

# **The Battle of the Neighborhoods: Scarborough, Toronto vs Brooklyn, New York**

**By Milton Suggs**

## **INTRODUCTION/BUSINESS PROBLEM**

The Horizon Corporation is mid-size company with approximately 350 employees and is growing rapidly. It is a company that specializes in developing educational curriculum, tools, software, and methodology across a wide variety of mediums and disciplines. The company is looking to expand by opening a second North American office. The stakeholders of the company have narrowed down this new location to either Brooklyn, New York or Scarborough in Toronto, CA.

Because about 150 employees will be relocating to open the new location, the company is looking to gain insights about the surrounding businesses and venues in Brooklyn and Scarborough to determine which location will be most beneficial to its employees and to the company as a whole, while also leaving room for potential expansion into small scale manufacturing.

This project will use neighborhood data from each city to make comparisons and contrasts and make a determination on which location will be a good fit for the culture of the organization.

## **DESCRIPTION OF DATA**

The data used in the analysis of the above problem was taken from the wikipedia pages of the aforementioned cities. The data consists of the names, latitude, longitude, and postal codes for each neighborhood.

The Foursquare API was used to ascertain information on the surrounding venues in each neighborhood. This information is instrumental in determining which neighborhoods are the most compatible with the needs of the employees and of the company.

**Toronto Neighborhoods** - [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

**Toronto Latitude and Longitude** - [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

**New York City neighborhoods** - [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) New York City

**Latitude and Longitude** - Python Geopy.geocoders

## **METHODOLOGY SECTION**

In order to find a solution to the problem it was necessary to collect and arrange data for both cities. This was done through utilizing the following steps:

**Import libraries and packages:**

- requests

- pandas - library for data analysis
- numpy - library to handle data in a vectorized manner
- random - library for random number generation
- json - library to handle JSON files
- geopy, Nominatim - module to convert an address into latitude and longitude values
- IPython - libraries for displaying images
- Folium - library for generating maps
- Matplotlib - libraries for plotting
- SKLearn - library for clustering

**The following steps were applied for the Toronto data and repeated for the New York data:**

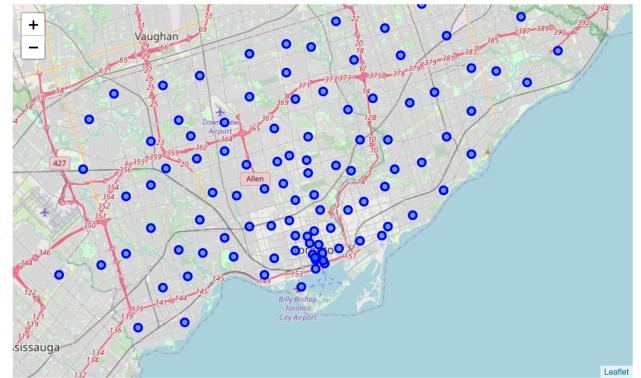
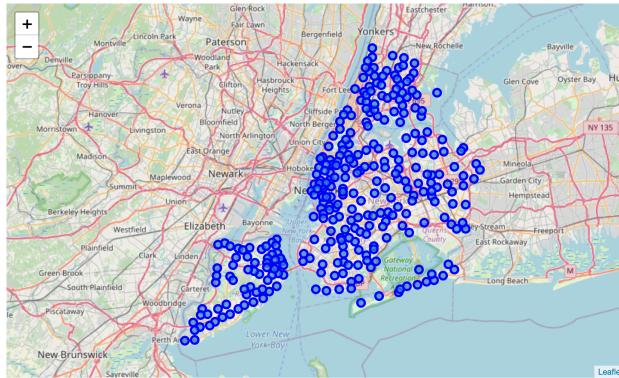
### **STEP 1: DOWNLOAD AND WRANGLE THE DATA SETS**

A dataset containing a list of postal codes, boroughs, and neighborhood names within the city of Toronto, CA is found at [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

A dataset containing the 5 boroughs and their neighborhoods as well as the the latitude and longitude coordinates of each neighborhood was downloaded from  
[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

After the initial datasets were downloaded new data frames were created containing the borough names, neighborhood names, and longitude and latitude information which was gathered through additional data sources.

Because visualization of location data is very helpful in gaining insight into the data, the Folium package was used to create maps of the Toronto and New York with their respective neighborhoods super-imposed on top:



### **STEP 2: USE FOURSQUARE API TO EXPLORE THE NEIGHBORHOODS OF SCARBOROUGH AND BROOKLYN**

A function was written to retrieve data for the surrounding venues within a 500 meter radius of each neighborhood. A data frame was created with the data of each venue consisting of:

- Venue Name
- Venue Category
- Venue Latitude
- Venue Longitude

Because our primary concern was of the venue category rather than the name, one hot encoding was used to create a data frame that displays the mean frequency of each venue category's occurrence in each neighborhood.

### STEP 3: USE K-MEANS TO CLUSTER NEIGHBORHOODS

K-means clustering is used to form clusters of the different venue categories within each neighborhood. These clusters are then analyzed to reach the conclusion of which location will be most compatible for the company.

