**Comparing Cities: New York with Toronto**

**Muhammad Badran**

**January 26, 2020**

1. **Introduction**
   1. **Background**

When you move to a new place and you don’t know anything about nearby restaurants and cafeterias around you, it is convenient to know some information and read other neighbors’ opinions and reviews about these places. When you visit a restaurant for the first time, you need to know the best dish they offer and try it yourself. Most cities in developed countries have many restaurants that compete with each other to provide new and best dishes. In most cases, these restaurants are close from each other and they need to get the best reviews from their customers which encourages others to visit those places. Foursqaure provide a great services to collect all information in one place. It can be used by developers to make applications that can be used to analyze these restaurants.

* 1. **Problem**

Comparing between cities of developed countries such New York City in the United States of America and Toronto in Canada and cluster the restaurants in them into categories would give insights about how the market performs in them and also after classifying these restaurants into clusters will give more information about similarities and differences between cities like New York and Toronto.

* 1. **Interest**

Owners of these restaurants, businessmen who would like to invest their money in that business in these cities, and even customers who visit or planning to visit these restaurants are interested, also tourists from other countries would be interested in these information.

1. **Data acquisition and cleaning**
   1. **Data sources**

New York data are from a json file that was provided during our course, and Toronto data are from table in Wikipedia page which contains list of postcodes in Canada. These data was used with Foursquare API for finding data about restaurants in different neighborhoods and clustering them.

* 1. **Data cleaning**

For New York data, after loading the json file which contains data about of 306 neighborhoods, some features are extracted from the json file such as Borough, Neighborhood, Latitude, and longitude. Then, a for loop is used for filling the neighborhood data frame. Now neighborhood data frame can be used with geocoder to find the latitude and longitudes.

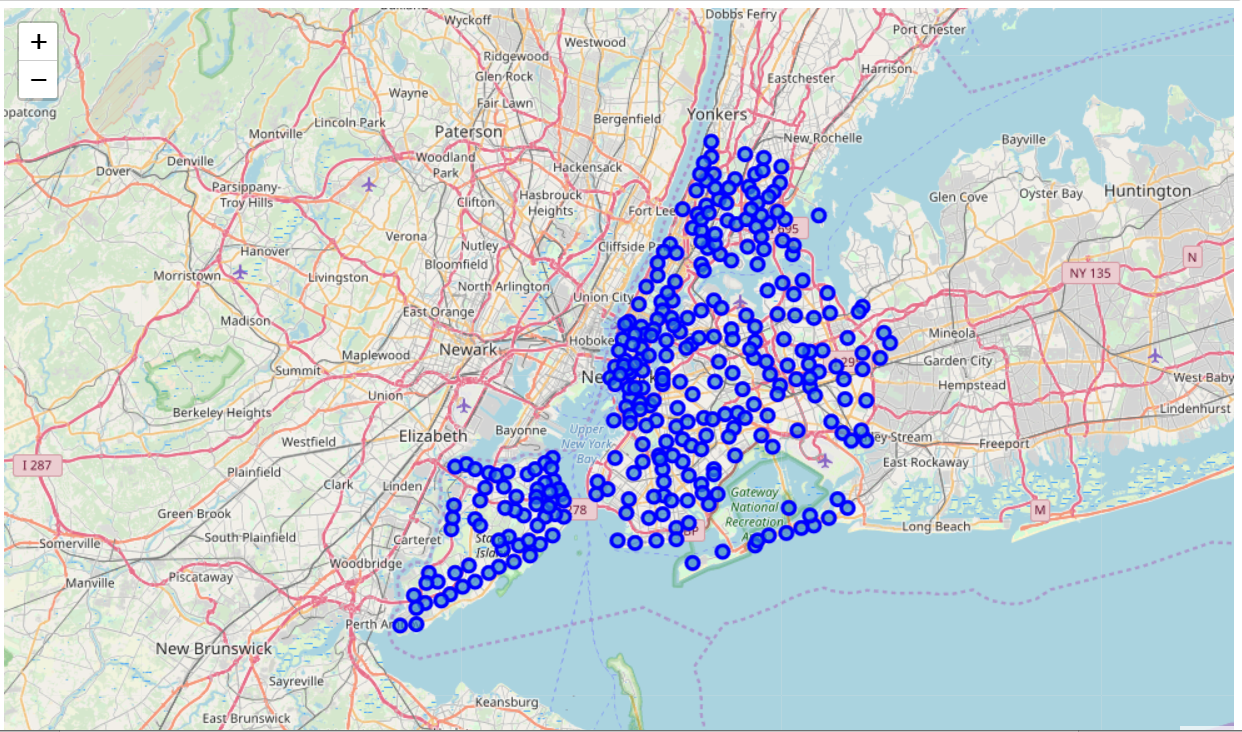
For Toronto data, after scraping postcodes from list of postcodes of Canada from Wikipedia page, all boroughs and neighborhoods are cleaned up by removing all not assigned values from the table. Also, geocoder liberary was used for finding the latitude and longitude of neighborhoods in Toronto.

