**Coffee Shop in Toronto?**

Aishwarya Guntoju

February 04, 2021

**1. Introduction**

**1.1 Background**

Canadians drink more coffee than others in the world. Canada scores 3rd when tallying coffee consumption across the globe. Out of 80 countries, Canada ranked No. 1 in 2015 when tallying up how many litres of coffee per capita people gulped down at food service joints like cafes. The country also scored third highest for the total amount of brewed coffee people consumed both inside and outside the home which is  an average of 152 litres per person. The statistics were compiled by global marketing research company, Euromonitor.

Long, cold winters are certainly a factor for Canada’s coffee binge. Euromonitor research analyst, Mark Strobel said that there's definitely a correlation between ambient weather and hot drinks consumption. But it is not the only factor. Other factors include Java consumption at food service locations, Tim Hortons everywhere in the country. Tim Hortons is a quick service restaurant which is known in particular for its coffee and baked goods.There are currently 3,692 Tim Hortons locations across Canada. That's about one for every 9,000 Canadians. There's no restaurant in the world that has this type of penetration. Moreover, Tim Hortons has become part of Canada's cultural identity. Also, Coffee being a part of Canadians life highlights the fact that there are many Coffee Shops in Canada.

**1.2 Problem**

Coffee is a part of Canadians daily routine.Canada also has many coffee shops.But it doesnot mean that there is no possibility of starting up a new one.

Moreover starting a coffee shop in Toronto will definitely work as people love coffee but we should find a proper location.

**Objective** : To find the possible locations in the neighbourhoods of Toronto to open a new Café based on population and density.

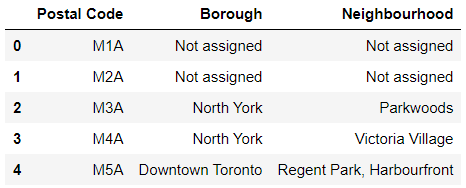
**Business question** : In the neighbourhoods of Toronto, where an entrepreneur can start a new Coffee brand or a franchise to enjoy competitive advantage?

**2. Data Acquisition and Preparation**

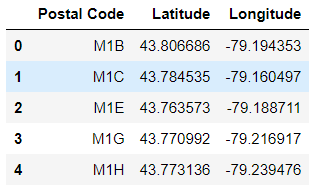
**2.1 Data sources**

Data is scraped from two web pages and also from FourSquare. From [List of postal codes of Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), the postal codes of all the neighbourhoods is extracted and demographics are extracted from [Demographics of Toronto neighbourhoods](https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods#:~:text=Etobicoke%20%20%20%20Name%20%20%20,%20%202249%20%2011%20more%20rows)​. Location data is extracted from Geospatial\_Coordinates.csv. Venues present in the neighbourhoods are extracted from FourSquare. Dataframes are:

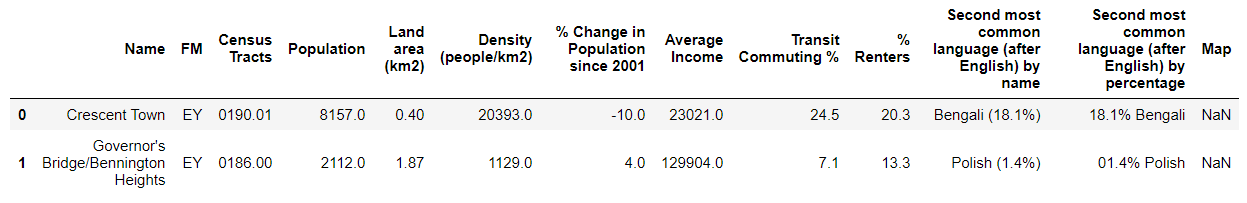
1. df ([List of postal codes of Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))



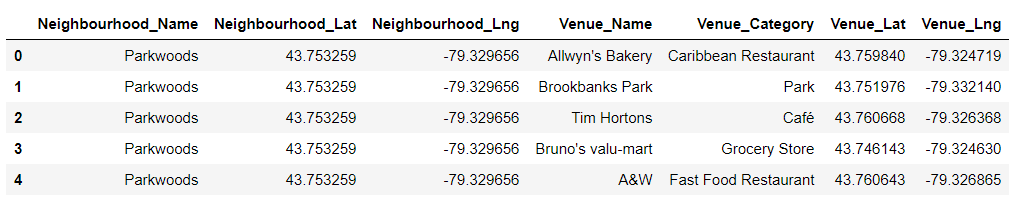
1. ll\_data (Geospatial\_Coordinates.csv)



1. pop\_den ([Demographics of Toronto neighbourhoods](https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods#:~:text=Etobicoke%20%20%20%20Name%20%20%20,%20%202249%20%2011%20more%20rows)​)

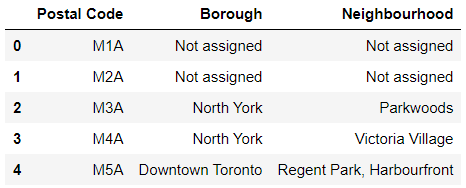


1. neigh\_venues (FourSquare API)

