Battle of the Neighbourhood

Introduction

Toronto is the capital city of Canadian province of Ontario and the most populous city in Canada. With an estimated population of over 6.2 million, there are diversity of different ethnic groups living in the city. Toronto is a cosmopolitan city that is well known of a diverse cuisine. Downtown Toronto is the heart of the city that brings many opportunities for entrepreneurs to start their business.

The objective of this project is to use Foursquare location data and regional clustering of venue to determine the best neighbourhood in Toronto to open a restaurant. With a population of 500k Italians, there are ample of opportunities to open a new Pizza Place. Pizza is a type of fast food that enables to accommodate to the larger number of busy commuters in the city during the peak hours.

Target Audience

Entrepreneur who wants to open a Pizza Place or franchise

Data Overview:

- **Neighbourhood list:** Wikipedia
- **Geographical Location Data:** Geocoder package
- **Venue Data:** Foursquare

Methodology

The data is extracted from the following sources.

The first source of data Wikipedia provided the postal code, borough and name of Neighbouhood in Toronto. Scrapping was performed on the site and transformed into Pandas dataframe.

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 1: Neighbourhood data from Wikipedia

The second source of data consists of the geographical coordinates of the neighbourhoods with the postal codes.

	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

Figure 2: Geographical data of Neighbourhoods in Toronto

The third source of data is collected via Foursquare explore API to obtain the various venue data. It requires a secret key and client id to pull any data. The data are grouped by the neighbourhoods to perform data clustering.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant

Figure 3: Foursquare Venue Data

The following assumptions were made for Boroughs that were not assigned to any neighbourhood

1. Borough that is not assigned will be ignored
2. There can be more than one neighbourhood that exist in one postal code area. M1B has two neighbourhood – Malvern and Rouge and they will be combined into one row with a comma.
3. If the cell has a borough but not assigned neighbourhood, the neighbourhood will be the same as Borough

After the implementation of the assumptions, the rows were grouped based on the borough below.

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 4: Borough Grouping

By merging the Borough group and Geocode package, the neighbourhood data will consist of the latitude and longitude.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 5: Merged table with Postal Code

Foursquare API is used to get the list of all the Venues in Toronto to analyse the number of Pizza Places all over Toronto. There was a total of 50 Pizza Places in Toronto. The Foursquare Venue data is merged with Neighbourhood data to give the nearest Venue for each of the Neighbourhoods.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant

Figure 6: Local Venues based on respective Neighbourhood

One hot encoding is used to transform categorical data into numerical data. Based on each neighbourhood, individual venues were turned into the frequency at how many Venues were located at each neighbourhood.

	Neighbourhoods	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Malvern, Rouge	0	0	0	0	0	0	0	0
1	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	0	0	0
2	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0

Figure 7: One Hot Encoding

The average frequency of occurrence of each Venue Category were taken based on each Neighbourhood.

	Neighbourhoods	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04

Figure 8: Average frequency of each venue based on grouped neighbourhoods

A new data frame was created to analyse the data for Pizza Places in the Neighbourhood.

	Neighbourhood	Pizza Place
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.250000
2	Bathurst Manor, Wilson Heights, Downsview North	0.047619
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.040000

Figure 9: Pizza Place data frame

K-Means clustering is used to cluster the neighbourhoods that have similar averages of Pizza Places in the Neighbourhood. The Elbow Point Technique will obtain the optimal K value to prevent overfitting or underfitting. A test was ran using different K values and accuracy value. The best K value of 4 was chosen because it is the point that has the sharpest turn. Hence, there will be 4 clusters of Pizza Places.

