# **The battle of the neighborhoods Toronto vs New york**

**Problem statement**

A major real estate company is looking at projects in North America and looking to build a 500-unit apartment complex in either New Yok or Toronto. The senior manager of the project management team approaches the data scientist to do analysis to identify which City the apartment should be built based on the surrounding venues and amenities. The areas identified which have the most promise are Queens in NY and Scarborough in Toronto. However, the project manager wants the data scientist to prove that these two are the best boroughs in both respective cities and which one should be chosen for the project. This analysis will focus on the best place to build an apartment complex based on the surrounding amenities. The data scientist will then present the findings at a conference to both senior leadership and the project manager – to inform the decision on which city they would be spending millions of dollars on this project.

**Data**

The data that will be used for this project

* Toronto data - https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M
* New York data - <https://geo.nyu.edu/catalog/nyu_2451_34572>
* Foursquare API data
* Longitude and Latitude - http://cocl.us/Geospatial\_data

We will use the above data sources to extract data for our model of comparison. The raw data of the boroughs in both New York and Toronto will be extracted from Wikipedia and the public NYU catalog. We will also use the Foursquare API to get valuable locations and venue information about the surrounding amenities for each of the boroughs that we identify as suitable for this project.

The algorithm that we will use to compare the cities will be K means clustering to identify categories of places around the neighborhoods in question. Queens and Scarborough will be analyzed side by side and as comparatively to make sure a decision is made on the best area to build this apartment complex.

**Methodology**

Multiple Python Packages were used to help with this analysis as shown below

1. Pandas - pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming
2. NumPy - library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
3. Scikit –library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN,
4. Folium - Folium is a powerful Python library that helps you create several types of Leaflet maps
5. Matplotlib - Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.

Foursquare API

* The Foursquare API will be used to pull in data for both Queens and Scarborough neighborhoods using the same methodology that we did in the previous lab.
* Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

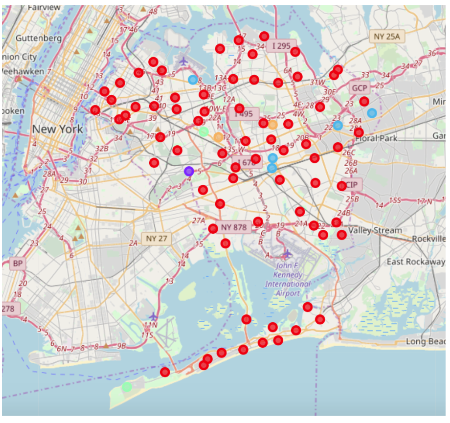
FOLIUM

* Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Seattle city over an interactive leaﬂet map.

K-Means Clustering

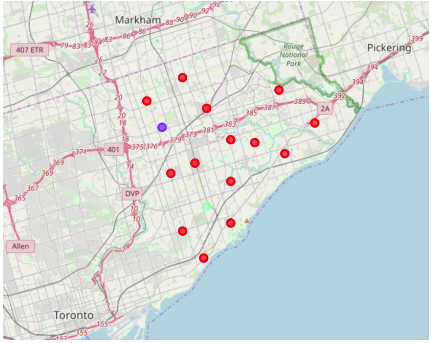
* Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python’s scientiﬁc libraries Pandas, NumPy and Scikit-learn.
* Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters 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, New york – K means clsutering analysis

|  |  |  |
| --- | --- | --- |
| **CLUSTER** | **NEIGHBORHOODS** | **MOST COMMONG VENUES** |
| **Cluster\_0** | **81** | **Pizza / Delis / International Cuisines** |
| **Cluster\_1** | **1** | **Dance Studio** |
| **Cluster\_2** | **5** | **Doughnuts / International Cuisine** |
| **Cluster\_3** | **2** | **Bakery / Beach** |
| **Cluster\_4** | **2** | **Gyms** |

Scarborough, Tornoto – K means clsutering analysis



|  |  |  |
| --- | --- | --- |
| **CLUSTER** | **NEIGHBORHOODS** | **MOST COMMONG VENUES** |
| **Cluster\_0** | **15** | **Skating Rig / International Cuisine** |
| **Cluster\_1** | **1** | **Pizza / Noodles** |
| **Cluster\_2** | **1** | **Chinese restuarant / Discount stores** |

**Discussion**

Scarborough – 17 neighborhoods – 85 places of interest for potential renters to enjoy. However, outside of the skating rig there isn’t much that brings appeal to the area from an extracurricular standpoint.

The overall building cost of Scarborough could be lower than that of Queens NY – a new analysis for this may be warranted.

Queens has 81 neighborhoods and a whopping 2108 places of interest that renters can enjoy. This along with its proximity to the ocean makes it an outstanding place to build luxury apartments.

I would also like to add that I learnt a lot during this course and would like to thank the course instructors. My python and data science knowledge were minimal and this course has helped me to get on my way to a career in this subject area.

**Conclusion**

This was a good exercise in theory and proved what we knew in practice – there is really no comparison when it comes to the marketability and the venues. Queens NY is the overwhelming winner in this project and likely to also demand much higher rent prices.

However, it all depends on the cost of the build and how soon the company can break even – there may be a second analysis warranted to figure that out.

Thanks for taking the time to read this – hope you enjoyed the course as much as I did!