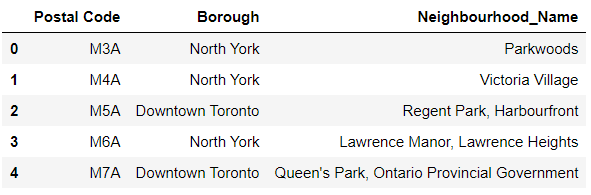


**2.2 Data cleaning**

Data scraped from multiple sources are stored into tables. Later,tables are merged as per requirements. This is the dataset from scraped web page : [List of postal codes of Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

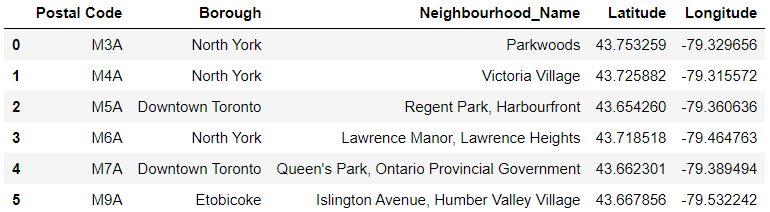


There are lot of missing values (“Not Assigned”). So, I dropped the rows where Neighbourhood = = “Not Assigned” as they don’t help in analysis.

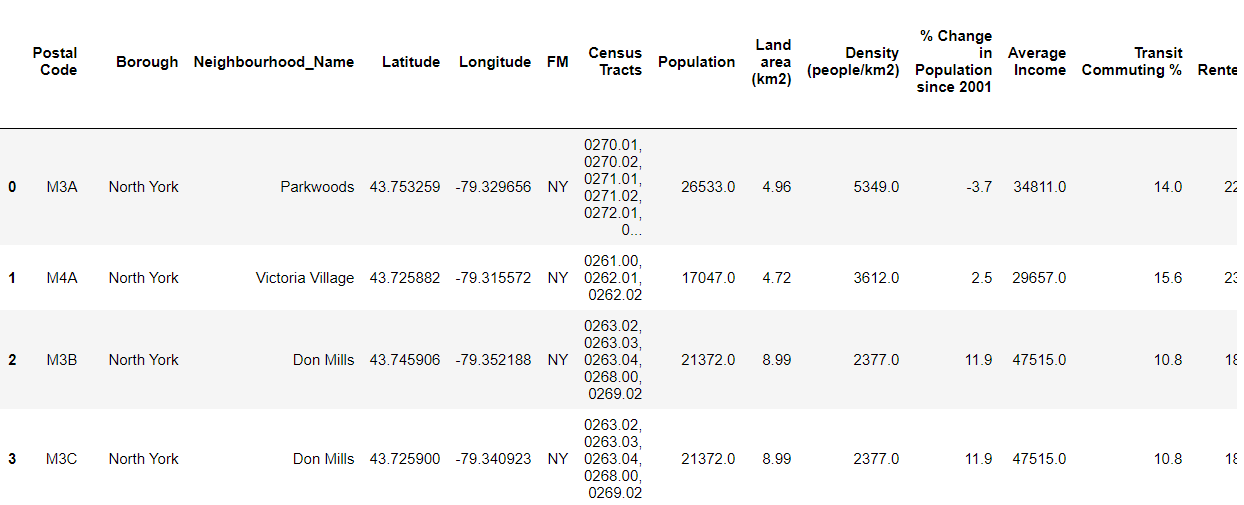


**2.3 Feature Engineering and Feature Transformation**

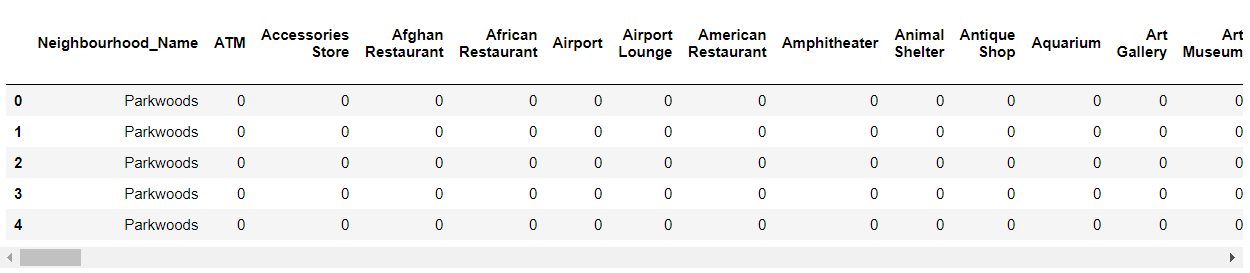
ll\_data is merged with df data and resultant is toronto\_df.



