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, the postal codes of all the neighbourhoods is extracted and demographics are extracted from Demographics of Toronto neighbourhoods. 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)

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

2) ll_data (Geospatial_Coordinates.csv)

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

3) pop_den (Demographics of Toronto neighbourhoods)

	Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
0	Crescent Town	EY	0190.01	8157.0	0.40	20393.0	-10.0	23021.0	24.5	20.3	Bengali (18.1%)	18.1% Bengali	NaN
1	Governor's Bridge/Bennington Heights	EY	0186.00	2112.0	1.87	1129.0	4.0	129904.0	7.1	13.3	Polish (1.4%)	01.4% Polish	NaN

4) neigh_venues (FourSquare API)

	Neighbourhood_Name	Neighbourhood_Lat	Neighbourhood_Lng	Venue_Name	Venue_Category	Venue_Lat	Venue_Lng
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	Caribbean Restaurant	43.759840	-79.324719
1	Parkwoods	43.753259	-79.329656	Brookbanks Park	Park	43.751976	-79.332140
2	Parkwoods	43.753259	-79.329656	Tim Hortons	Café	43.760668	-79.326368
3	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	Grocery Store	43.746143	-79.324630
4	Parkwoods	43.753259	-79.329656	A&W	Fast Food Restaurant	43.760643	-79.326865

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

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

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

	Postal Code	Borough	Neighbourhood_Name
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

2.3 Feature Engineering and Feature Transformation

ll_data is merged with df data and resultant is toronto_df.

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242

A dataframe neigh_pop_den is formed by Toronto_df and pop_den.

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	Rente
0	M3A	North York	Parkwoods	43.753259	-79.329656	NY	0270.01, 0270.02, 0271.01, 0271.02, 0272.01, 0...	26533.0	4.96	5349.0	-3.7	34811.0	14.0	22
1	M4A	North York	Victoria Village	43.725882	-79.315572	NY	0261.00, 0262.01, 0262.02	17047.0	4.72	3612.0	2.5	29657.0	15.6	23
2	M3B	North York	Don Mills	43.745906	-79.352188	NY	0263.02, 0263.03, 0263.04, 0268.00, 0269.02	21372.0	8.99	2377.0	11.9	47515.0	10.8	18
3	M3C	North York	Don Mills	43.725900	-79.340923	NY	0263.02, 0263.03, 0263.04, 0268.00, 0269.02	21372.0	8.99	2377.0	11.9	47515.0	10.8	18

In neigh_venues, dummies are generated from the column 'Venue_Category' and neigh_venues_onehot is formed.

	Neighbourhood_Name	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Art Gallery	Art Museum
0	Parkwoods	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Parkwoods	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Parkwoods	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Parkwoods	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Parkwoods	0	0	0	0	0	0	0	0	0	0	0	0	0

neigh_venues_group is formed grouping by 'Neighbourhood_Name'.

	Neighbourhood_Name	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Art Gallery	Art Museum
0	Agincourt	0	0	0	0	0	0	1	0	0	0	0	0	0
1	Alderwood, Long Branch	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bathurst Manor, Wilson Heights, Downsview North	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Bayview Village	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Bedford Park, Lawrence Manor East	0	0	0	0	0	0	1	0	0	0	0	0	0

Now, merged is formed by merging neigh_pop_density(formed by dropping unrequired columns) and neigh_venues_group .

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude	Population	Density (people/km2)	Average Income	Second most common language (after English) by name	ATM	Accessories Store	Afghan Restaurant	African Restaurant	A
0	M3A	North York	Parkwoods	43.753259	-79.329656	26533.0	5349.0	34811.0	Unspecified Chinese (3.4%)	1	0	0	0	

neigh_pop_den_venues is formed by dropping unrequired columns from merged.

