**Toronto Neighborhoods**

Abhilash Mohapatra, October 5th, 2019

1. Introduction
   1. Background

Toronto is Canada’s business and financial capital, a growing financial hub in North America, and a top 10 global centre. Annual business operating cost are significantly lower than in many leading international centres. Toronto is growing with strong development projects helping to bring more people and jobs in the city. Globally connected, Toronto’s highly educated talent pool is the driving force behind city’s innovation and prosperity. It’s the home to largest start-up ecosystem in Canada being the destination for wide range of traditional and advanced Banking, Insurance and Manufacturing organization.

* 1. Scope

The city is known for excellent shopping, beautiful parks, delicious food, thriving sports culture, beaches, microbreweries, wineries, exotic food trucks, live music venue, speciality boutiques, global and local retail stores, internationally renowned universities and restaurants with thousands of loving fans cheering on city’s professional sports teams. Being the largest start-up ecosystem in Canada, there is a huge scope of business and developments.

* 1. Interest

Growth being a never-ending process, Construction Transport Health Education Tourism Banking and Insurance organizations need to study the existing ecosystem to broaden their scope of operations. Hence, clustering Toronto into different regions based on interest will be first step in journey.

Here we collected data from available open sources and tried to classify Toronto into 5 different classes based on different types of venues available with 500 meters of a neighborhood.

1. Data acquisition and mining
   1. Acquisition

Borough neighborhood data was scrapped from web. The venue categories gathered from json files using API calls and latitudes longitudes using geocoders. Python was used for the whole project.

Web - <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Foursquare API - <https://api.foursquare.com/v2/venues/search>

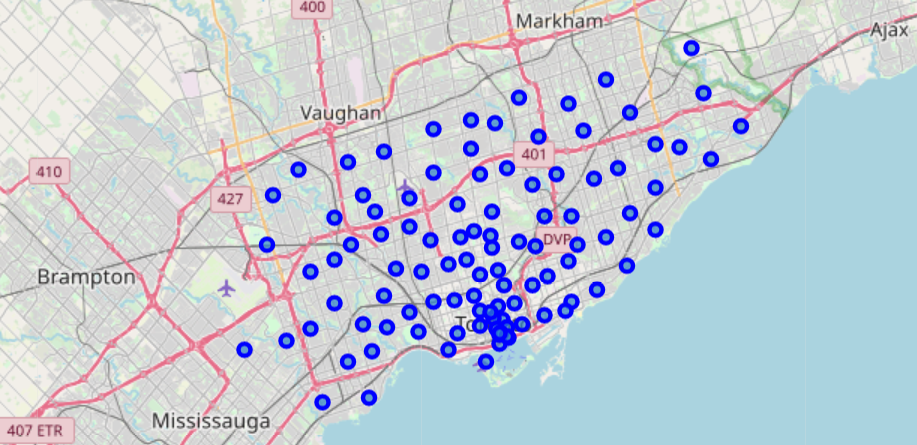
Geospatial data - <http://cocl.us/Geospatial_data>

* 1. Mining

Data was collected from the above sources and was cleaned for making understanding and making it input to ML model for clustering Toronto into 5 clusters. All boroughs with was not assigned has been neglected. All neighborhoods which was not assigned has been assigned the same name as borough. One hot encoding was used to convert the venue categories into model acceptable forms. While grouping the neighborhoods based on one hot encode, mean was used. 100 venues were considered within 500 meters of a neighborhood for classifying Toronto. All available venue categories available was considered here in the experiment. Total number of unique venue categories considered were 506. Total neighborhoods which is equal to total post codes in Toronto was 103.







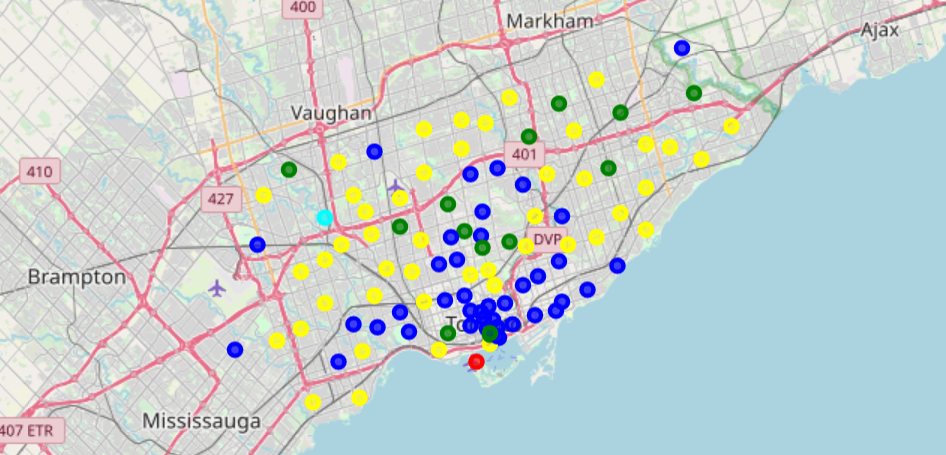
1. ML Model

kMeans clustering was used to cluster neighborhoods based on the venues nearby. Cluster Labels which is output of ML model is below as shown.



1. Conclusions

The output data from ML model was used to draw a basic insight about Toronto. Below are some final visualizations.





1. Future

The above is very minute introduction to Machine learning.

Such clustering models with optimized statistics and data sources can be used to classify and study Toronto for new scopes of development in fields like Insurance Plans (Life& vehicle), retail banking plans, restaurants, Coffee shop, Spa, Event space, Pharmacy, Uber, Amazon and others.

There is extremely huge scope of improvement from the above using techniques like Deep Learning and Statistics.