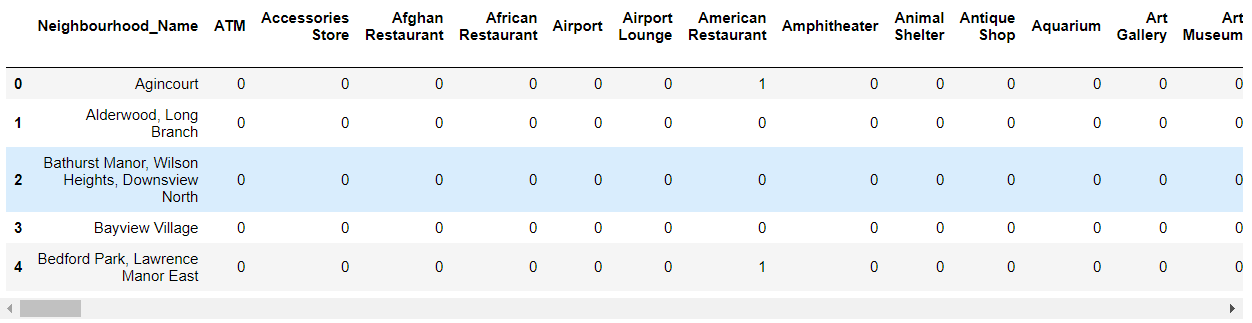
A dataframe neigh\_pop\_den is formed by Toronto\_df and pop\_den.



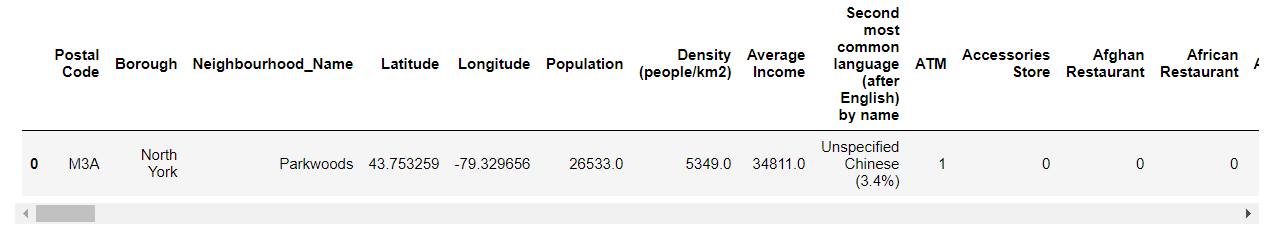
In neigh\_venues, dummies are generated from the column ‘Venue\_Category’ and neigh\_venues\_onehot is formed.



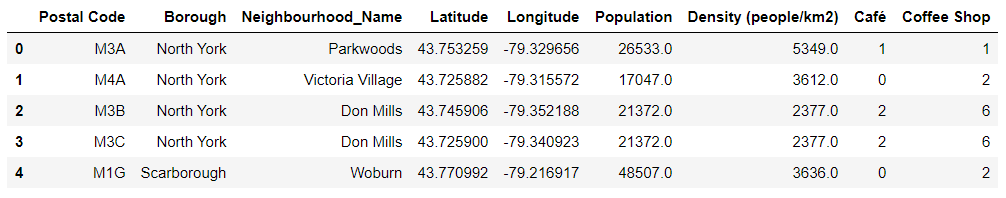
neigh\_venues\_group is formed grouping by ‘Neighbourhood\_Name’.



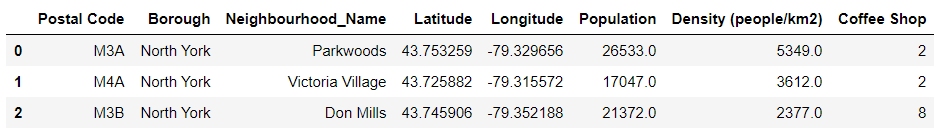
Now, merged is formed by merging neigh\_pop\_density(formed by dropping unrequired columns) and neigh\_venues\_group .



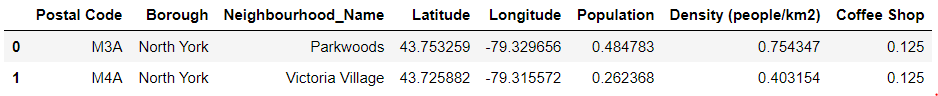
neigh\_pop\_den\_venues is formed by dropping unrequired columns from merged.



As Café and Coffee Shop are same, one column is formed with the name Coffee Shop by concatenating both the columns.



In the above dataframe ; Population, Density (people/km2) and Coffee Shop are normalised using standard normalizer which is also called as MinMaxScaler().



