**Introduction to problem:**

It is quite common now a days that we might have to shift out of place that we are currently staying .The reasons that we might have to shift can be multifold: may be we have to shift as current place is far away from work, or we might have to shift if current place is far away from our kid’s school, or we might have to shift as the rent we have to pay at current locality might have increased. It is quite common that we might feel out of place if we shift to a locality, which is not very similar to our neighborhood. So in this project we are trying to identify similar neighborhoods in city of Toronto and we are trying to recommend a list of similar localities a person can shift in city of Toronto if at all he has to shift from his current neighborhood.

**Data we are using in this project:**

Using the following link,

[https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M,](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" \t "_blank) we are going to gather data of neighborhoods and their corresponding postal codes and boroughs

Using the following link,

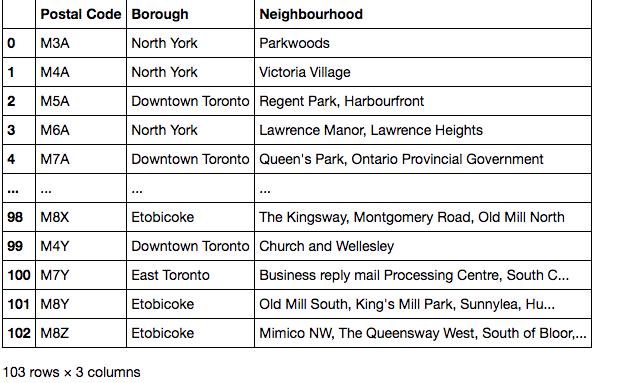
<http://cocl.us/Geospatial_data>, we gonna gather data of corresponding latitude and longitude

Then using four square API, we are going to gather near by venues in each neighborhood of Toronto and use this data to cluster neighborhoods in city of Toronto

Each cluster represents similar neighborhood and using this criteria we will define similar neighborhoods into which a person can move

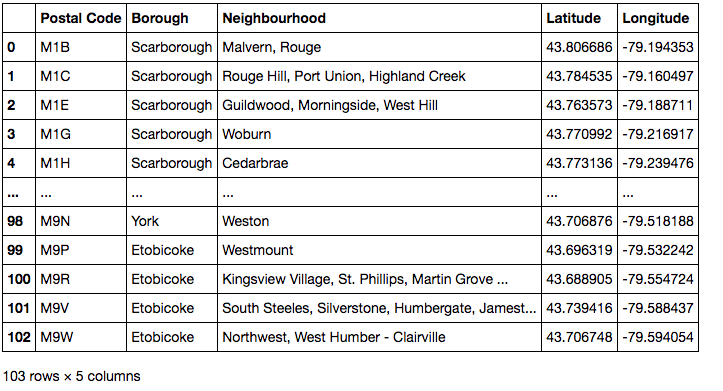
**Methodology:**

* The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
* Only process the cells that have an assigned borough. Ignore cells with a borough that is **Not assigned.**
* If a cell has a borough but a **Not assigned**neighborhood, then the neighborhood will be the same as the borough.



csv file that has the geographical coordinates of each postal code: [http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data" \t "_blank)

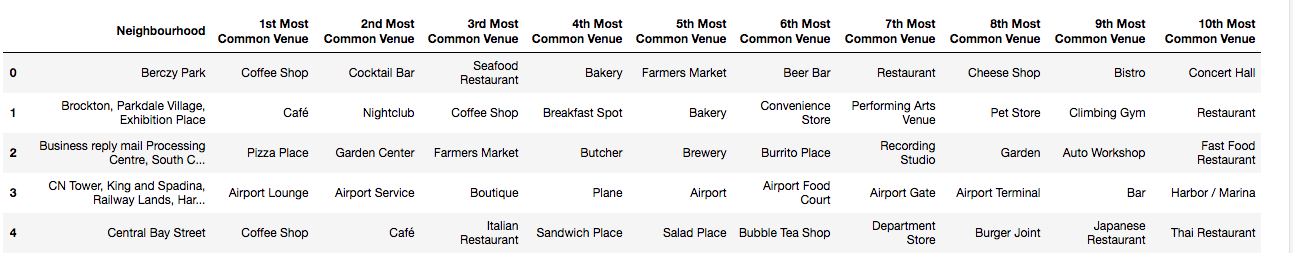
Use the Geocoder package or the csv file to create the following dataframe:



Explore and cluster the neighborhoods in Toronto using k-means and four square API

