

# The Battle of Neighborhoods Report

## Problem

Which city is recommended for investors to open a restaurant? Toronto or New York?

## Introduction

If an investor wants to open a restaurant, he or she needs to consider the market, target customers, competitors and many other factors of the neighborhoods. In this project, the data will show how neighborhoods in the city are classified and what characteristics they have.

People who intend to run business in foodservice industry or other related industries would be interested in this project.

## Data

*New York:*

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

This contains the Borough, Neighborhood and geographic location information in New York.

*Toronto:*

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

This contains the Postal Code, Borough and Neighborhood information in Toronto. This can be copied and pasted into an excel file. After that, it can be loaded into notebook.

[http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

This contains the latitude and longitude of the Postal Code in Toronto.

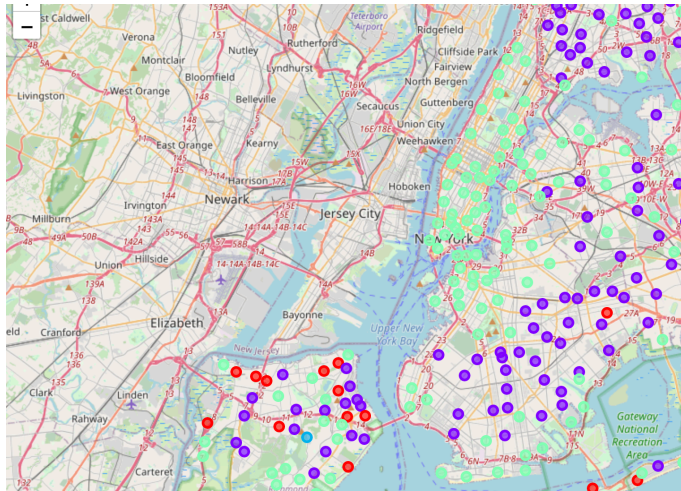
## Methodology

1. Clean the data with unassigned neighborhoods and match the postal codes with corresponding latitude and longitude values.
2. Exploratory Data Analysis to figure out the overall distribution of neighborhoods.

3. Use Foursquare API to find out the venues of neighborhoods.
4. Use one hot encoding to separate categories into columns and transpose the data frame.
5. Use unsupervised machine learning algorithm K-means clustering to classify neighborhoods into 5 clusters based on the surrounding interesting places.
6. Use Folium library to visualize the map of New York and Toronto and point out the clusters with markers in different colors.

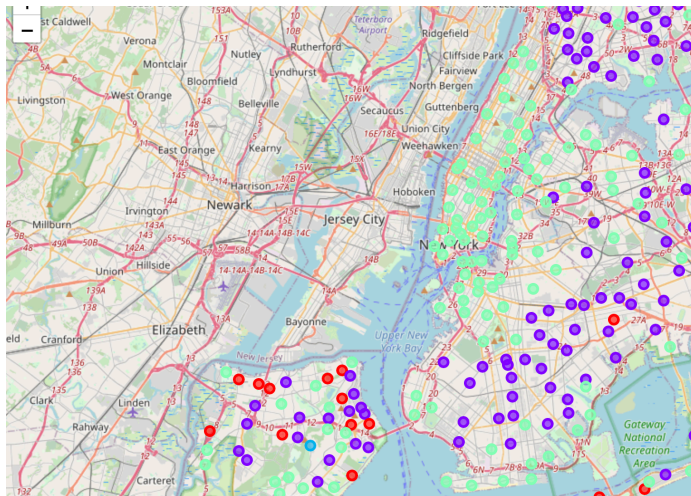
## Results

### *New York*



In New York, there are 427 unique categories. The amount variance between clusters is relatively large. The largest cluster contains mainly about restaurants like pizza place, Chinese restaurants and café.

### *Toronto:*



In Toronto, there are 269 unique categories.

The clusters are more evenly distributed compared to New York. However, there is not an obvious sign about which cluster is composed mostly of restaurants.

## **Discussion**

Maybe focusing on several famous boroughs and neighborhoods can help investors have a clearer understanding about opening restaurants in these two cities.

## **Conclusion**

Since there are enough categories and restaurants related venues in New York, the market does not have enough room for new producers. The venues in Toronto are not that scattered. The data suggests the investors that it would be more profitable to open a restaurant in Toronto.