**2.5 Feature Selection**

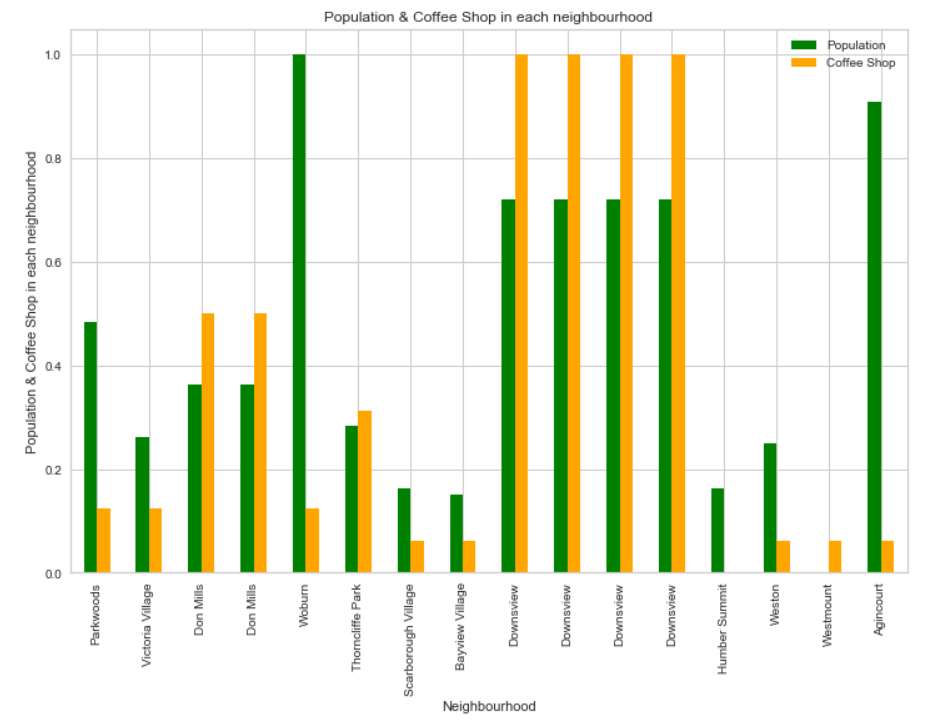
After cleaning,transformation,engineering and also by dropping all the unrequired columns, a dataframe neigh\_pop\_den\_venues is formed with the columns : Postal Code, Borough, Neighbourhood\_Name, Latitude, Longitude, Population, Density (people/km2), Coffee Shop.

Another dataframe named clusters is formed with only one column ‘Coffee Shop’ as we are clustering Toronto neighbourhoods based on Coffee Shop.

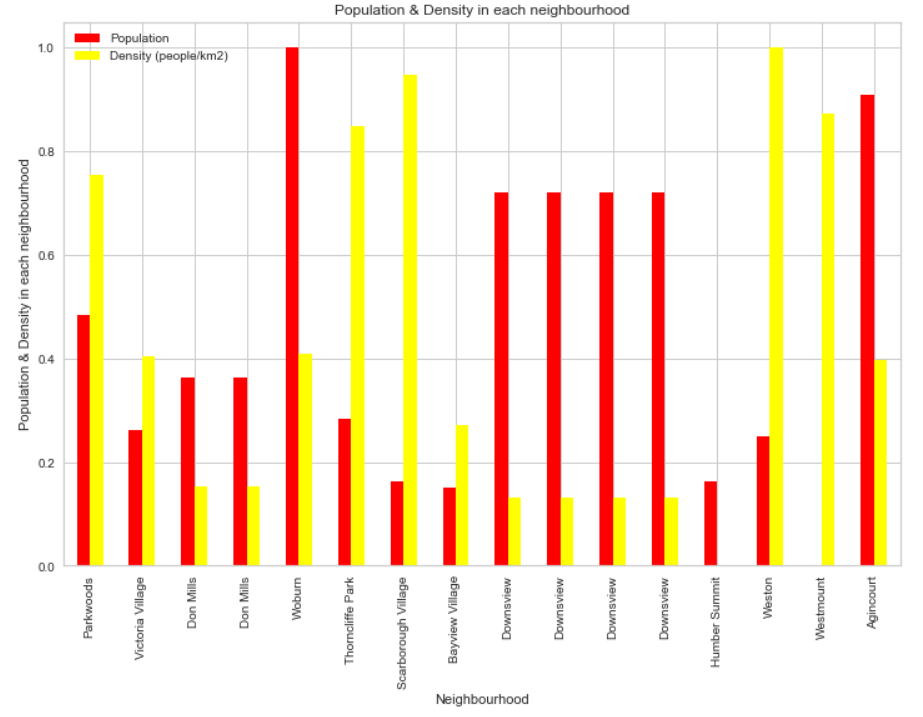
**3. Exploratory Data Analysis**

As am going to search the locations for starting up a Coffee Shop based on population and density, I first checked for the correlation between ‘Population’ and ‘Coffee Shop’ as well as ‘Population’ and ‘Density(people/km2)’.

**3.1 Relationship between Population and Coffee Shop**



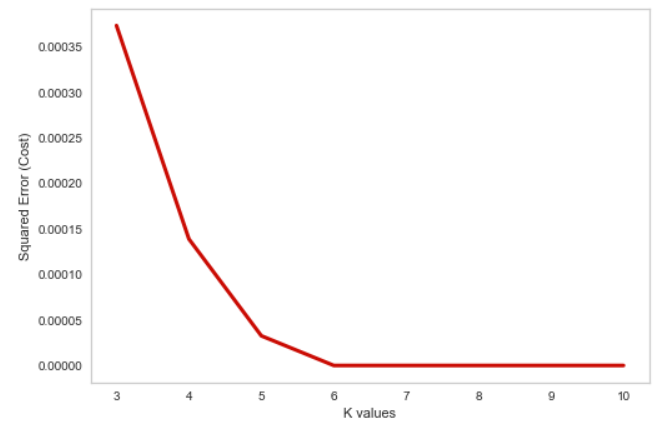
**3.2** ​​**Relationship between Population and Density**



