Find Best Place to Open a Cafe in Toronto Canada

ASWINI CHINNENAHALLI SIDDAREDDY

Introduction

- Toronto is capital city of the Ontario, Canada with a population of more than 6 Million

- Toronto is international center for business, finance, arts, and culture, and is recognized as one of the most multicultural cities in the world

- One of my friend is planning to open a Cafe business in Toronto, Canada region.

- Finding a location is one of the most important things in starting a cafe.

- The location search could take months if we start searching and analyzing the area manually.

- We can reduce this time to few hours/ days by using machine learning techniques and Four-square location data to find best suitable location to open a Cafe.

Data Acquisition

- Toronto data that contains Borough, Neighborhoods along with there latitudes and longitudes is obtained from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Geographical location of the neighborhoods is obtained from https://cocl.us/Geospatial_data
- Venue Data is obtained using Foursquare API

Data Cleaning

- Toronto data scraped from Wikipedia page has Boroughs that were not assigned to any neighborhood.

- The following assumptions were made to clean the Toronto Wikipedia dataset.

  - Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned.

  - More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.

  - If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

Data Cleaning

We merged the neighborhood dataset and geographical coordinates dataset into single dataset and the merged dataset is shown below.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
1	M4P	Central Toronto	Davisville North	43.712751	-79.390197
2	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678
3	M4S	Central Toronto	Davisville	43.704324	-79.388790
4	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160

Exploratory Data Analysis

We have grouped the neighborhoods by borough and the count of neighborhoods by borough can be seen below.

```
Borough
Central Toronto      9
Downtown Toronto    19
East Toronto         5
East York            5
Etobicoke           12
Mississauga          1
North York          24
Scarborough         17
West Toronto         6
York                 5
Name: Neighborhood, dtype: int64
```

Exploratory Data Analysis

We then grouped the venues by neighborhood to see the total number of venues in each neighborhood.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	3	3	3	3	3	3
Alderwood, Long Branch	9	9	9	9	9	9
Bathurst Manor, Wilson Heights, Downsview North	23	23	23	23	23	23
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	23	23	23	23	23	23
...	...	...	...	...	...	...
Willowdale, Willowdale East	34	34	34	34	34	34
Willowdale, Willowdale West	5	5	5	5	5	5
Woburn	3	3	3	3	3	3
Woodbine Heights	7	7	7	7	7	7
York Mills West	3	3	3	3	3	3

Exploratory Data Analysis

Group the data by neighborhood to see average number of Café's present in each neighborhood.

	Neighborhoods	Café
0	Agincourt	0.000000
1	Aldenwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000
3	Bayview Village	0.250000
4	Bedford Park, Lawrence Manor East	0.043478

Exploratory Data Analysis

We have clustered the neighborhoods based on the similar number of cafés in that neighborhood using K-means clustering