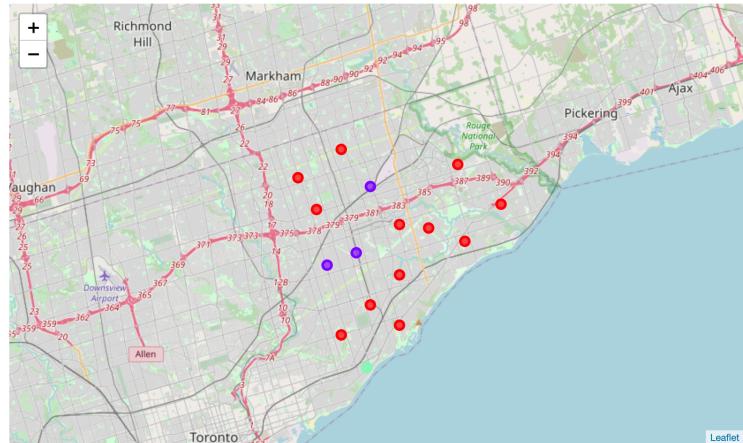
## RESULTS SECTION

### Scarborough, Toronto, Canada

The neighborhoods of Scarborough were grouped into 3 clusters using K-Means.

#### Scarborough Cluster 0

Number of neighborhoods: 15  
Most Common Venues: various international cuisine restaurants, department stores, construction & landscaping



#### Scarborough Cluster 1

Number of neighborhoods: 3  
Most Common Venues: various international cuisine restaurants, skating rinks, auto garage

#### Scarborough Cluster 2

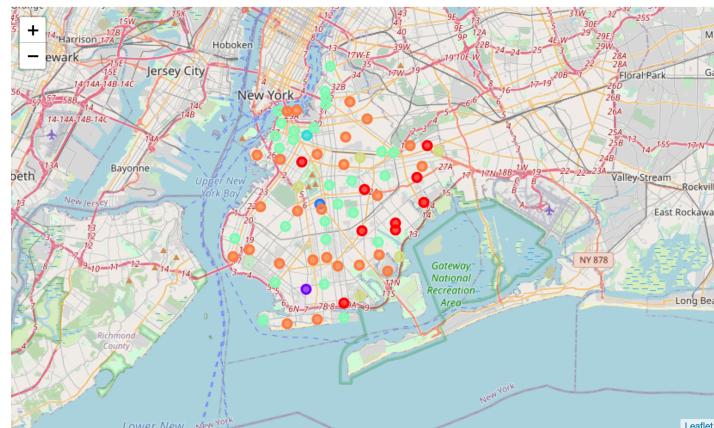
Number of neighborhoods: 1  
Most Common Venues: general entertainment, skating rink, cafe

### Brooklyn, New York

The neighborhoods of Brooklyn were grouped into 7 clusters using K-Means.

#### Brooklyn Cluster 0

Number of neighborhoods: 12  
Most Common Venues:  
Various food and restaurant venues, various recreational venues including playgrounds and pools



#### Brooklyn Cluster 1

Number of neighborhoods: 1  
Most Common Venues:  
Various food and restaurant venues, various fitness & recreational venues including playgrounds and gyms

### **Brooklyn Cluster 2**

Number of neighborhoods: 1

Most Common Venues: Chinese & Japanese restaurants, pizza place

### **Brooklyn Cluster 3**

Number of neighborhoods: 1

Most Common Venues: Caribbean restaurant, mobile phone shop, pizza place

### **Brooklyn Cluster 4**

Number of neighborhoods: 27

Most Common Venues: restaurants of varying cuisine, pharmacy

### **Brooklyn Cluster 5**

Number of neighborhoods: 5

Most Common Venues: restaurants of varying cuisine, pharmacy

### **Brooklyn Cluster 6**

Number of neighborhoods: 26

Most Common Venues: restaurants of varying cuisine, pharmacy

## **DISCUSSION SECTION**

### **Toronto Overview:**

Toronto contains 11 boroughs with 103 neighborhoods.

The geographical coordinates of Toronto: Lat 43.7170226, Long -79.4197830350134.

Scarborough borough contains 90 venues in 17 neighborhoods

There are 80 distinct venues of 55 categories.

### **New York City Overview:**

NYC contains 5 boroughs with 306 neighborhoods.

The geographical coordinate of New York City: Lat 40.7308619, Long -73.9871558.

Brooklyn borough contains 2712 venues in 70 neighborhoods

There are 2213 distinct venues in 281 categories

While Brooklyn contains a far greater number of neighborhoods than Scarborough, the distribution of neighborhoods within the clusters are very similar; Scarborough has one cluster containing a majority of the neighborhoods and the remaining clusters containing one to five neighborhoods; Brooklyn has two clusters containing the majority of neighborhoods and the others containing 12 or less.

## **CONCLUSION SECTION**

Although Brooklyn has a far greater number of venues and variety of venue categories, based on the size of the company and the type of business it is my recommendation that the Horizon Corporation open its new location in Scarborough, Toronto, CA.

This conclusion was reached because in the neighborhoods of Scarborough there is enough variety of venues to ensure that the needs of the employees of the company are adequately met, and there is also potential room for expansion as the business grows.