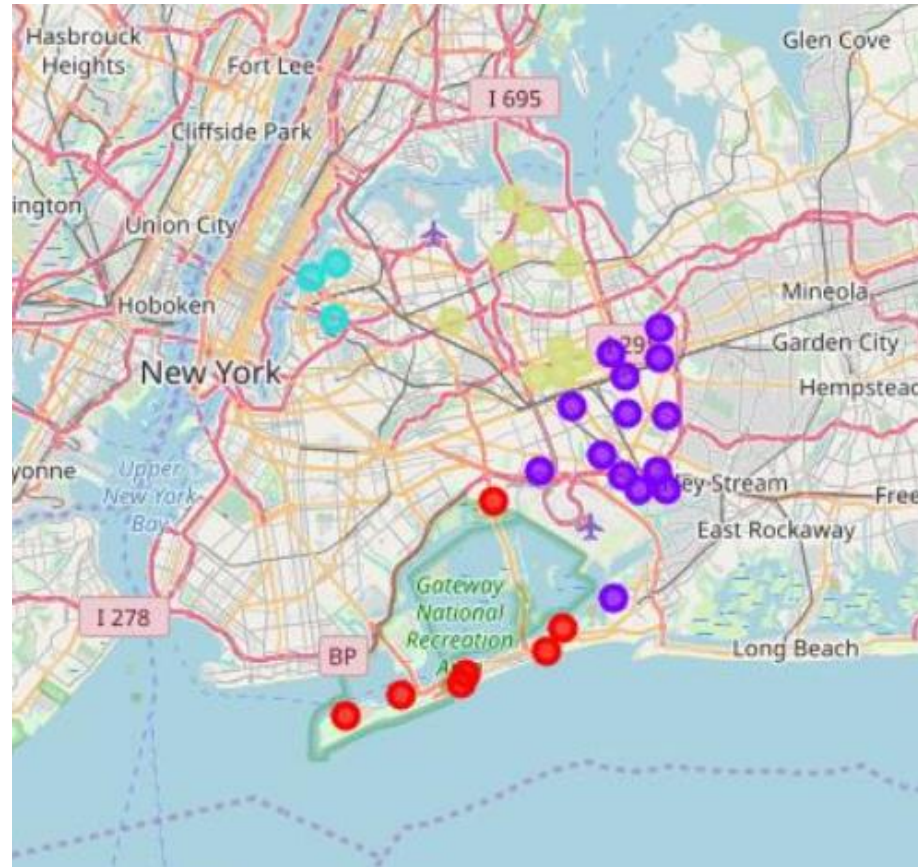
- Cluster Analysis for Brooklyn



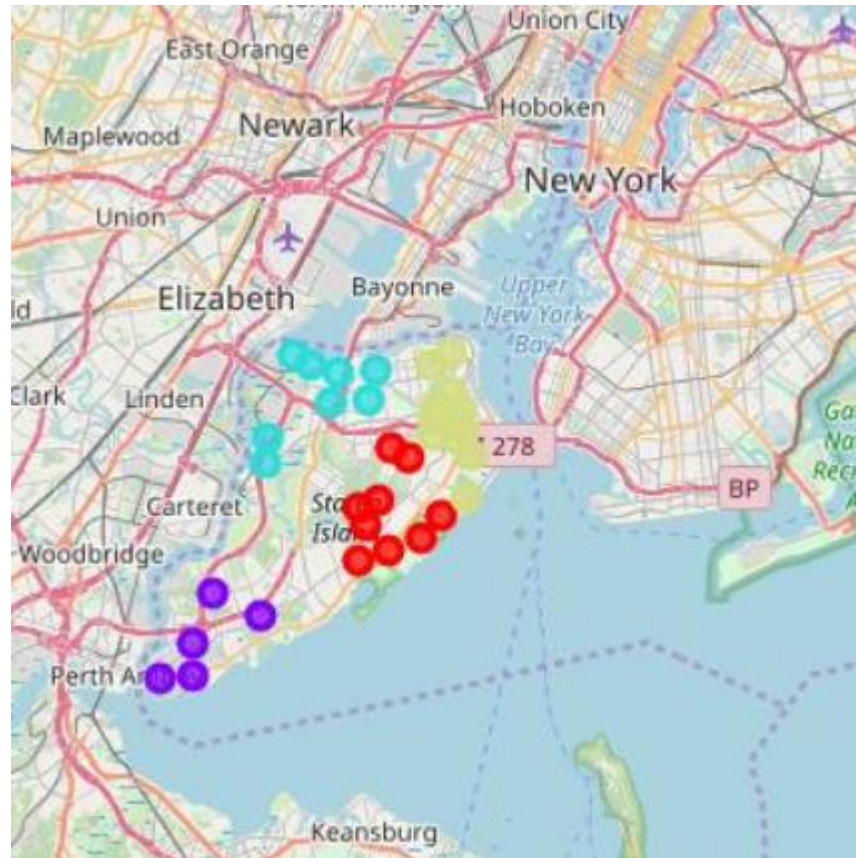
- Cluster Analysis for Manhattan



- Cluster Analysis for Queens



- Cluster Analysis for Staten Island

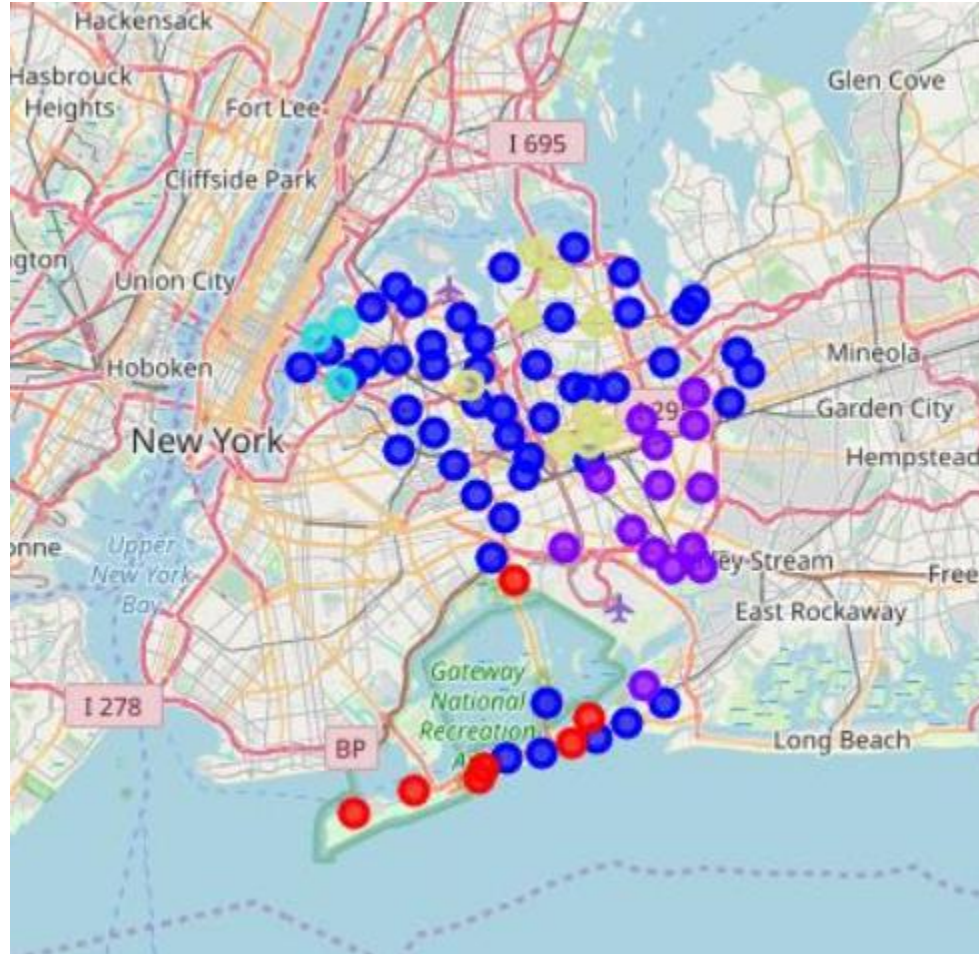


Comparing clusters in each Borough