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude	Population	Density (people/km2)	Café	Coffee Shop
0	M3A	North York	Parkwoods	43.753259	-79.329656	26533.0	5349.0	1	1
1	M4A	North York	Victoria Village	43.725882	-79.315572	17047.0	3612.0	0	2
2	M3B	North York	Don Mills	43.745906	-79.352188	21372.0	2377.0	2	6
3	M3C	North York	Don Mills	43.725900	-79.340923	21372.0	2377.0	2	6
4	M1G	Scarborough	Woburn	43.770992	-79.216917	48507.0	3636.0	0	2

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

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude	Population	Density (people/km2)	Coffee Shop
0	M3A	North York	Parkwoods	43.753259	-79.329656	26533.0	5349.0	2
1	M4A	North York	Victoria Village	43.725882	-79.315572	17047.0	3612.0	2
2	M3B	North York	Don Mills	43.745906	-79.352188	21372.0	2377.0	8

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

	Postal Code	Borough	Neighbourhood_Name	Latitude	Longitude	Population	Density (people/km2)	Coffee Shop
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.484783	0.754347	0.125
1	M4A	North York	Victoria Village	43.725882	-79.315572	0.262368	0.403154	0.125

2.5 Feature Selection

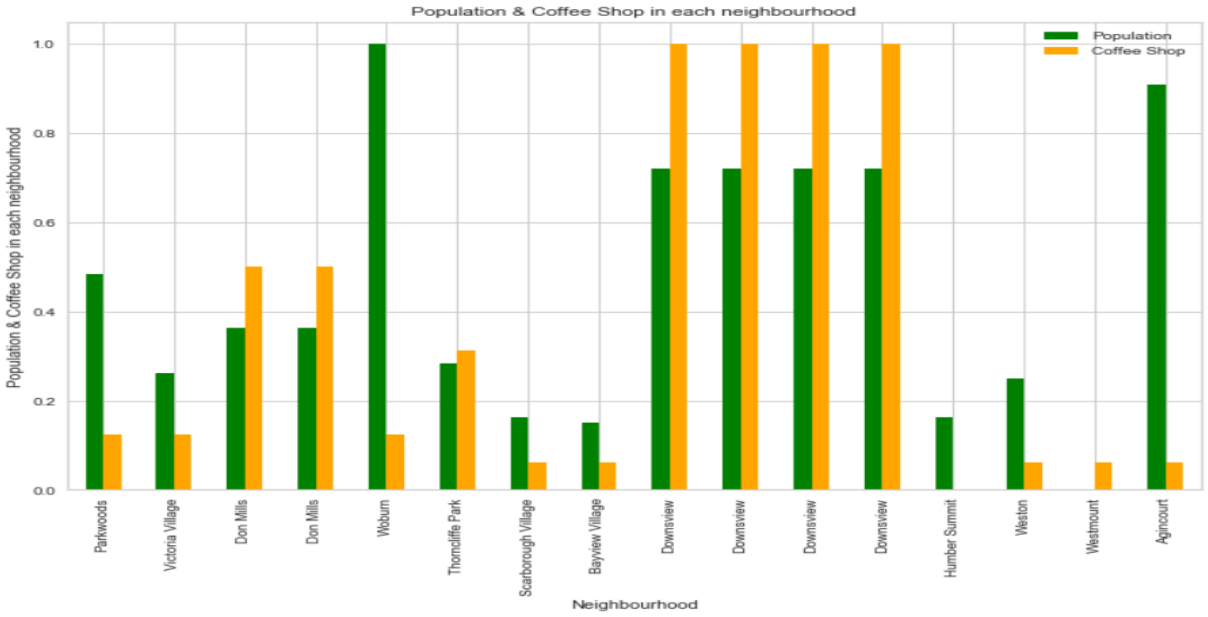
After cleaning,transformation,engineering and also by dropping all the unrequired columns, a dataframe neigh_pop_den_venues is formed with the columns : Postal Code, Borough, Neighbourhood_Name, Latitude, Longitude, Population, Density (people/km2), Coffee Shop.