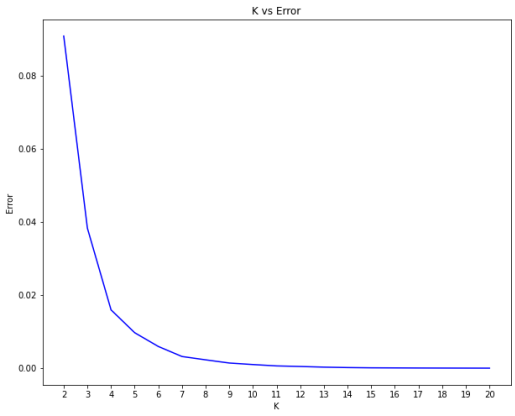

Figure 10: K vs Error

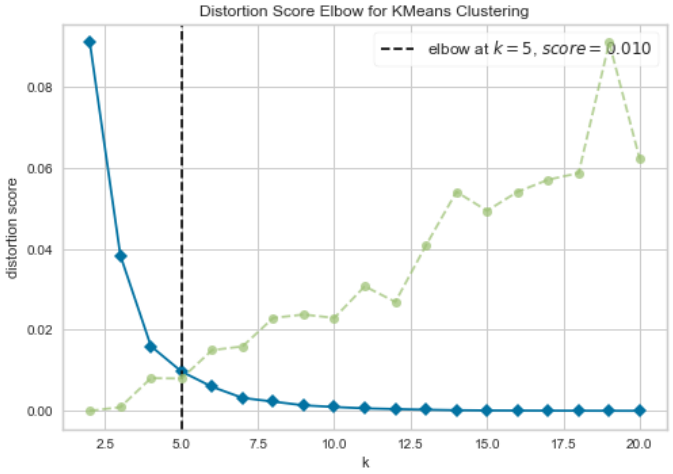

Figure 11: KMeans Clustering using Elbow Point

Based on the dotted line, the Elbow is at K=5. In K-Means clustering, neighbourhoods that have similar mean frequency of Pizza Places were divided into 5 clusters. Each of the clusters were labelled from 0 to 4 as the indexing of labels begin with 0 instead of 1.

	Neighbourhood	Pizza Place	Cluster Labels
0	Agincourt	0.000000	0
1	Alderwood, Long Branch	0.250000	2
2	Bathurst Manor, Wilson Heights, Downsview North	0.047619	1
3	Bayview Village	0.000000	0
4	Bedford Park, Lawrence Manor East	0.040000	1

Figure 12: Cluster Labels

Using Folium, a map was created to show each neighbourhood based on each of the cluster label. This will analyse opportunities in opening a new Pizza Place in Toronto.

