The Battle of the Neighbourhoods

Introduction: Business Problem

New York and Toronto are the financial capital of the USA and Canada resprectively. As such, they attrack many companies from different business sectors, amongst other banking, finance, retailing, world trade, tourism, real estate. The main purpose of this project is to use the Foursquare location data and regional clustering of venues information of New York and Toronto to find out what might be the best city that provides the optimum living standard and quality of life of the employees of a company that want to move its headquarter into those cities.

Data

For the analysis, datasets containing the postal codes, borough and neighbourhoods names have been retrieved from the respective wikipedia website pages. They have been combined with the Geographical coordinates of the neighbourhoods with the respective postal code. Information pertaining to the local venues are obtained by resorting to the Foursquare API. The venue data will provide vital information that can be used by the target audiance when opting for a city to move in their headquarters. In addition to Foursquare, various python packages will be used to create maps and machine learning models to further provide insights into our neighborhood battle project.

As way of example, a screenshot of the wikipedia webside showing the postal codes, borough and neighbourhoods names of Toronto city

WIKIPEDIA
The Free Encyclopedia

Article Talk Read

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed,

Canada Post provides a free postal code look-up tool on its website,[1] via its applications for such smartphones as the iPhone and BlackBerry,[2] and sells hard-copy directories and CD-ROMs. Many

match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0; however, the postal code M0R 6T0 is assigned to an Amazon warehouse in Mississauga, suggesting that Canad

Postal Code	Borough	Neighborhood
M1A	Not assigned	
M2A	Not assigned	
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens

By resorting to the web scraping functionality of Pandas the data can be cast into the appropriate format for the analysis

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

After the above dataframe to the Geographical coordinates of the neighbourhoods, we obtained

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Upong using Foursquare API, we obtain the venue data

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

Methodology

After all the data was collected and put into data frames, cleansing and merging of the data was required to start the process of analysis. When getting the data from Wikipedia, there were Boroughs that were not assigned to any neighbourhood therefore, the following assumptions were made:

- Only the cells that have an assigned borough will be processed. Borough's that were not assigned get ignored.

- More than one neighbourhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighbourhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighbourhoods separated with a comma .

- If a cell has a borough but a Not assigned neighbourhood, then the neighbourhood will be the same as the borough.

Using the Latitude and Longitude collected from the Geocoder package, we merged the two tables together based on Postal Code. After, the venue data pulled from the Foursquare API was merged with the table above providing us with the local venue within a 500-meter radius shown below.

Now after cleansing the data, the next step was to analyze it. We then created a map using Folium and colour-coded each Neighborhood depending on what Borough it was located in.

Next, we used the Foursquare API to get a list of all the Venues in Toronto and New York which included Parks, Schools, Café Shops, Asian Restaurants etc.

Because categorical data are not optimal for machine learning algorithms, we performed a technique in which Categorical Data is transformed into Numerical Data for Machine Learning algorithms. This technique is called one-hot-encoding For each of the neighbourhoods, individual venues were turned into the frequency at how many of those Venues were located in each neighbourhood.

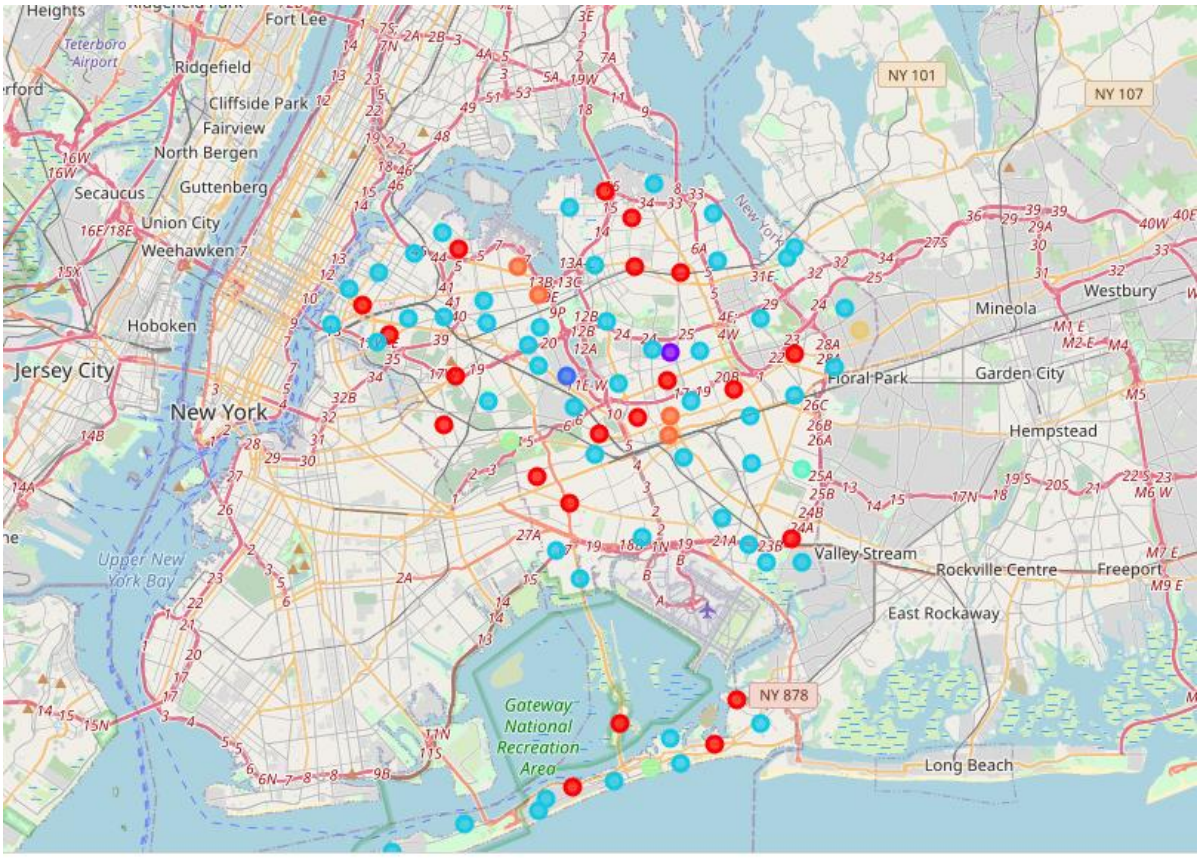
Then we grouped those rows by Neighborhood and by taking the average of the frequency of occurrence of each Venue Category.