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

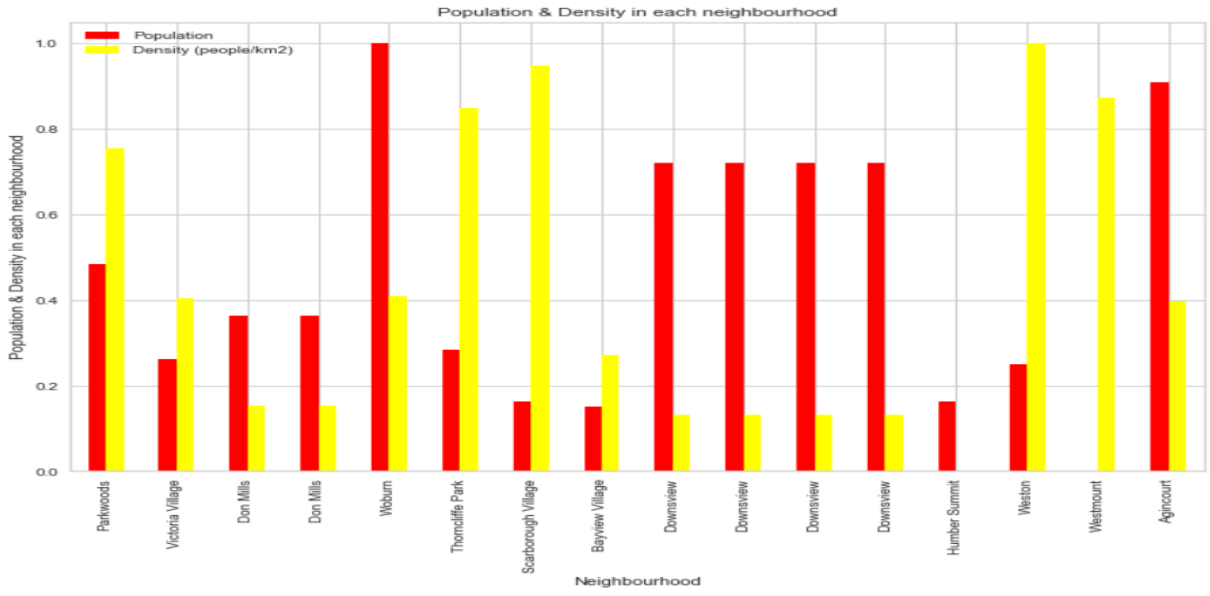
3. Exploratory Data Analysis

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



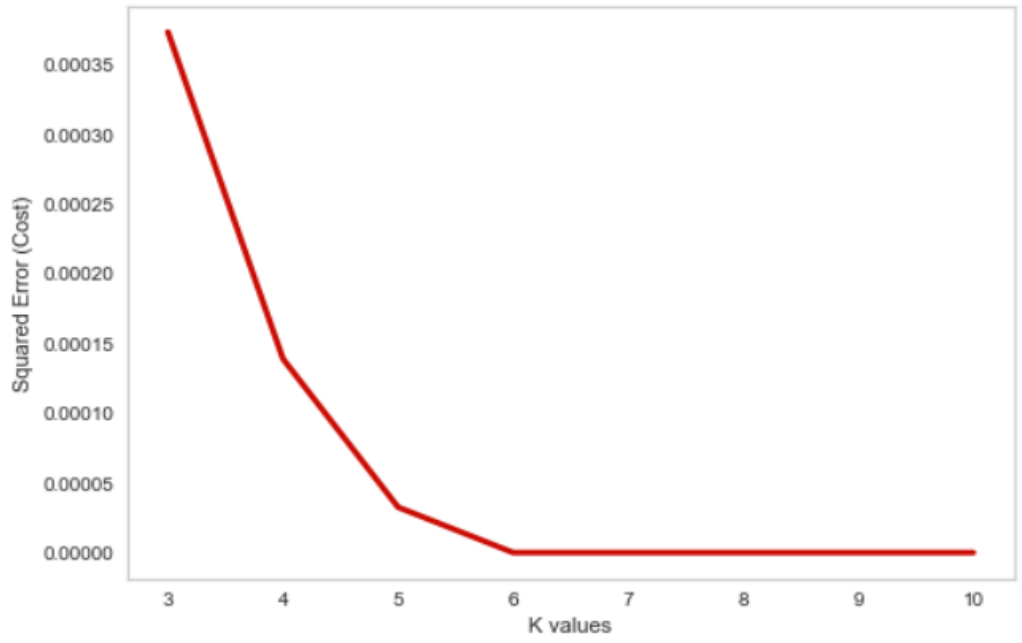
3.2 Relationship between Population and Density



