**The Battle of Neighborhoods: New York Vs Toronto**

**Table of Contents**

* Introduction
* Data
* Methodology
* Result
* Discussion
* Conclusion

## INTRODUCTION

### Business Problems

* George is a chef from England, and he wants to expand his restaurant chain which he started with his friends three years ago. He always wanted to do business in the land of opportunities North America, so he finally comes down to two cities New York and Toronto as he has some contacts in both the places. He gets started with his research for comparing how similar or dissimilar are both the cities. This would help him in taking informed decision and help him in expanding his business effectively.
* Hence the problem is how to determine the right neighborhood for George to open a restaurant.

### Audiences

The apt audience for this work is the George and his friends who may want to start up a business in the neighborhoods.

The project or solution is important to the stakeholders and other audiences because it provides the best neighborhood to be considered in citing a business in the city. Besides, it helps the audience or stakeholders in making or reaching decision quickly.

## DATA

## Source of Data

Firstly, he needs to fetch and explore neighborhoods data of those two cities from Foursquare API. The information he wants to focus on are restaurants, coffee shops and entertainment venues. We will choose one neighborhood from each city: Manhattan for New York and Downtown for Toronto. We need to apply Neighborhood Segmentation and Clustering to analyzing the neighborhood data and prioritize the best restaurant location in both cities based on foot traffic and type of restaurants. Lastly, we can decide which city is the better place for George to get start.:

* [Demographics of New York City from U.S. Census of 2000 and the New York Department of City Planning](https://www1.nyc.gov/site/doh/data/health-tools/neighborhood-statistics-demographics.page)
* [Demographics of New York City from wikipedia](https://en.wikipedia.org/wiki/Demographics_of_New_York_City)
* [List of Postal Codes of Canada from wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
* [Boroughs of Nework City from wikipedia](https://en.wikipedia.org/wiki/Buroughs_of_New_York_City)
* [Demographics of Toronto Neighbourhoods from wikipedia which is in turn taken from 2006 Canadian Census](https://en.wikipedia.org/wiki/Demographics_of_Toronto_Neighbourhoods)
* [Foursquare API Geolocation data](https://foursquare.com/)

### Description of Data

The attributes or characteristics of the data are:

* identity numbers,
* longitudes,
* latitudes,
* boroughs,
* neighbourhoods,
* venue,
* categories,
* population,
* average income,
* gross domestic product,
* tips 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 Green Wich meridian also measured in numeric. The borough is a subset of a city. The neighbourhood is the community area name.

The target value (variable) or label of the data set is the neighbourhood which is the dependent variable. The other variables is like longitudes, latitudes, venues, etc are the independent (predictor) variables.

## METHODOLOGY

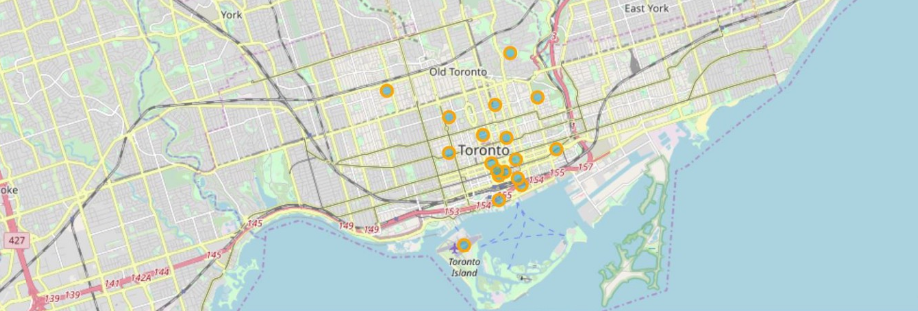
### Data Analysis (Exploratory Data Analysis) of the Manhattan and the Downtown

* The data sets of the two neighbourhoods, the Downtown and the Manhattan were visualized to gain knowledge of the distribution of variables and the neighbourhoods in the Boroughs, the Downtown and the Manhattan, using folium library to see how the neighbourhoods in this borough are spatially distributed.
* The resulting data is subjected to initial step of data analysis – 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.

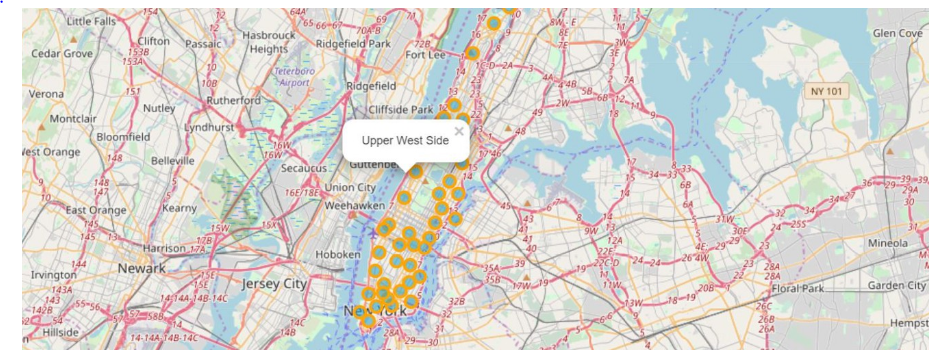
### Inferential Statistical Test (Machine Learning) of the Manhattan and the Downtown

* Here the encoded data during the analysis are fed into the machine learning algorithm - K-mean Clustering - to group or segment the neighbourhoods.
* At the end, the neighbourhoods that have similar characteristics are grouped together. Hence the similarities and dissimilarities of the Downtown and the Manhattan are observed.

RESULTS

Toronto Neighborhoods

Number of venues in downtown-207



New York Neighborhoods

Number of venues in manhatten-341

## DISCUSSION

### Observations

* The two neighborhoods in the Manhattan 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 listed above and in the table 2.5 shown above in the result section.
* 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, labor 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.

### Recommendations

Even though there are factors or variables that are not available during the research, the research still recommend that the Manhattan neighborhoods should be considered in siting a restaurant business if the two boroughs given to be considered are the Manhattan and the Downtown.

## CONCLUSION

* In the end, the Manhattan neighborhoods are better than the Downtown neighborhoods in siting a restaurant based on the available data. because even though the two neighborhoods are similar in factor that influence the location of restaurant business, yet they are very dissimilar in the population densities and hence the market shares of a potential restaurant business.
* Any potential restaurant businesses sited in the Manhattan 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.