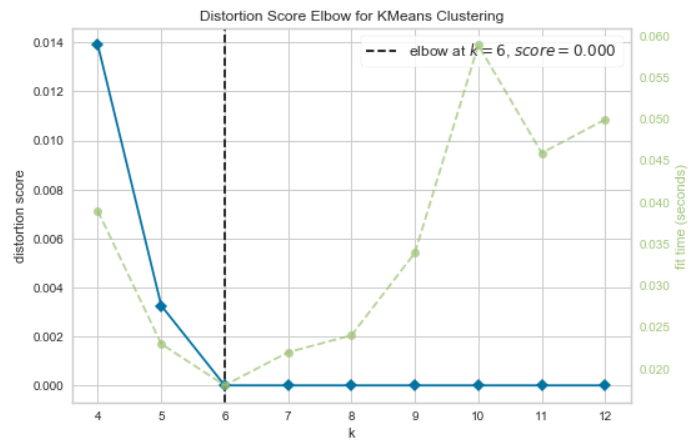
**4. Predictive Modeling**

**4.1 Clustering neighbourhoods of Toronto**

Firstly I identified best k value (number of clusters) to perform clustering and for the same I used elbow method on clustering dataset with Coffee Shop frequencies.

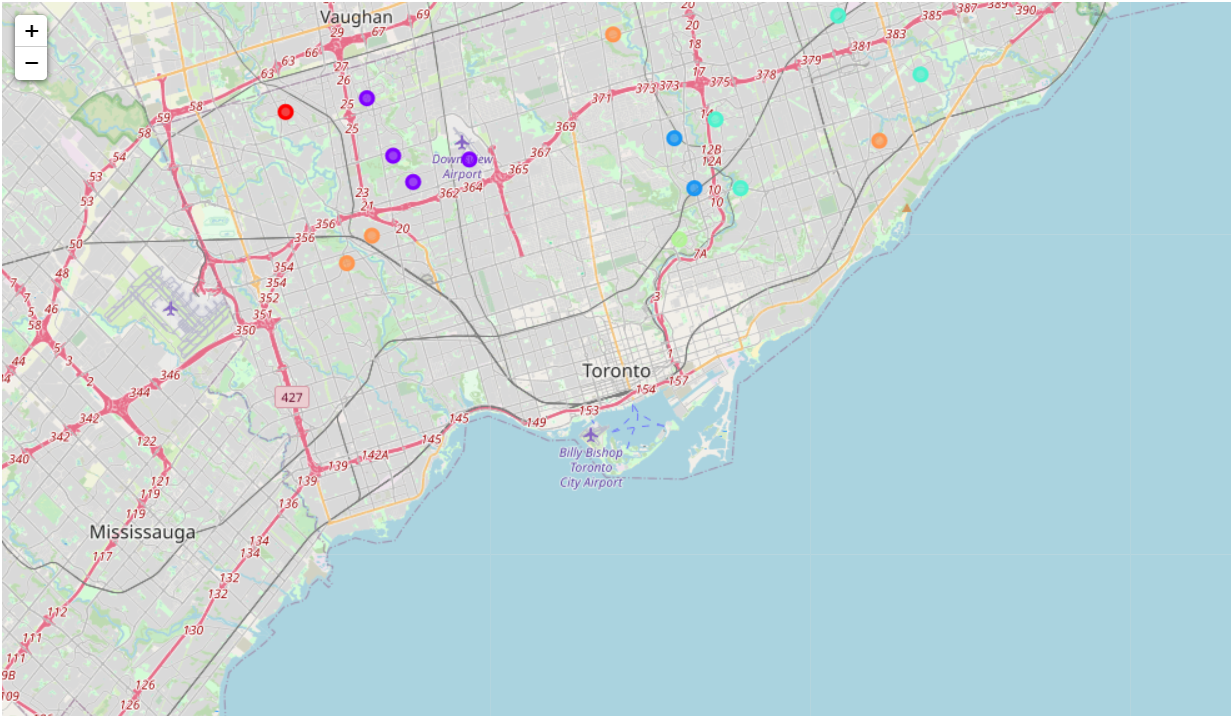


We can understand the best k from the above graph , but to have a clear idea I used KElbowVisualizer.



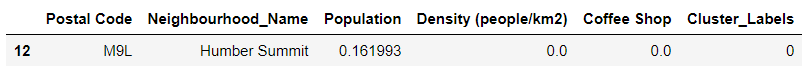
From the above graph , the best k is 6.