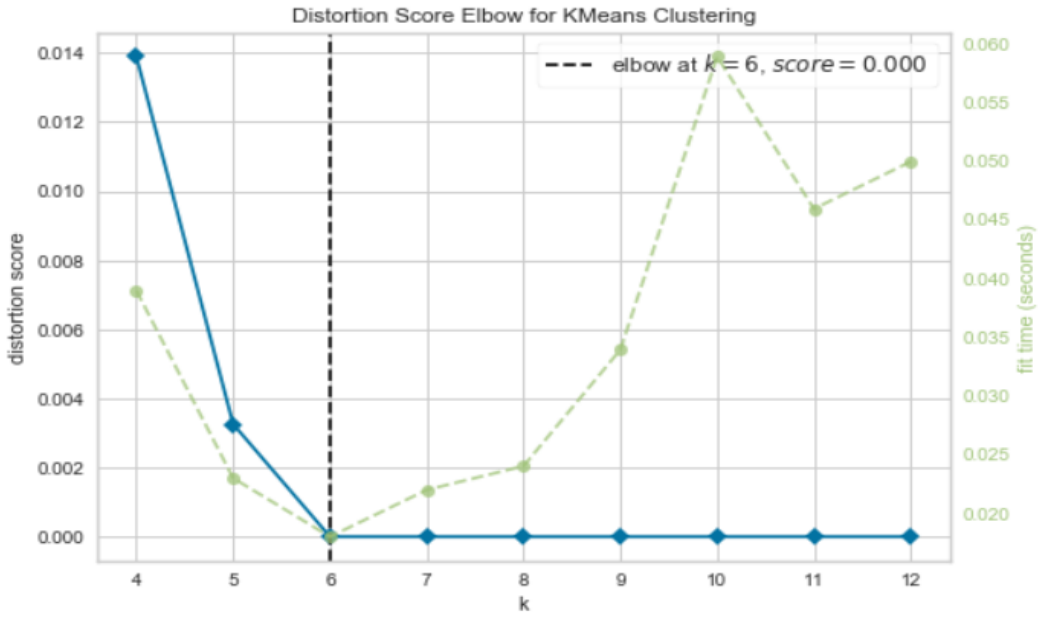
4. Predictive Modeling

4.1 Clustering neighbourhoods of Toronto

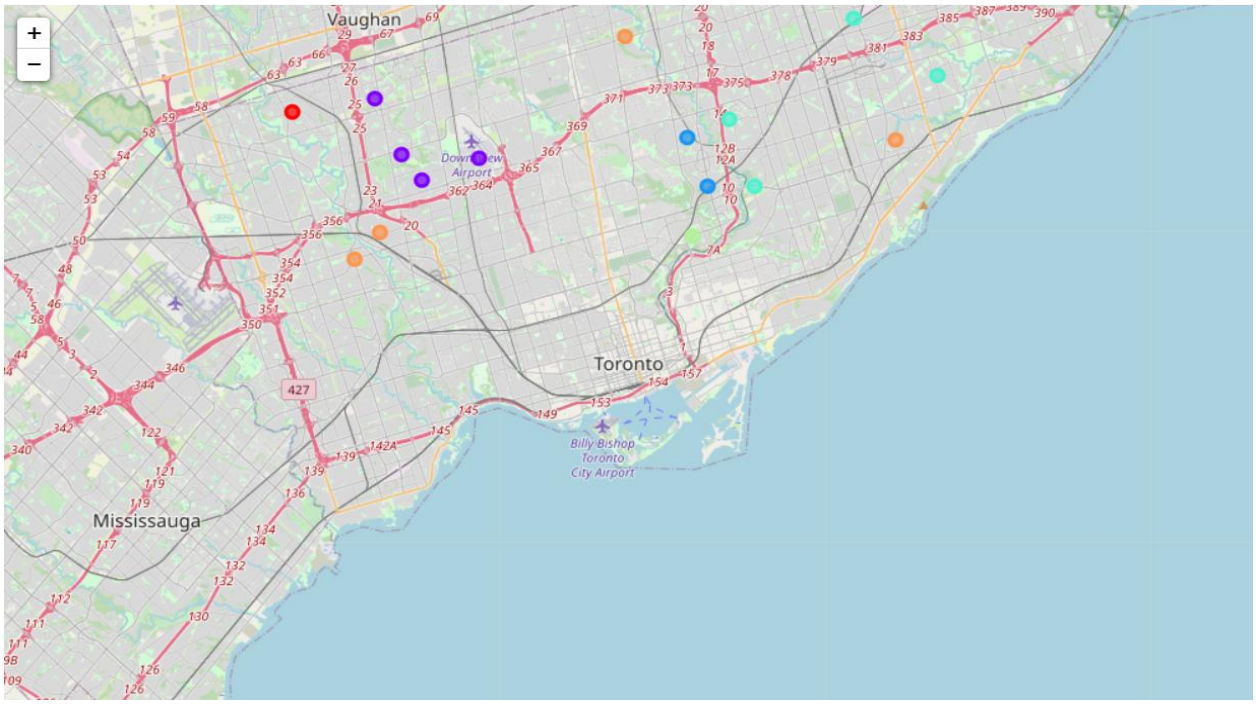
Firstly I identified best k value (number of clusters) to perform clustering and for the same I used elbow method on clustering dataset with Coffee Shop frequencies.