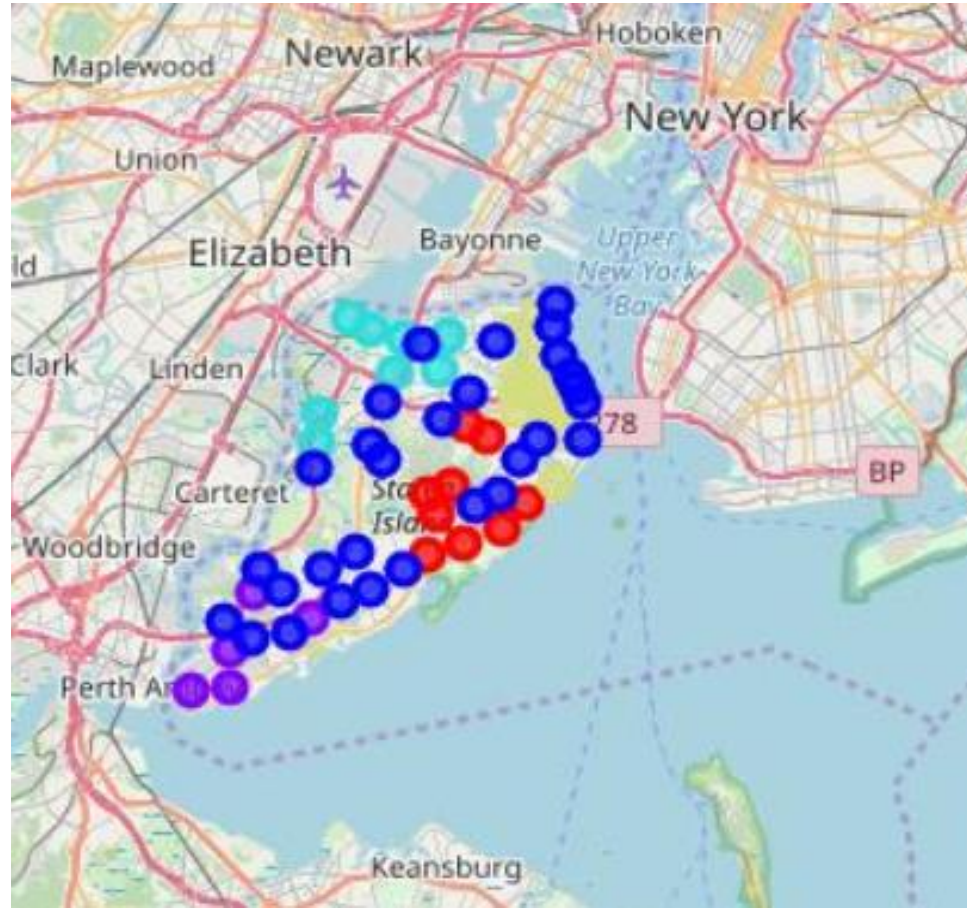
From the above visual representations we can see that **Queens and Staten Island** both are not much crowded with pizza places. In Queens' **cluster 1** consists majority of neighbourhoods so it can be considered for optimal location. Similarly in Staten Island we can see that all clusters are highly dense and **cluster 3** is more appropriate cluster to be considered for optimal location.

Let's see which cluster is optimal for consideration by adding neighbourhoods which are having pizza place in Queens and Staten Island. In below figures blue circle represents neighbourhoods which are having pizza places.

