FINAL REPORT

Jabin Choi

Coursera – IBM Data Science

12/30/2019

CAPSTONE PROJECT



# Intro – Problem description

It is no doubt that Canada is one of many countries with immigrants. Scarborough, in particular, is one of the cities that recently attracted a lot of attention among immigrants. This project is useful for people planning to move to Scarborough by showing the results of comparing and analyzing dataset related to their average housing price and school rates among the venues registered in the app called Foursquare. For more details on Foursquare, please visit the following url:

\*\* Foursquare URL: https:// foursquare.com

# wHAT IS FOURSQUARE API?

In this project, dataset related to the average of housing price and school ratings among the venues registered in the app called Foursquare.

Foursquare is a social location service that provides users to explore the location around the world. The app enables users to download it through iPhone, Blackberry or Android devices and then connects their account to other social media accounts.

Foursquare API itself is a RESTful set of addresses to which developers can send requests. Users has nothing to download onto their server and it just allows developers to interact with the Foursquare platform.

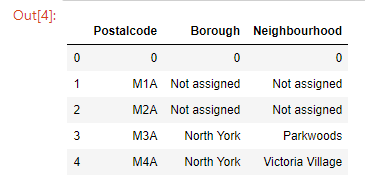
The major purpose of using Foursquare API is to gather data source as it has a database of millions of places and the API provides the ability to perform location search and sharing as well as details on a business.

# DATA approach - METHDOLOGY

First, data will be loaded from Wiki url:

‘ <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> ’

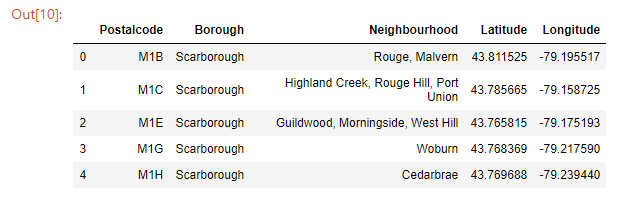
The data table looks like the followings:



Data cleaning:

* Drop meaningless contents in the table – meaning handle ‘Not assigned’ and None values
* Group and re-index the table
* Add two columns called ‘Latitude’ and ‘Longitude’

Once data cleaning is completed, the final data table before analyzing might look as follows:



Data from Foursquare API:

To gain datasets of various venues in different neighborhoods of that specific borough – Scarborough, we need to use ‘Foursquare’ locational information. The information contains venues name, locations, menus and photos.

* Neighbourhoods
* Neighbourhoods Latitude / Longitude
* Venues
* Venue names such as restaurants or stores
* Venue Latitude / Longitude
* Venue Category

In this project, we will extract information that we need such as venue names, categories and latitude / longitude. Once the columns are obtained from through the API, then it might look as follows if having chosen the radius to be 100 meters:



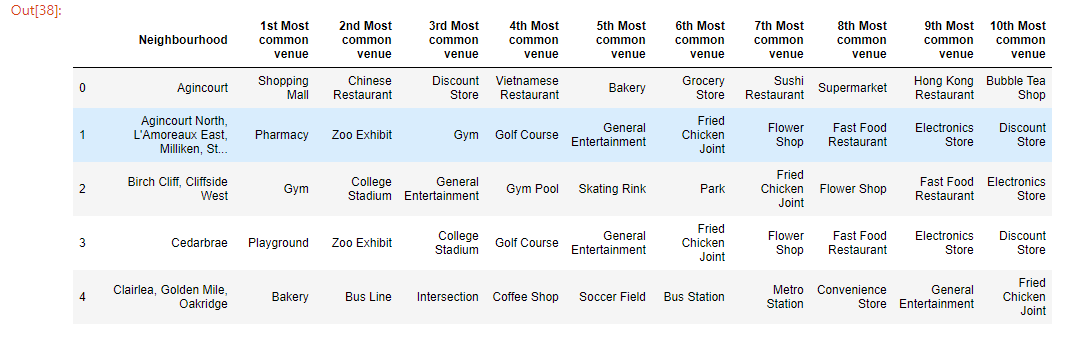
Methodology – K Means Clustering:

In this project, K-Means clustering that is a form of unsupervised machine learning is used for data analysis. By grouping similar data points (here, ‘Neighbourhoods’ column) together, we will discover underlying patterns and analyze dataset related to the average housing price and school ratings among the venues.