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



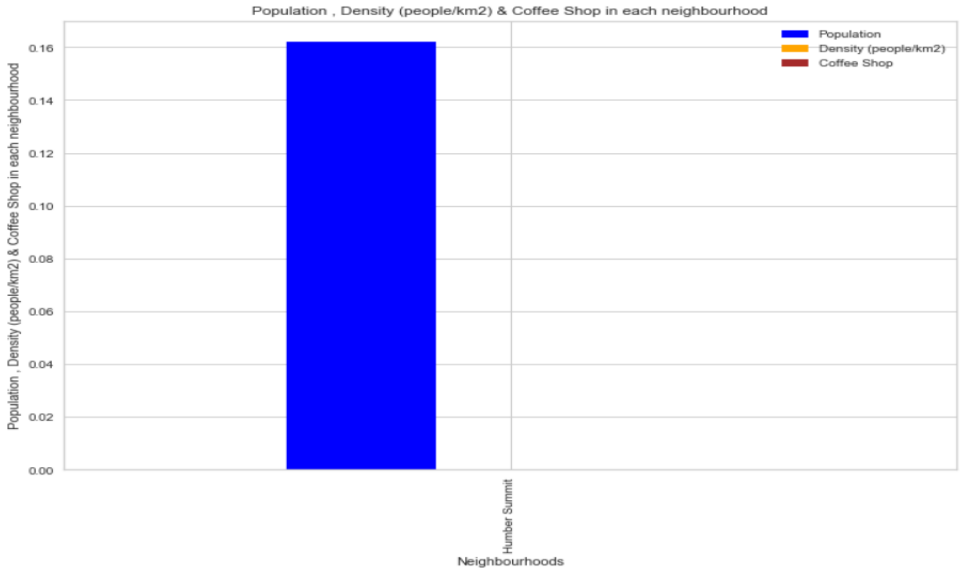
From the above graph , the best k is 6.
After fitting data to KMeans clustering, labels are generated.I used folium map to view the clustered neighbourhoods of Toronto.