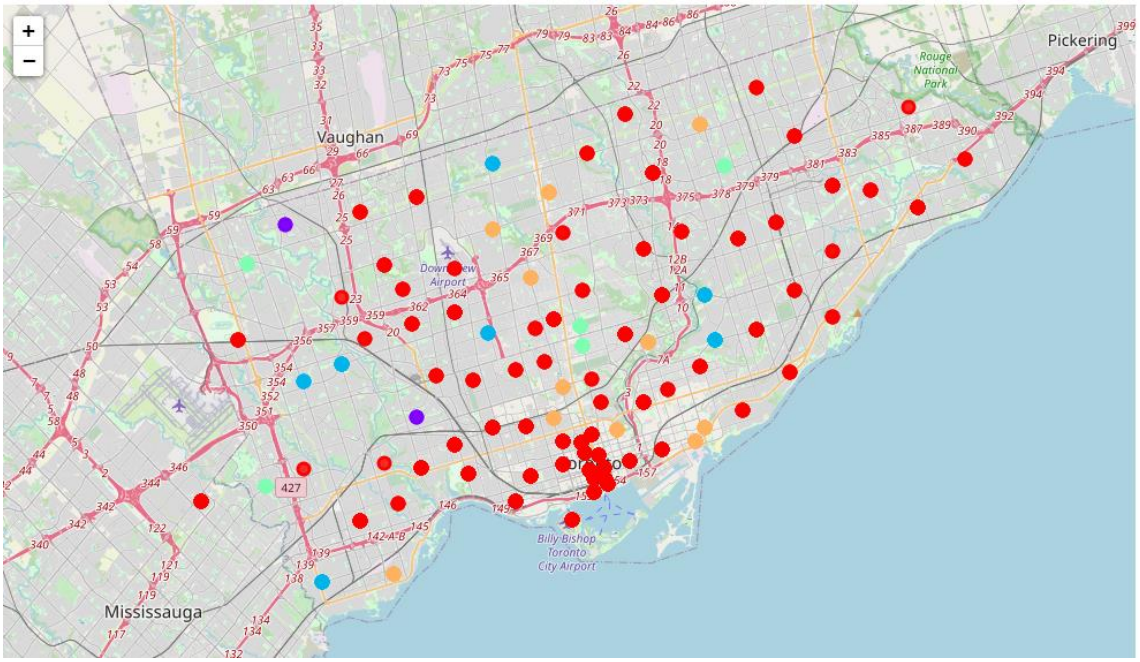

Figure 13: Map with 5 Clusters

Cluster Analysis

With a total of 5 clusters, a bar graph was used to analyse the number of neighbourhoods in each cluster. Cluster 1 has the highest number of neighbourhoods of 74 and cluster 3 has the lowest number of neighbourhoods of 1.

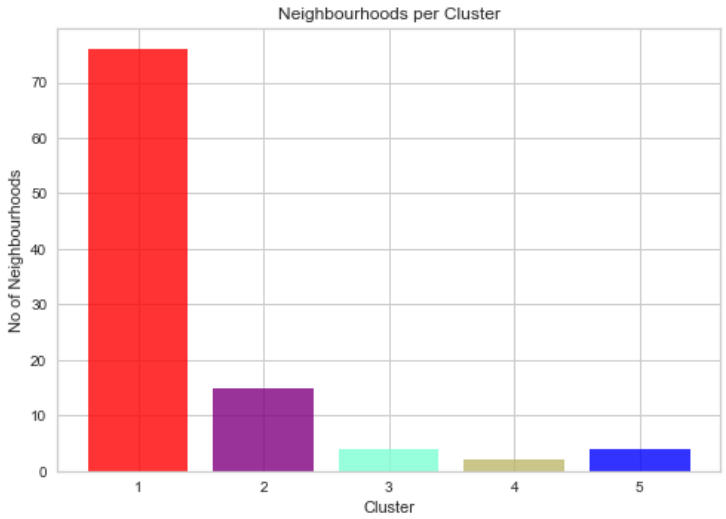

Figure 14: Neighbourhoods per Cluster

Even though cluster 1 has the highest number of neighbourhoods, it has the lowest average number of Italian restaurants of 0.01. It can also be observed that neighbourhoods in cluster 1 is sparsely populated. However, cluster 2 has the highest average number of Italian restaurants of 0.5.

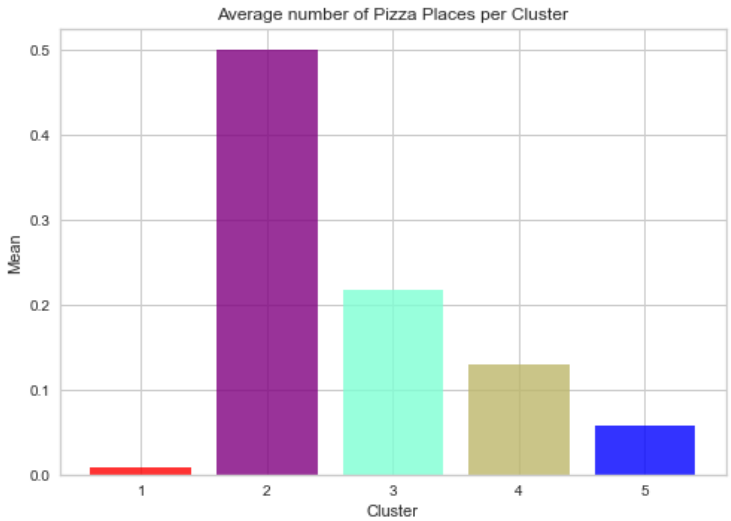

Figure 14: Average number of Pizza Places per cluster

Cluster 1

Cluster 1 is concentrated in Downtown Toronto area. It has 76 Neighbourhoods and 256 venue categories. Although it has the highest number of neighbourhoods, the neighbourhoods are sparsely population, and it has the least average of Pizza Places of 0.01.

	Borough	Neighbourhood	Pizza Place	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Parkwoods	0.0	0	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
1	North York	Parkwoods	0.0	0	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
4	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop

Cluster 2

Cluster 2 is concentrated in North York and East York area. It has 2 Neighbourhoods, 3 venue categories and 2 Pizza Places. The average number of Pizza Places is the highest of 0.5. However, the data size in the cluster is too small as it only has 2 records.