Extensive comparative analysis of two randomly picked neighborhoods world has been carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

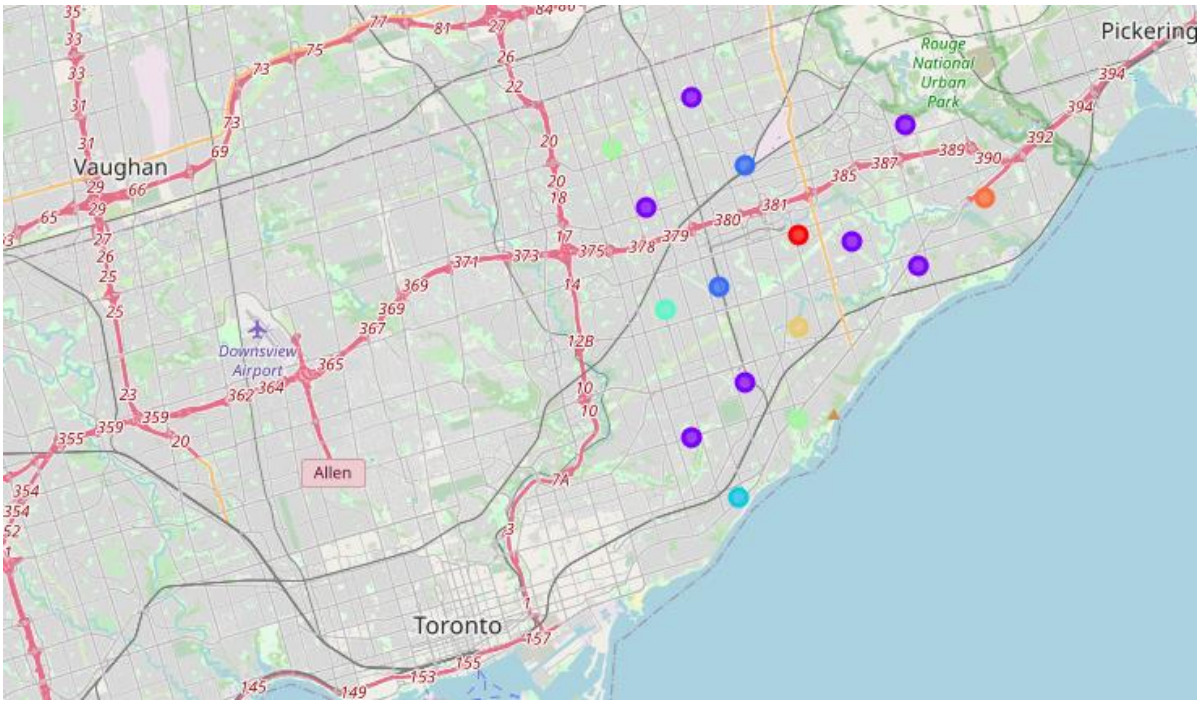
Results

Queens Borough in New York City



I used k-means to group the Queens borough into 5 clusters. Cluster_0 has 81 neighborhoods and consist of many international cuisine restaurants and grocery stores. The most common venues are pizza places, deli, and Chinese restaurants.Cluster_1 has 1 neighborhood and the most common venue is a dance studio. Cluster_2 has 5 neighborhoods and the most common venue are donut shops and international cuisine restaurants. Cluster_3 has 2 neighborhoods and the most common venues are the beach and a bakery. Cluster_4 has 2 neighborhoods and the most common venues are gyms and donut shops.

Scarborough Borough in Toronto, Canada



I use k-means to group the neighborhoods in Scarborough into 3 clusters. Cluster_0 has 15 neighborhoods and the most common venues are skating rinks, international cuisine restaurants and breakfast spots. Cluster 1 has 1 neighborhood 1 neighborhood , and the most common venues are pizza place and noodle house. Cluster 2 has 1 neighborhood, and the most common venues are Chinese restaurants and discount stores.

Discussion

Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinate of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough borough, found 88 venues in 17 neighborhoods In Scarborough borough, the neighborhoods with the most venues are L'Amoreaux West and Steeles West. There are 79 distinct venues in 52 categories.

New York City has 5 boroughs and 306 neighborhoods. The geographical coordinate of New York City are 40.7308619, -73.9871558. Foursquare found 2097 venues in 81 neighborhoods in Queens borough.

Many of the neighborhoods are homogenous and are very similar to each other. Both Scarborough and Queens borough consist of neighborhood cluster that contain majority of the neighborhoods, and the remaining cluster had 1-5 neighborhoods. Queens borough had a significant more number of neighborhoods and venues than Scarborough.

Conclusion

In conclusion, based on the quantity of venues and variety of venues, I would choose Queens over Scarborough as a choice to relocate the headquarters of the Fortune 500 company. Queens offer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families of the company's employees.