Predicting the best Neighborhood and Perfect location for a Restaurant

Akande Adebambo

December 18, 2020

**1. Introduction**

A friend wants to open restaurant but doesn't know which one of these two Neighborhoods, the Brooklyn and the Downtown in the New York City and the Toronto City respectively, is better to open or site a company or business. My friend is faced with the problem of how to determine the right Neighborhood and perfect location for his business.

To choose the restaurant, we will need to access data of Neighborhoods in [New York](https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City) (Brooklyn) And [Toronto](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto) to build our analysis. We will also access coordinates latitude and longitude of the Localities of these two cities from Foursquare APIs and also identify our venues locations with maps. From Foursquare we will need different restaurant information. We will then leverage the data in order to determine which neighborhood is the most appropriate for the restaurant.

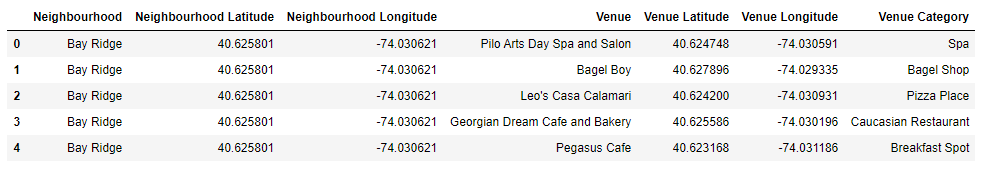
**2. Source of Data**

* Demographics of New York City from U.S. Census of 2000. Demographics of New York City from Wikipedia
* Demographics of Toronto Neighborhoods from Wikipedia 2006
* Foursquare API Geolocation data [Foursquare-Developer-API](https://foursquare.com/)

### Description of Data

The attributes of the data includes:

* identity numbers,
* longitudes,
* latitudes,
* boroughs,
* Neighborhoods,
* venue,
* categories,
* population,
* average income,
* gross domestic product,etc.



The latitude is the line running east and west of the earth measured in numeric while the Longitude is an imaginary line running north and south of Greenwich meridian also measured in numeric. The borough is a subset of a city. The neighborhood is the community area name. The target value (variable) or label of the data set is the neighborhood which is the dependent variable. The other variables are longitudes, latitudes, venues, etc are the independent (predictor) variables.

## 3. Methodology

### Data Analysis of the Brooklyn and the Downtown

* The data sets of the two neighborhoods, Downtown and Brooklyn were visualized to gain knowledge of the distribution of variables and the neighborhoods in the Boroughs, the Downtown and Brooklyn, using folium library to see how the neighborhoods in this borough are spatially distributed.



Exploratory data analysis of Brooklyn

* The resulting data is subjected to initial step of data analysis and the exploratory data analysis (EDA) to gain insight on how correlated or trending the data are. After the data sets have been subjected to exploratory data analysis using Folium, Seaborn and Matplotlib libraries to see how the variables of the data are distributed, trending and insight gained, the data was then subjected to pre-processing stage.

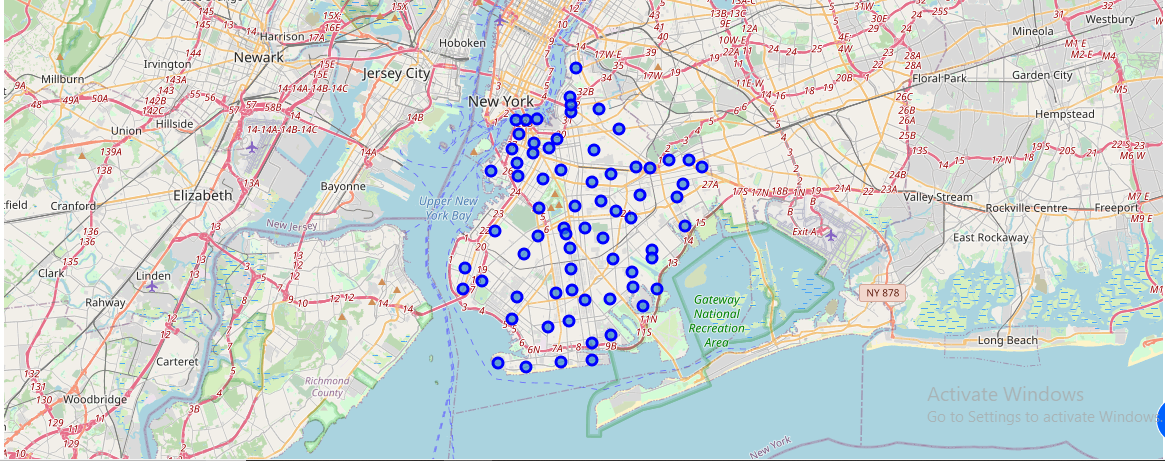
### Inferential Statistical Test (Machine Learning) of the Brooklyn and the Downtown