	Borough	Neighbourhood	Pizza Place	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Humber Summit	0.5	1	43.756303	-79.565963	Pizza Monza	43.755043	-79.567195	Pizza Place
1	North York	Humber Summit	0.5	1	43.756303	-79.565963	Faab Fitness	43.758156	-79.570442	Gym
2	York	Runnymede, The Junction North	0.5	1	43.673185	-79.487262	195 Jane Rocket	43.672335	-79.492634	Bus Line
3	York	Runnymede, The Junction North	0.5	1	43.673185	-79.487262	241 Pizza	43.672351	-79.492631	Pizza Place

Cluster 3

Cluster 3 has the second highest average number of Pizza Places of 0.21. It is situated in the North York Area and Etobicoke area. It has 7 Neighbourhood and 8 restaurants.

	Borough	Neighbourhood	Pizza Place	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Victoria Village	0.2	2	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
1	North York	Victoria Village	0.2	2	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
2	North York	Victoria Village	0.2	2	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
3	North York	Victoria Village	0.2	2	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
4	North York	Victoria Village	0.2	2	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place

Cluster 4

Cluster 4 is concentrated in Central Toronto, Etobicoke and Scarborough area. It has 5 Neighbourhoods, 44 Venue Categories and 10 Pizza Places. It has the third highest average number of Italian restaurants of 0.13.

	Borough	Neighbourhood	Pizza Place	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.142857	3	43.643515	-79.577201	Pet Valu	43.641667	-79.577050	Pet Store
1	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.142857	3	43.643515	-79.577201	LCBO	43.642099	-79.576592	Liquor Store
2	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.142857	3	43.643515	-79.577201	Starbucks	43.641312	-79.576924	Coffee Shop
3	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.142857	3	43.643515	-79.577201	The Beer Store	43.641313	-79.576925	Beer Store
4	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.142857	3	43.643515	-79.577201	Shoppers Drug Mart	43.641312	-79.576924	Pharmacy

Cluster 5

Cluster 5 had one of the lowest average number of Pizza Places of 0.06. There was a total of 79 neighbourhood and 258 unique venues. There were 14 Pizza Places within the cluster.

	Borough	Neighbourhood	Pizza Place	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Parkwoods	0.0	0	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
1	North York	Parkwoods	0.0	0	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
4	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop

Hence, the ordering of the average Pizza Places in each cluster are shown below.

Rank	Cluster	Average Pizza Places
1	2	0.5
2	3	0.22
3	4	0.13
4	5	0.06
5	1	0.01

Analysis

Most of the Pizza Places are in cluster 3 that are represented by the purple cluster. The neighbourhood in North York and Etobicoke has the highest average number of Pizza Places are Westmount and Parkview Hill. Even though cluster 1 has the highest number of neighbourhoods, it has the lowest average number of Italian restaurants. Compared to the other clusters, it can be observed that Central Toronto in cluster 4 has the second lowest average number of Pizza Places.

After analysing the clusters, the optimum place to open a new Pizza Place is West Toronto or East Toronto as there are many neighbourhoods in the area but limited Pizza Places. This offers new opportunities to achieve a competitive advantage. The second-best Neighbourhood to open a new Pizza Place will be North York in areas such as Bedford Park or Humber Summit. As shown in cluster 2 and 3, Pizza Places are getting popular in the North York and there are a lower number of Pizza Places compared to other Neighbourhoods.

The drawback of the analysis is that the clustering is based on the data from Foursquare API. It does not take into consideration of the food preference of the population within the neighbourhoods as it will play an essential factor to choose which place to open a Pizza Place. The project can be improved by using more information and different Machine Learning algorithms.

In conclusion, this assignment has provided an opportunity to be a data scientist to work on a complex business problem. Python libraries were utilised for data analysis and data visualisation. The KMeans and KElbow Machine Learning algorithm has been applied to calculate the error. We can use this data set to analyse other situation such as another food and beverage venture or opening a movie theatre.