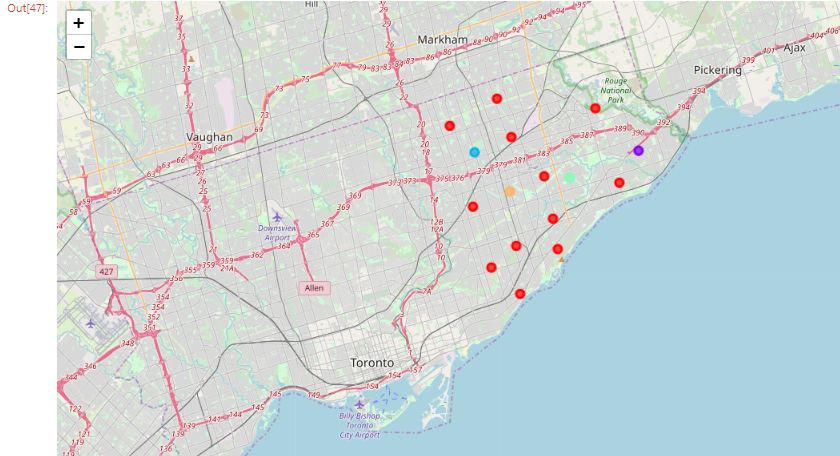
From the methodology, we are able to see top 5 most common venues per neighbourhood in Scarborough by setting 5 clusters (k=5).

# sORTED DATA & RESULTS OF k-mEANS

[Sorted data table]

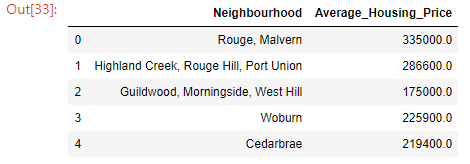


[Result of K-Mean Clustering, k=5]

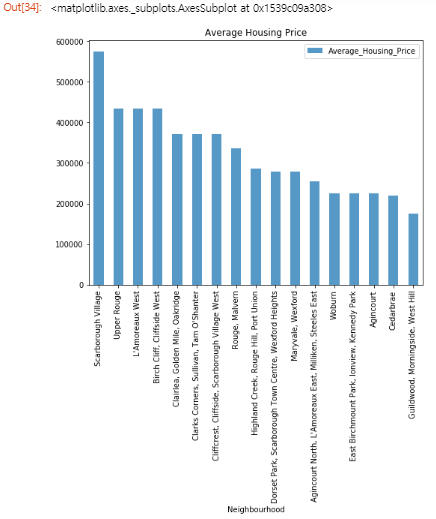


# dISCOVERY THROUGH DATA ANALYSIS

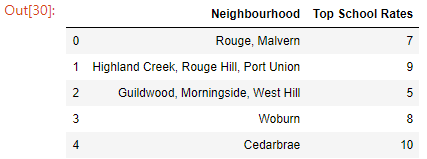
Once combine Scarborough neighbourhoods dataset with average of housing price, the result of the table is as follows:



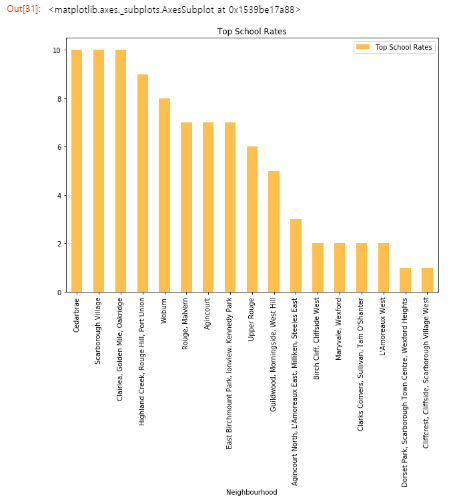
The result of bar plot after sorting values by the average of housing price is as follows:



This time, we will combine Scarborough neighbourhoods data with top school rates. The table looks like the followings:



If we sort values by top school rates and then plot the table, then it looks like the following:



# Conclusion

[Result of Barplot 1]

From a bar plot named ‘Average Housing Price’, we are notified that ‘Scarborough Village’ is the neighbourhood that has the highest average of housing price and ‘Upper Rounge’ / ‘L'Amoreaux West’ are going after.

[Result of Barplot 2]

From a bar plot named ‘Top School Rates’, we are notified that ‘Cedarbrae’ is the neighbourhood that got the highest school rates and ‘Scarborough Village’ / ‘Clairlea, Golden Mile, Oakridge’ are going after.