* Here the encoded data during the analysis are fed into the machine learning algorithm using K-mean Clustering to segment the neighborhoods. The neighborhoods that have similar characteristics are grouped together. Therefore the similarities and dissimilarities of the Downtown and the Brooklyn are observed.

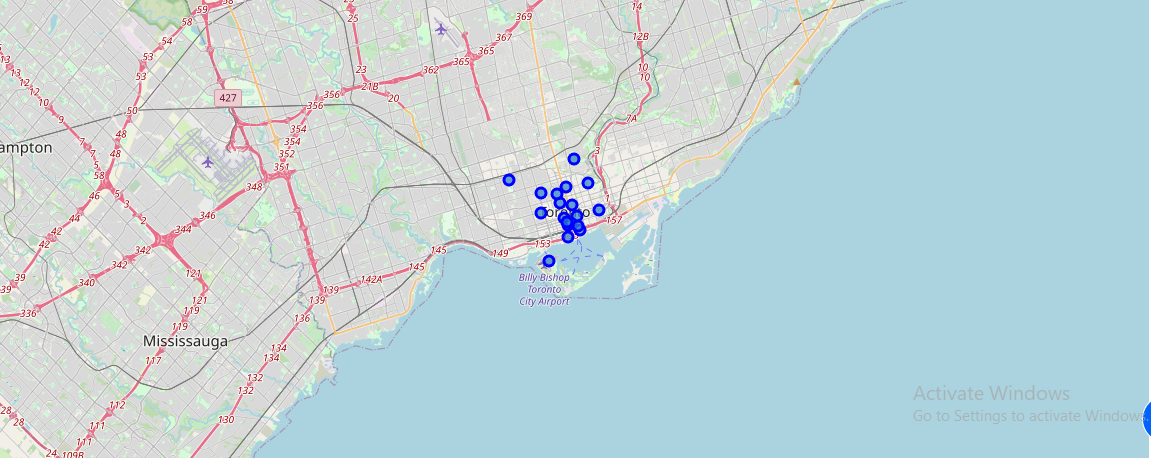


## 4. Discussion

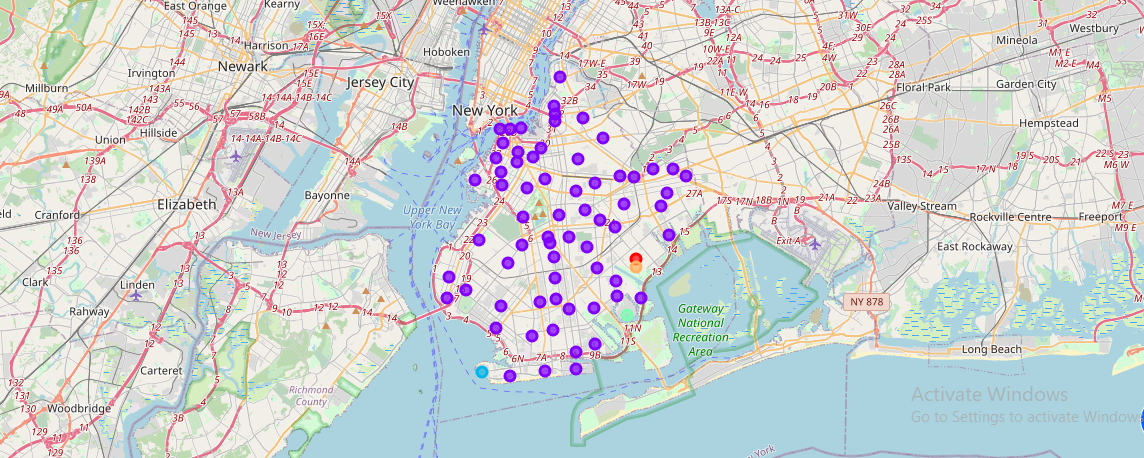
* The two neighborhoods in the Brooklyn of the New York City and the Downtown Toronto of the Toronto City are similar in having banks, grocery, market, farmers market, waterfall, college, transport station and buildings but dissimilar in other venues



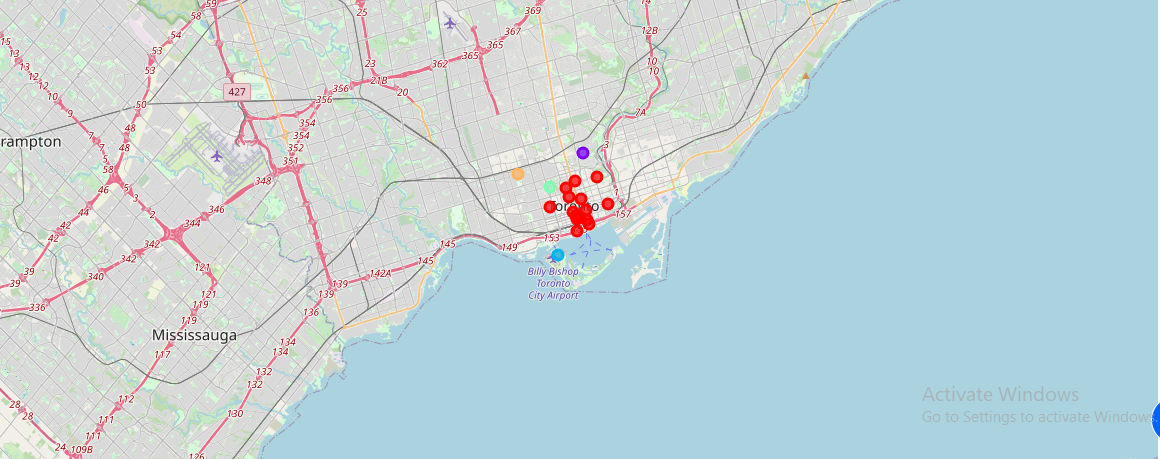
Map of Brooklyn using latitude and longitudm values



Map of Downtown Toronto using latitude and longitude values



Brooklyn map cluster



Downtown map cluster

* These venues that are similar indirectly represent availability of capital made available by bank, raw materials provided by the farmers markets and groceries, water provided by waterfall, labour provided by the colleges and the universities, efficient management provided by the college and the competitors which are in the restaurant businesses in the neighborhoods, transport facilities provided by the presence of bus or train stations in the neighborhoods.

### 5. Recommendations

Although there are factors or variables that are not available during this project, the project still recommend that the Brooklyn neighborhoods should be considered in siting a restaurant business of the two boroughs given to be considered are the Brooklyn and the Downtown.

## 6. Conclusion

* In the end, the Brooklyn neighborhoods are better than the Downtown neighborhoods in siting a restaurant business based on the available data. The two neighborhoods are similar in factor that influences the location of restaurant business yet they are glaringly dissimilar in the population densities and hence the market shares of a potential restaurant business.
* Any potential restaurant businesses sited in the Brooklyn will have large market share because of the population or population density of the area than if it is to be sited in the Downtown.