We have used elbow technique to find the best k value that neither overfit nor under fits our model
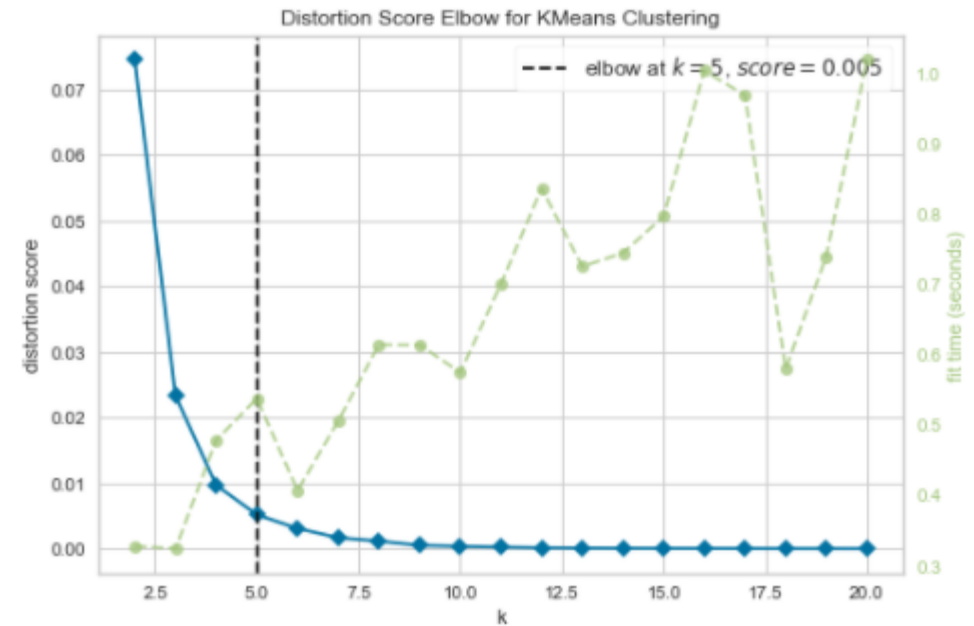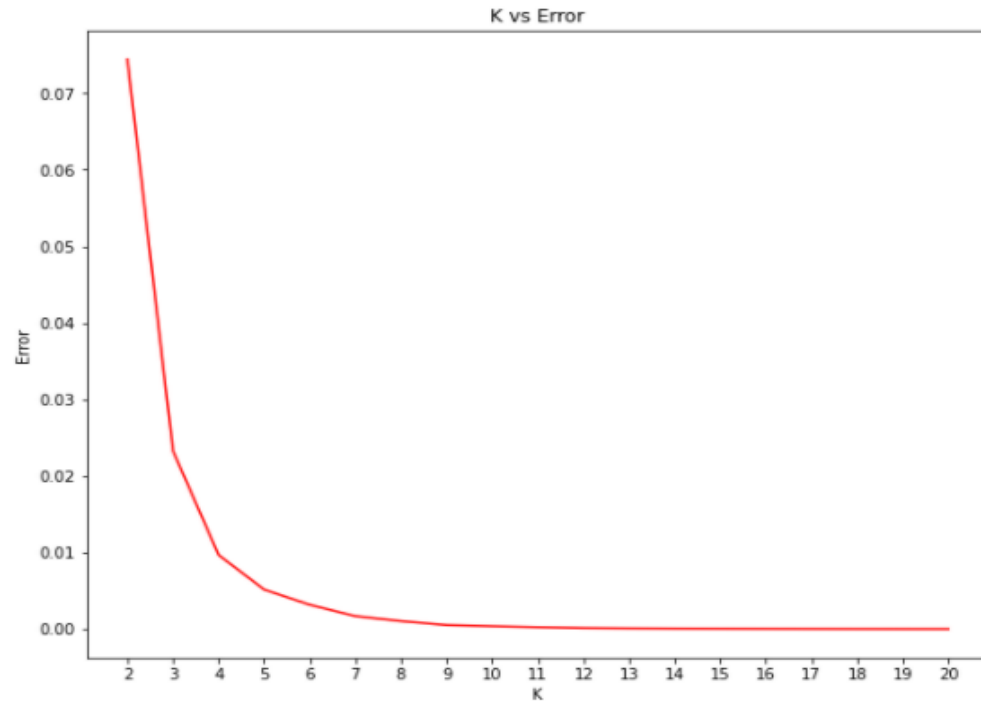
Exploratory Data Analysis

We will put the similar objects of a variable in the same cluster in k-means clustering.

 The neighborhoods that has similar frequency of café's were added to 5 clusters and are labelled from 0 to 4.

	Neighborhood	Café	Cluster Labels
0	Agincourt	0.000000	0
1	Alderwood, Long Branch	0.000000	0
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0
3	Bayview Village	0.250000	2
4	Bedford Park, Lawrence Manor East	0.043478	4

Analysis

We analyze the five clusters by finding the total number of neighborhoods in each cluster and average number of cafés in that cluster