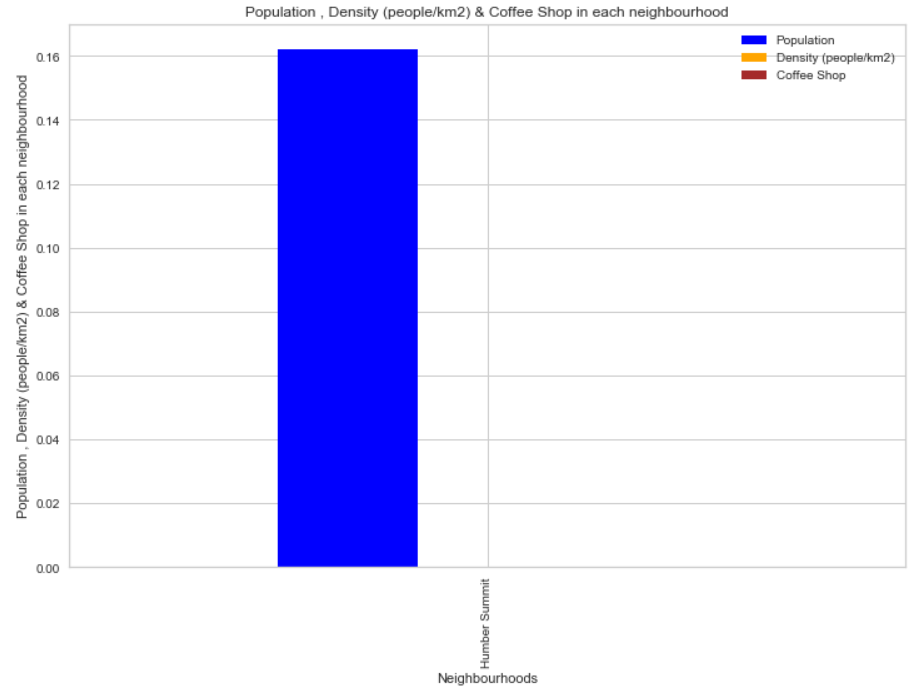
After fitting data to KMeans clustering, labels are generated.I used folium map to view the clustered neighbourhoods of Toronto.



**4.2 Examine Clusters**

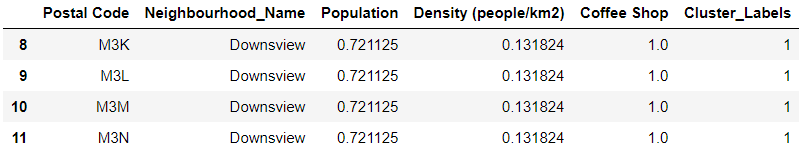
**Cluster 1**

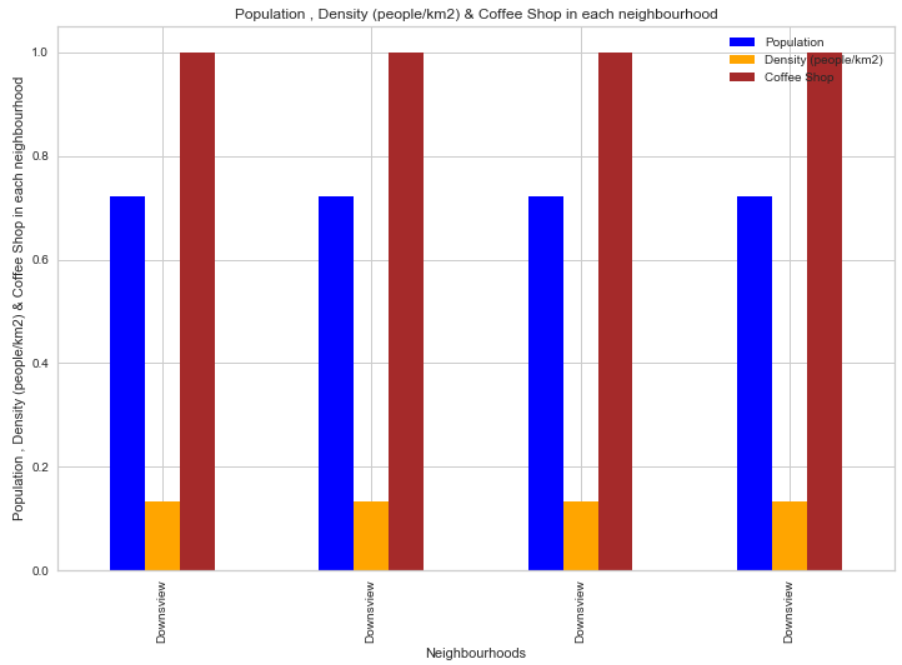




The values of Density and Coffee Shop are very low that they are not viewed on the graph.Cluster1 has very low coffee shops as per population.So, Coffee Shop can be started in Humber Summit.