1. **Method**

Importing the following important libraries to accomplish the following tasks:

* numpy to handle data in form of vectors which fasten the computations
* pandas to deal with data frames
* json library to handle json files
* requests library to handle requests
* matplotlib for plotting charts
* k-means for clustering
* folium for map rendering

Geocoder library is used to find latitude and longitude of New York City and finally a map of 306 neighborhood is formed by folium libraray.



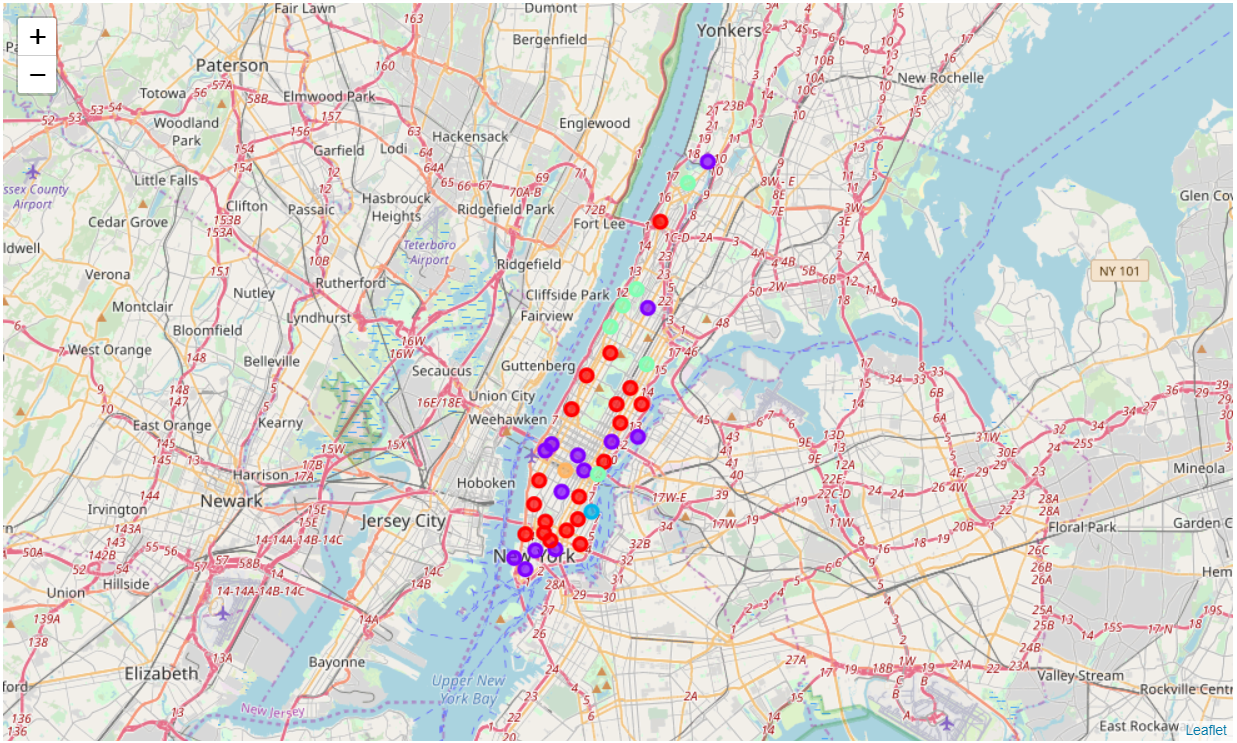
Only locations of Manhattan are selected from the data frame, and then another map of Manhattan is created.