4.2 Examine Clusters

Cluster 1

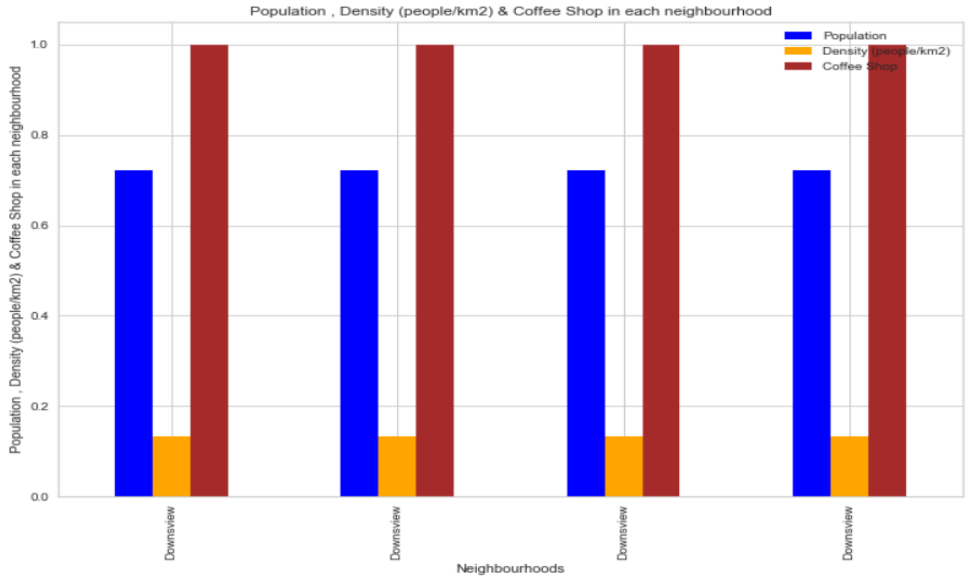
	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
12	M9L	Humber Summit	0.161993	0.0	0.0	0



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

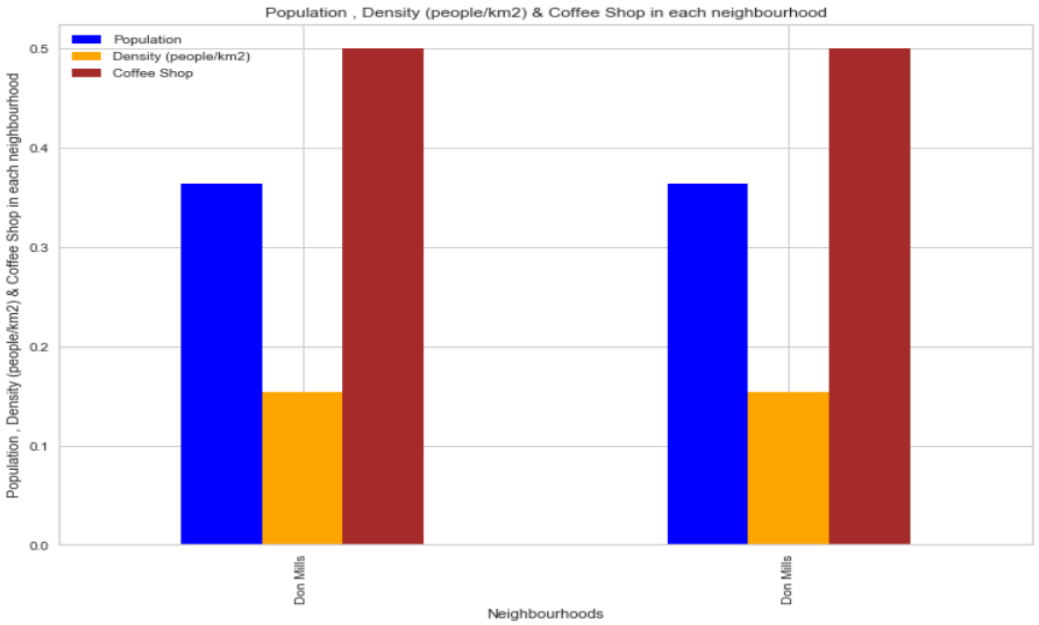
	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
8	M3K	Downsview	0.721125	0.131824	1.0	1
9	M3L	Downsview	0.721125	0.131824	1.0	1
10	M3M	Downsview	0.721125	0.131824	1.0	1
11	M3N	Downsview	0.721125	0.131824	1.0	1