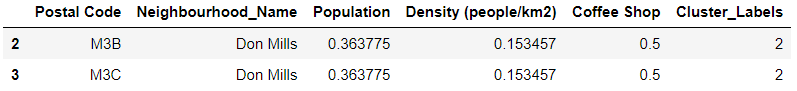
**Cluster 2**

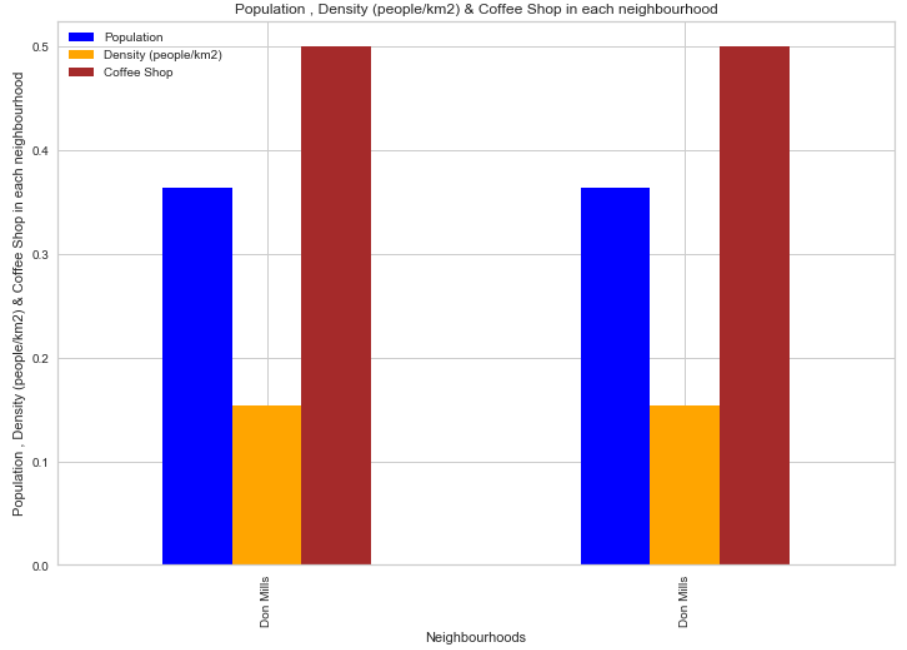




Cluster2 has large number of coffee shops.So, there is no possibility of starting one more.

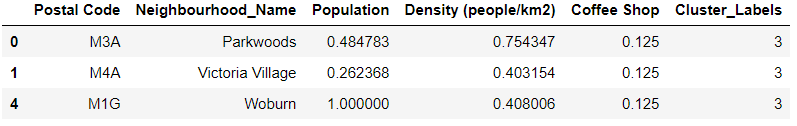
**Cluster 3**

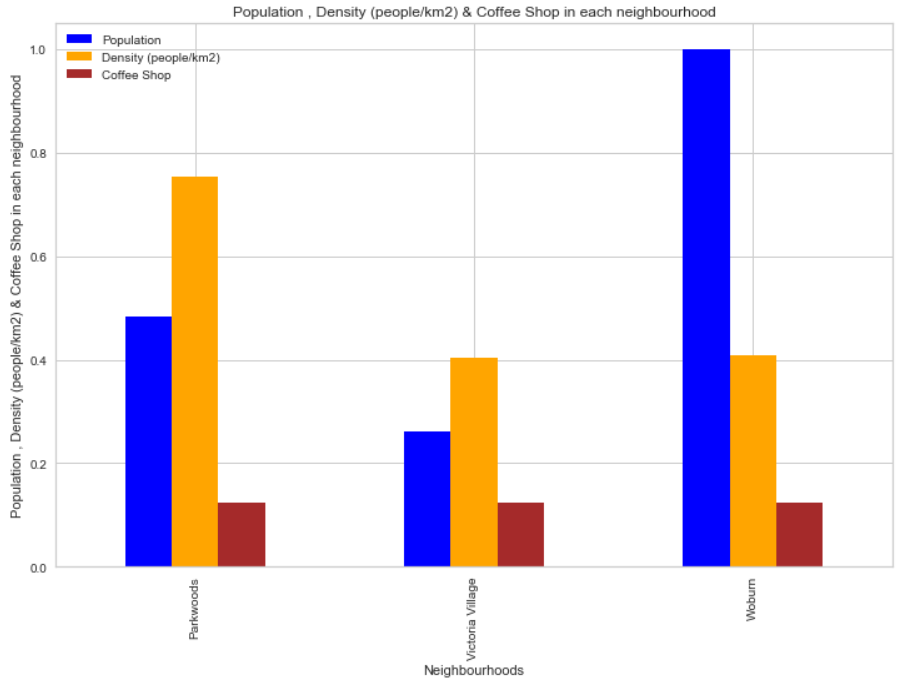




Cluster3 also has many Coffee Shops which can serve the population of the neighbourhoods so, no possibility for one more.