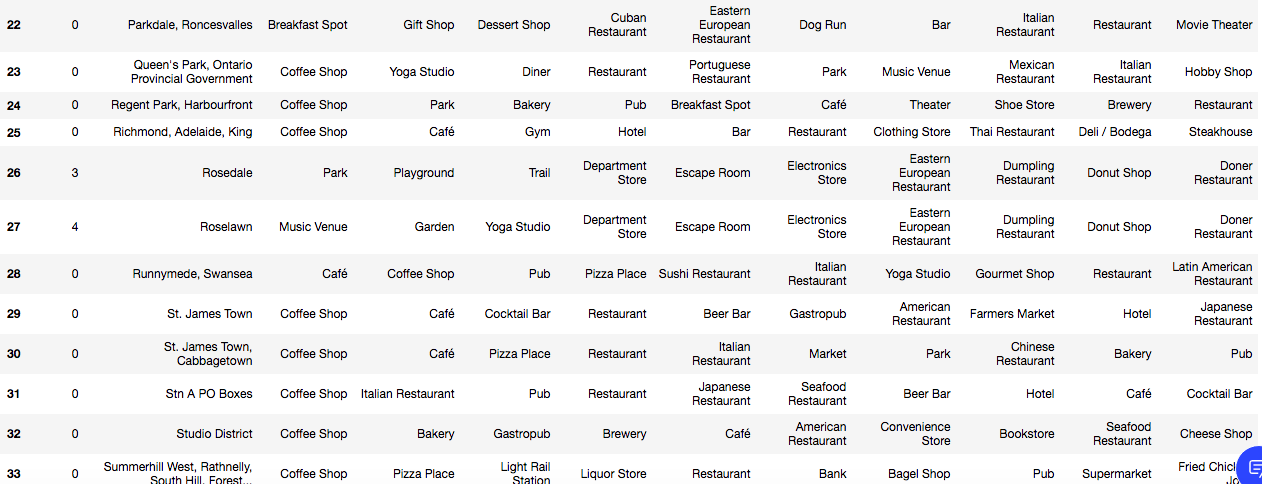
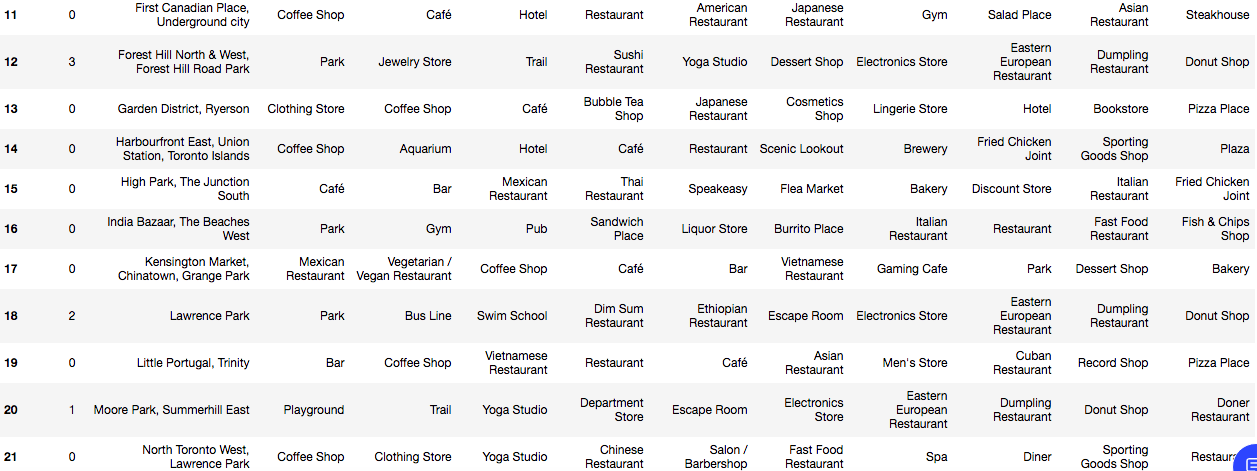
Work with boroughs that contain the word Toronto.

No of clusters=5



Results:







discussion section:

As per above table ,

all neighborhoods with cluster label 0 are similar

all neighborhoods with cluster label 1 are similar

all neighborhoods with cluster label 2 are similar

all neighborhoods with cluster label 3 are similar

all neighborhoods with cluster label 4 are similar

People in same cluster neighborhood should prefer similar neighborhood in same cluster if they shouldn’t feel out of place

Most neighborhoods are in cluster 0

Conclusion:

Using k means clustering we were able to divide neighborhoods in Toronto city into 5 clusters .We found that most number of neighborhoods are in cluster 0