We create a bar chart to see the total number of neighborhoods per cluster and average number of cafés per cluster
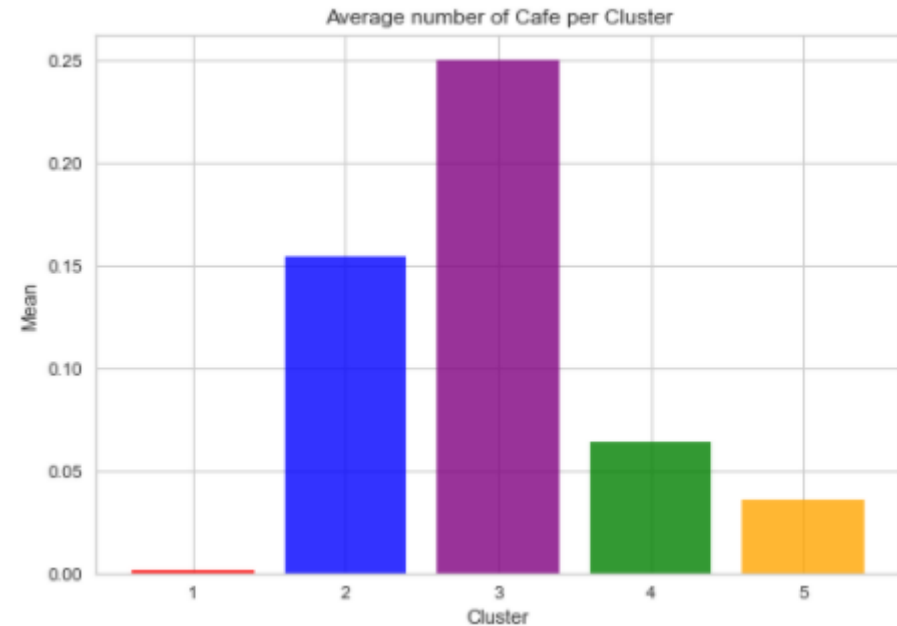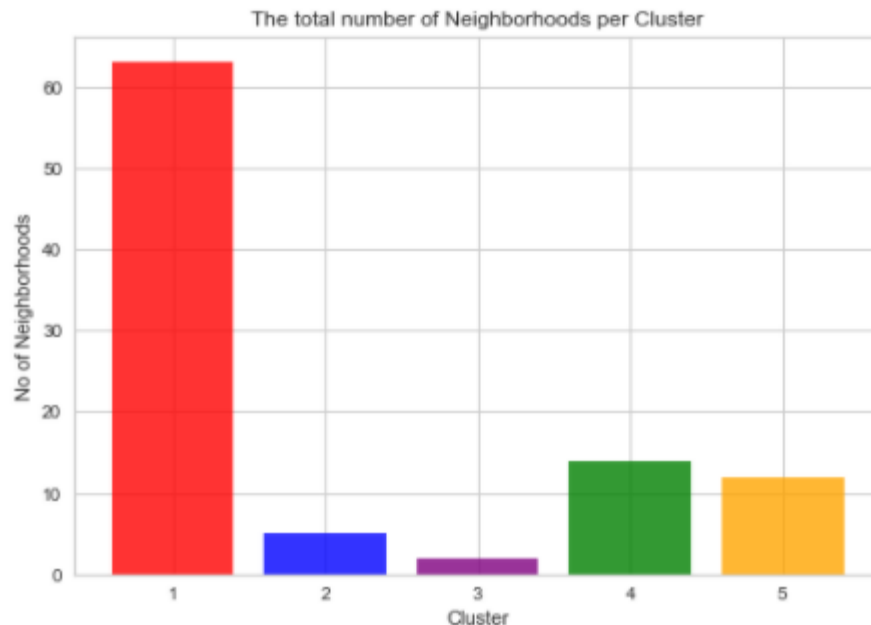
Analysis

Let us analyze the individual clusters by looking at their data frames.

Cluster 1:

	Borough	Neighborhood	Café	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Toronto	Lawrence Park	0.0	0	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Central Toronto	Lawrence Park	0.0	0	43.728020	-79.388790	HYC Design Inc.	43.726793	-79.391681	Business Service
2	Central Toronto	Lawrence Park	0.0	0	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
3	Central Toronto	Lawrence Park	0.0	0	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
4	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	Summerhill Market North	43.715499	-79.392881	Food & Drink Shop
5	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	Winners	43.713236	-79.393873	Department Store
6	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	Best Western Roehampton Hotel & Suites	43.708878	-79.390880	Hotel
7	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	Subway	43.708474	-79.390674	Sandwich Place
8	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	900 Mount Pleasant - Residents Gym	43.711671	-79.391767	Gym / Fitness Center
9	Central Toronto	Davisville North	0.0	0	43.712751	-79.390197	Provocative Pizza Series	43.708293	-79.389546	Pizza Place

- We can see that cluster 1 is in Central Toronto area and it has 65 neighborhoods.