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

Cluster 3

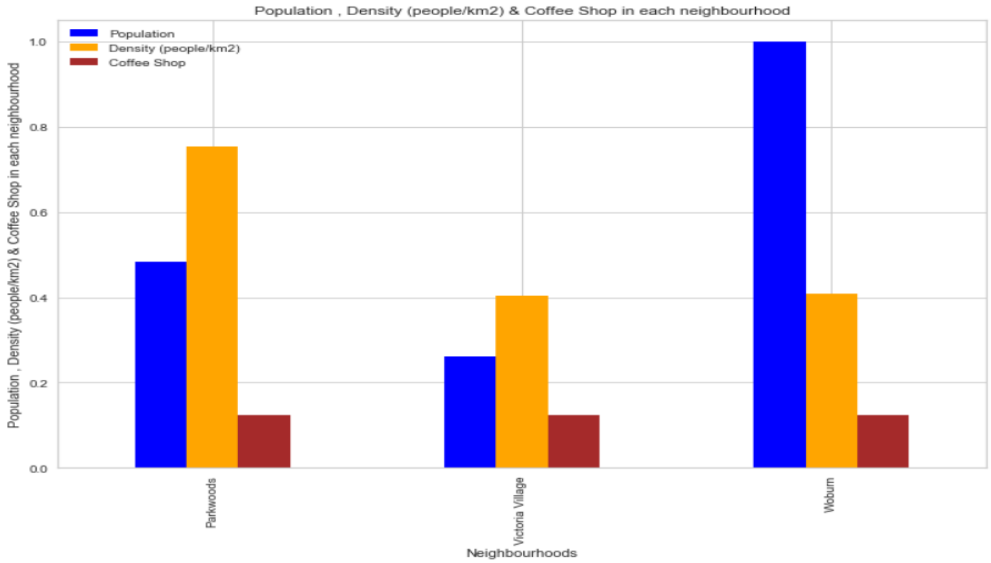
	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
2	M3B	Don Mills	0.363775	0.153457	0.5	2
3	M3C	Don Mills	0.363775	0.153457	0.5	2