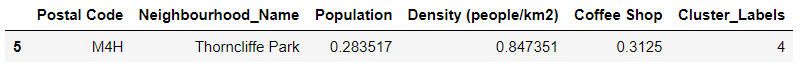
**Cluster 4**

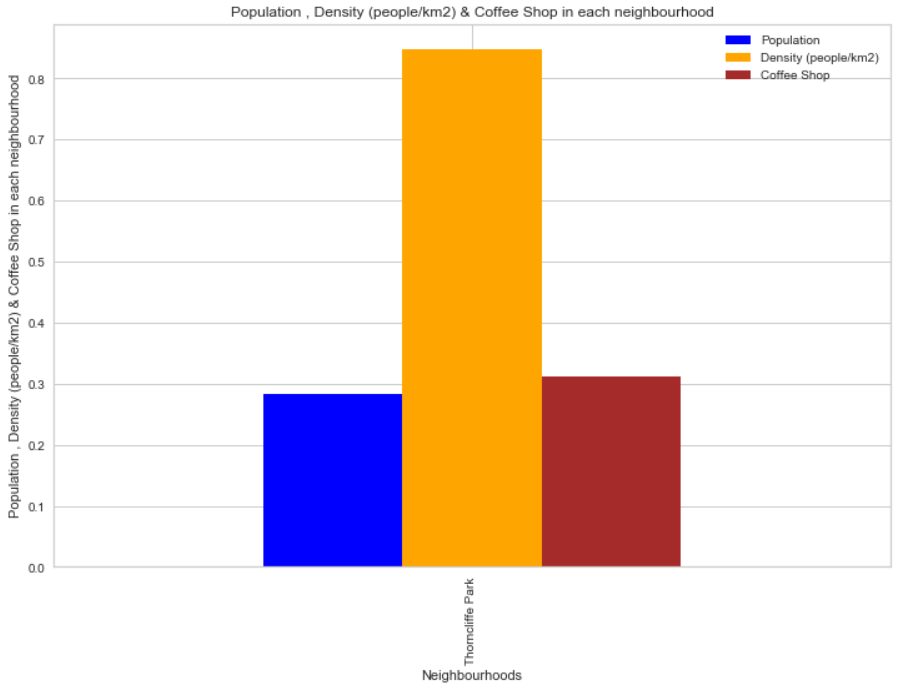




Cluster4 has less Coffee Shops. As density is high although population is low in Parkwoods , Victoria Village and the Woburn has high population, a new Coffee Shop can be started.

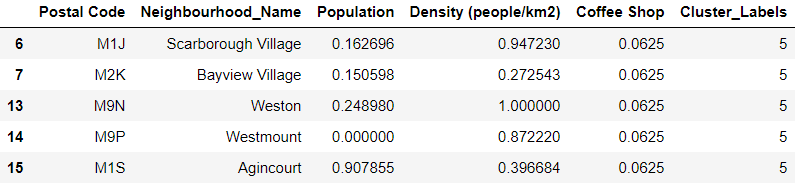
**Cluster 5**

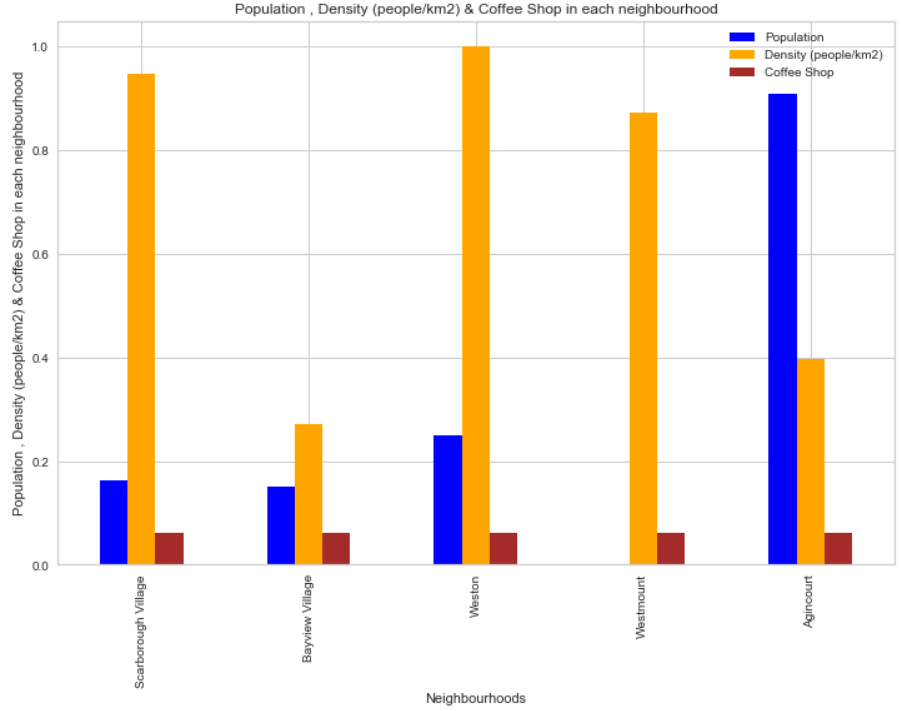




Although population is less in cluster5, Thorncliffe Park has high density. So, Coffee Shop can be started.

**Cluster 6**





In cluster6; Scarborough Village, Weston, Westmount have high density compared to population .BayView has enough number of Coffee Shops. Agincourt has high population and comparable density.So, Coffee Shop can be started in ScarboroughVillage, Weston, Westmount and also Agincourt.

**5. Conclusions**

In this study, I used clusters to find locations where I can start up a Coffee Shop based on population and density of neighbourhoods. After examining all the clusters, few locations are suitable to start Café.They are Humber Summit, Parkwoods , Victoria Village, Woburn, Thorncliffe Park, Scarborough Village, Weston, Westmount, BayView and Agincourt.

**6. Future directions**

I could find locations only in 3 boroughs and also the dataframe is small due to lack of source availability. In future I would like to work on more boroughs to find few more possibilities and also add tips,ratings to compare Coffee Shops as well as to outline the design and working of a new Coffee Shop.