- We have 22 unique venue locations in this cluster and there are very few cafés.

- Cluster 1 has highest number of neighborhoods.

Analysis

Cluster 2:

	Borough	Neighborhood	Café	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
177	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.10	1	43.643515	-79.577201	Carsbrooke Park	43.647051	-79.576682	Park
186	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.10	1	43.643515	-79.577201	LCBO	43.642099	-79.576592	Liquor Store
185	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.10	1	43.643515	-79.577201	Starbucks	43.641312	-79.576924	Coffee Shop
184	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.10	1	43.643515	-79.577201	The Beer Store	43.641313	-79.576925	Beer Store
183	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.10	1	43.643515	-79.577201	Shoppers Drug Mart	43.641312	-79.576924	Pharmacy
...	...	...	...	...	...	...	...	...	...	...
138	Downtown Toronto	First Canadian Place, Underground city	0.07	1	43.648429	-79.382280	Bosk at Shangri-La	43.649023	-79.385826	Asian Restaurant
139	Downtown Toronto	First Canadian Place, Underground city	0.07	1	43.648429	-79.382280	Bulldog On The Block	43.650652	-79.384141	Coffee Shop
140	Downtown Toronto	First Canadian Place, Underground city	0.07	1	43.648429	-79.382280	Petit Four Bakery	43.647744	-79.379588	Sandwich Place
141	Downtown Toronto	First Canadian Place, Underground city	0.07	1	43.648429	-79.382280	Shangri-La Toronto	43.649129	-79.386557	Hotel
123	Downtown Toronto	First Canadian Place, Underground city	0.07	1	43.648429	-79.382280	Maman	43.648309	-79.382253	Café

- We can see that cluster 2 is in Etobicoke, Downtown Toronto, etc. area and Erin gate, Bloor dale Gardens, Old Bumhamthrope, First Canadian Place, Underground city etc., are few neighborhoods in that cluster.

- We have 19 unique venue locations in this cluster and there are a greater number of Cafés in this cluster when compared to cluster 1.

Analysis

Cluster 3:

	Borough	Neighborhood	Café	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Bayview Village	0.25	2	43.786947	-79.385975	Kaga Sushi	43.787758	-79.381090	Japanese Restaurant
1	North York	Bayview Village	0.25	2	43.786947	-79.385975	Maxim's Cafe and Patisserie	43.787863	-79.380751	Café
2	North York	Bayview Village	0.25	2	43.786947	-79.385975	TD Canada Trust	43.788074	-79.380367	Bank
3	North York	Bayview Village	0.25	2	43.786947	-79.385975	Sun Star Chinese Cuisine 翠景小炒	43.787914	-79.381234	Chinese Restaurant
4	Scarborough	Birch Cliff, Cliffside West	0.25	2	43.692657	-79.264848	Birchmount Community Centre	43.695175	-79.262161	General Entertainment
5	Scarborough	Birch Cliff, Cliffside West	0.25	2	43.692657	-79.264848	Scarborough Gardens	43.694647	-79.262230	Skating Rink
6	Scarborough	Birch Cliff, Cliffside West	0.25	2	43.692657	-79.264848	Birchmount Stadium	43.695323	-79.261293	College Stadium
7	Scarborough	Birch Cliff, Cliffside West	0.25	2	43.692657	-79.264848	The Birchcliff	43.691666	-79.264532	Café

- We can see that cluster 3 is in North York, Scarborough area and Bayview Village, Birch Cliff, Cliffside West are few neighborhoods in that cluster.

- We have 8 unique venue locations in this cluster and there are highest number of cafés in this cluster.