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

Cluster 4

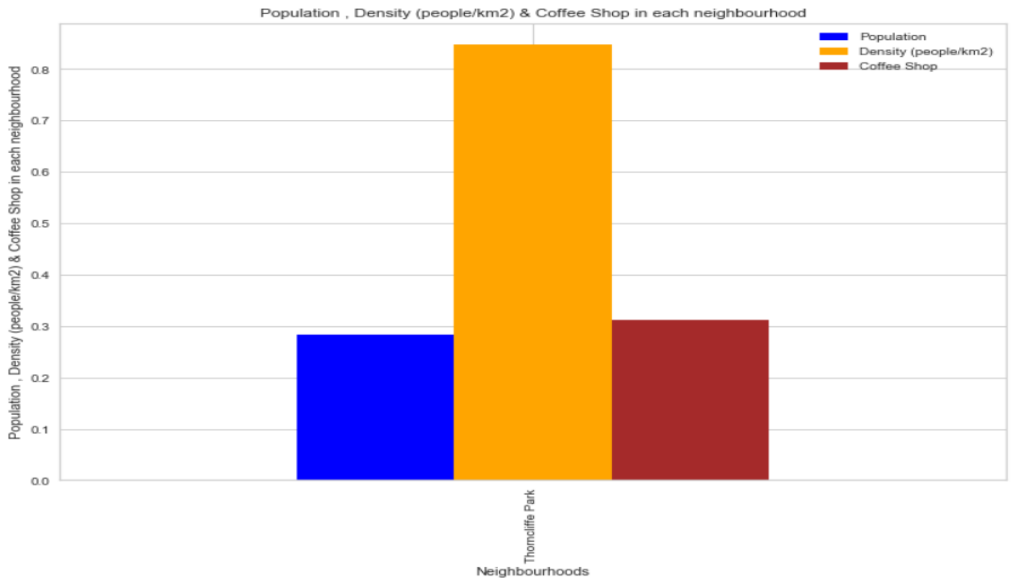
	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
0	M3A	Parkwoods	0.484783	0.754347	0.125	3
1	M4A	Victoria Village	0.262368	0.403154	0.125	3
4	M1G	Woburn	1.000000	0.408006	0.125	3



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

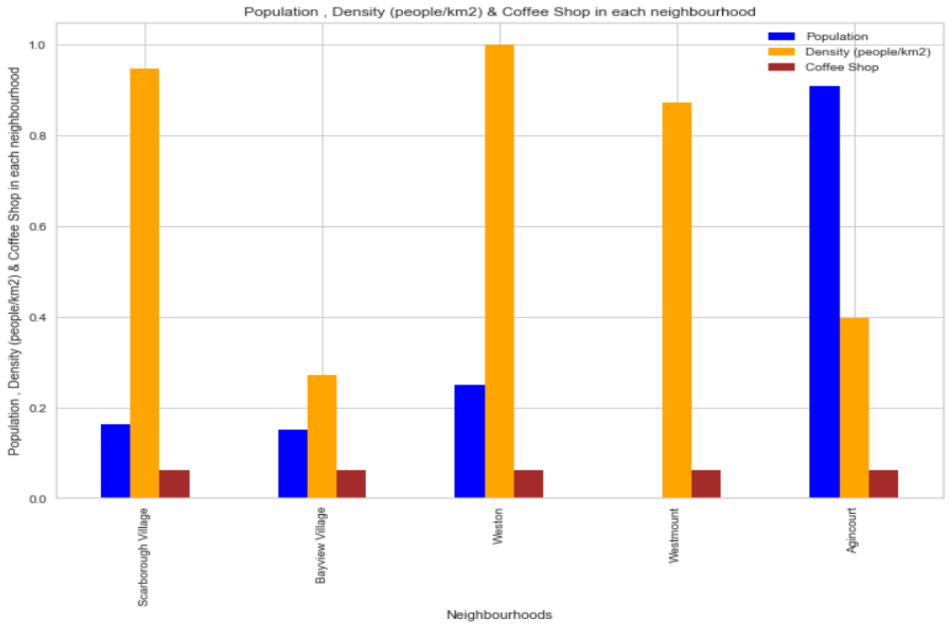
	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
5	M4H	Thorncliffe Park	0.283517	0.847351	0.3125	4



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

Cluster 6

	Postal Code	Neighbourhood_Name	Population	Density (people/km2)	Coffee Shop	Cluster_Labels
6	M1J	Scarborough Village	0.162696	0.947230	0.0625	5
7	M2K	Bayview Village	0.150598	0.272543	0.0625	5
13	M9N	Weston	0.248980	1.000000	0.0625	5
14	M9P	Westmount	0.000000	0.872220	0.0625	5
15	M1S	Agincourt	0.907855	0.396684	0.0625	5



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.