Queens Cluster with pizza places



Staten Island cluster with pizza places

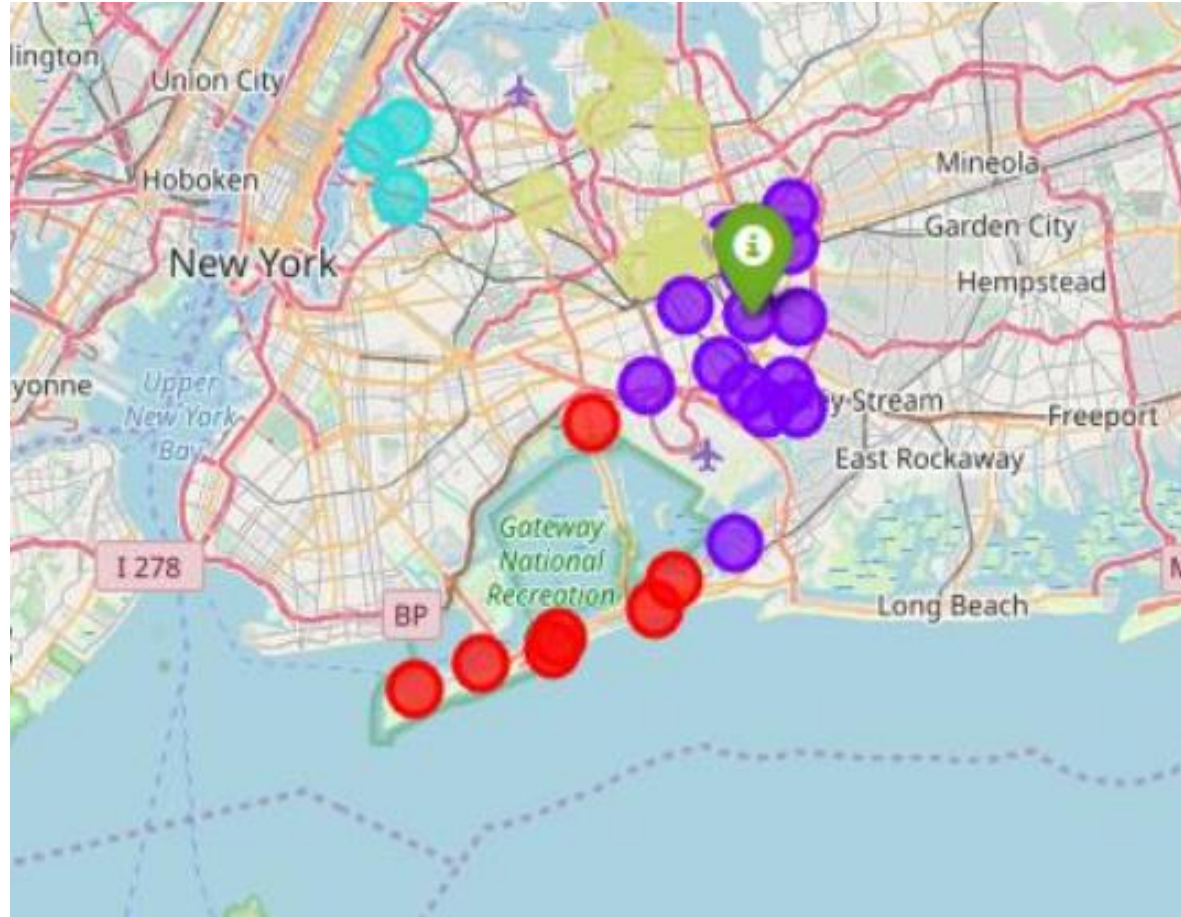


Optimal location

Here we can clearly see that in Staten Island all clusters are more closely surrounded by pizza place. While in Queens' cluster 1 is not as much crowded with pizza place. So we can consider **Queens' cluster 1** for optimal location.

Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Pizza Place	Cluster Labels	Distance
Queens	St. Albans	40.694445	-73.758676	0	1	0.714652

Visualization of optimal location



Results and Discussion

As our analysis and graphical visualization shows that there are number of different venues in each borough. But we have seen that each borough has mostly crowded with pizza places. And in each borough, there are neighbourhoods with pizza place and non-pizza place. Their ratio are mostly fair in each borough but geographical location is different. In some borough non pizza places are surrounded with pizza places and in some borough they are isolated.

For optimal location we first calculate the centre of that cluster and then calculate the distance of each neighbourhood in that cluster from centre. We sort the distance to find the closest location from centre which was **St. Albans** in **Queens borough**. This recommended location should therefore be considered only as a starting point. There may be other factors taken into account and other conditions may be introduced for optimal location.

Conclusion

Purpose of this capstone project was to identify an optimal location that is not crowded by most common venue in New York City. It gives stakeholders as a starting point for consideration. We calculated the number of occurrence of each venues in each borough which gives an idea about most common or popular venue in each borough. We found that pizza place is most common among all borough.

Then clustering of those neighbourhoods where there were no pizza place gave us a zone of interest that meet some basic conditions.

We considered the cluster of neighbourhoods which was not already crowded/surrounded by pizza places.

Final decision on optimal location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended cluster of neighbourhood, taking other factors and conditions into consideration.