Analysis

Cluster 4:

	Borough	Neighborhood	Café	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Toronto	North Toronto West, Lawrence Park	0.055556	3	43.715383	-79.405678	Blo Blow Dry Bar	43.714712	-79.400227	Health & Beauty Service
1	Central Toronto	North Toronto West, Lawrence Park	0.055556	3	43.715383	-79.405678	St. Clements - Yonge Parkette	43.712062	-79.404255	Park
2	Central Toronto	North Toronto West, Lawrence Park	0.055556	3	43.715383	-79.405678	Second Cup	43.714583	-79.400120	Café
3	Central Toronto	North Toronto West, Lawrence Park	0.055556	3	43.715383	-79.405678	GAP	43.715450	-79.400089	Clothing Store
4	Central Toronto	North Toronto West, Lawrence Park	0.055556	3	43.715383	-79.405678	Tim Hortons	43.714894	-79.399776	Coffee Shop
...	...	...	...	...	...	...	...	...	...	...
1182	West Toronto	Little Portugal, Trinity	0.047619	3	43.647927	-79.419750	Rotate This	43.648544	-79.420518	Record Shop
1183	West Toronto	Little Portugal, Trinity	0.047619	3	43.647927	-79.419750	Bazara	43.648535	-79.420521	Japanese Restaurant
1184	West Toronto	Little Portugal, Trinity	0.047619	3	43.647927	-79.419750	Get Well	43.649320	-79.421937	Bar
1185	West Toronto	Little Portugal, Trinity	0.047619	3	43.647927	-79.419750	The Communist's Daughter	43.649362	-79.420963	Bar
1186	West Toronto	Little Portugal, Trinity	0.047619	3	43.647927	-79.419750	Lost & Found	43.649378	-79.424149	Men's Store

- We can see that cluster 4 is in Central Toronto, West Toronto area and North Toronto West, Lawrence Park, Little Portugal, Trinity are few neighborhoods in that cluster.

- We have 56 unique venue locations in this cluster and there are only few cafés. Cluster 4 has third highest average for Café's.

Analysis

Cluster 5:

	Borough	Neighborhood	Café	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Toronto	The Annex, North Midtown, Yorkville	0.157895	4	43.672710	-79.405678	The Annex Hodgepodge	43.674975	-79.406543	Sandwich Place
1	Central Toronto	The Annex, North Midtown, Yorkville	0.157895	4	43.672710	-79.405678	Fet Zun	43.675147	-79.406346	Middle Eastern Restaurant
2	Central Toronto	The Annex, North Midtown, Yorkville	0.157895	4	43.672710	-79.405678	Toronto Archives	43.676447	-79.407509	History Museum
3	Central Toronto	The Annex, North Midtown, Yorkville	0.157895	4	43.672710	-79.405678	Dish Cooking Studio	43.674066	-79.410764	Café
4	Central Toronto	The Annex, North Midtown, Yorkville	0.157895	4	43.672710	-79.405678	Krispy Kreme Doughnut Cafe	43.674732	-79.407730	Donut Shop
...	...	...	...	...	...	...	...	...	...	...
84	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.136364	4	43.636847	-79.428191	Uma Cafe	43.637016	-79.424048	Café
85	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.136364	4	43.636847	-79.428191	Queen Supermarket	43.634265	-79.429112	Grocery Store
86	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.136364	4	43.636847	-79.428191	Cinema Nightclub	43.637614	-79.423890	Nightclub
87	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.136364	4	43.636847	-79.428191	Gardiner Expy & Dufferin St	43.633800	-79.425460	Intersection
88	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.136364	4	43.636847	-79.428191	Dunn Milk Variety	43.633667	-79.431549	Convenience Store

- We can see that cluster 5 is in Central Toronto and West Toronto area and The Annex, North Midtown, Yorkville are few neighborhoods in this cluster.

- We have 26 unique venue locations in this cluster and there are very few cafés.

Conclusion and Future Directions

- We have used several python libraries, Four Square API to find the best area to open a Café in Toronto, Canada

- We created several visualizations to better understand our dataset and analysis using seaborn, matplotlib libraries

- By using all these libraries, machine learning techniques, we are able say that the best place to open new café is Central Toronto area as there are many neighborhoods in that area and little to no cafés

- In this study, we mainly focused on analyzing neighborhoods, venue details. However, population, other demographic details might also contribute to evaluate Café business success criteria.

- By adding demographic data into analysis, we can increase the confidence and accuracy of our analysis in finding best place to open Café in Toronto.