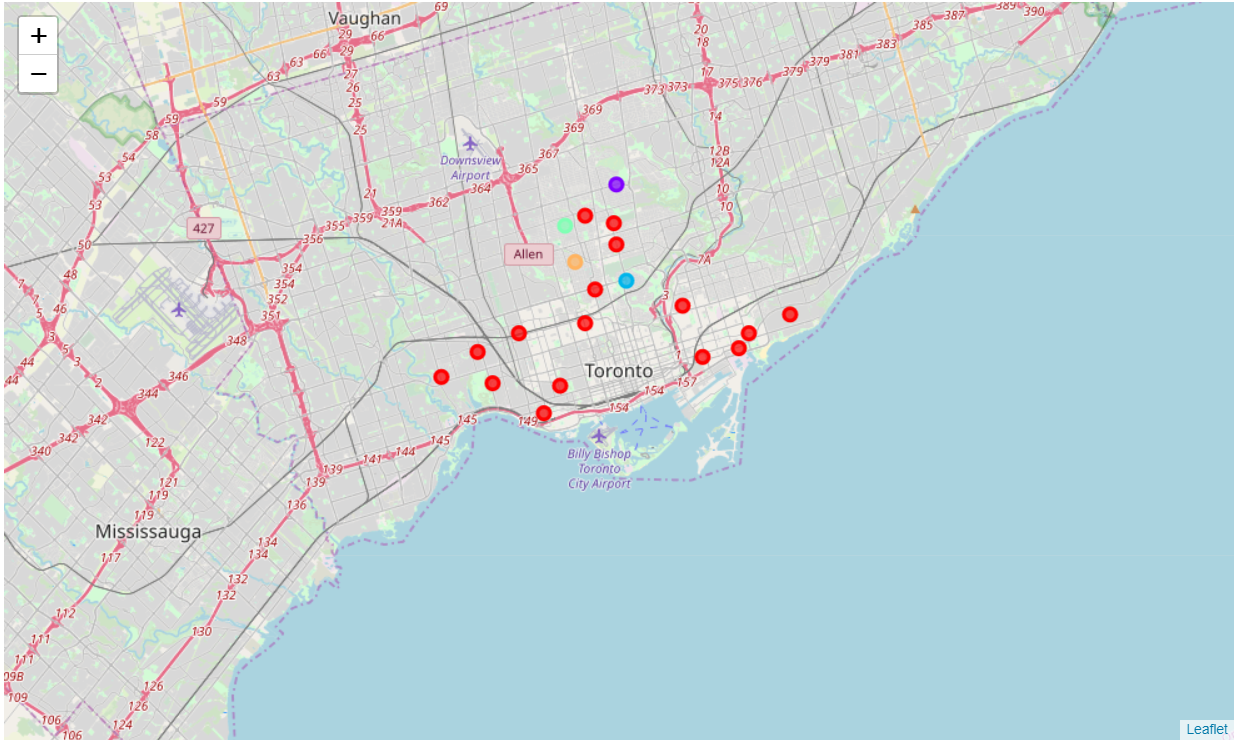
After providing Foursquare credentials and the url was built, it was used along with limit and radius variables for getting the requests from Foursquare API. The results are returned as json file.

Categories of restaurants of 24 venues with latitude and longitude are extracted from Foursquare data of Manhatten. A getNearbyVenues function was used get nearby venues. The categories were encoded by using One hot encoder then the column were arranged, then venues were grouped by their Neighborhoods. Only most common venues were returned.

The final step is clustering and forming the map with color circles which represent clusters.



The above map of New York shows the clustered locations with different colors to show each cluster. The same map of Toronto in the map below.



In case of Toronto, the data was extracted from postcodes from Wikipedia page. Then the same procedure described above was carried out to draw the map of the locations on Toronto. The map is shown below.

1. **Conclusion**

There is more diversity in clusters of New York than in Toronto. Both cities are great an full of life and are active for many businesses.

1. **Future Work**

Applying same procedure on different cities when their data is available.