

Capstone Project - Battle of Neighborhoods

Introduction

Background

East York, Etobicoke, North York, Scarborough, York, and Old Toronto are the 6 boroughs which combine together to form the City of Toronto.

One of these boroughs, Etobicoke is home to several lakefront parks, golf courses, and vast Centennial Park, with a conservatory featuring tropical plants. The 1830s Montgomery's Inn has a museum, tea room, and pub and hosts a weekly farmers' market. Islington - City Centre West area is a busy commercial hub, containing shopping complexes and casual chain eateries, plus history-themed murals along Dundas Street West.

Affordable housing has always been a problem in Toronto. Mr. X is very interested in buying a home in Etobicoke but is unsure about the neighbourhood. He has requested us to find and suggest the best neighbourhood suited as per his needs.

Business Problem

Mr. X is interested in a neighbourhood that meets the below criteria:

- price should be around \$400000 to \$500000
- nearby shopping centre
- nearby restaurants and eateries
- nearby park or green area

We need to find a neighbourhood in Etobicoke which fulfills the above conditions and make our suggestions.

Data Section

The data needed for our analysis will be collected from various sources.

1. Toronto neighbourhood geo location and boundaries -

This dataset will help us identify the latitude and longitude for all the neighbourhood.

Further we will use this data with Foursquare APIs to find nearby information.

We are using the Toronto Open data catalogue to support or needs.

Website url - <https://open.toronto.ca/dataset/neighbourhoods/>

	LONGITUDE	LATITUDE	Neighborhood
0	-79.425515	43.676919	Wychwood
1	-79.403590	43.704689	Yonge-Eglinton
2	-79.397871	43.687859	Yonge-St.Clair
3	-79.488883	43.765736	York University Heights
4	-79.457108	43.714672	Yorkdale-Glen Park

2. Toronto Housing data -

The data came from various sources including Toronto Community Housing Corporation, City of Toronto's Shelter, Support and Housing Administration, City of Toronto Affordable Housing Office and Statistics Canada. Average Home Price data was taken from Realosophy.com.

Website url - <https://open.toronto.ca/dataset/wellbeing-toronto-housing/>

	Neighborhood	Home Prices
0	West Humber-Clairville	317508
1	Mount Olive-Silverstone-Jamestown	251119
2	Thistletown-Beaumont Heights	414216
3	Rexdale-Kipling	392271
4	Elms-Old Rexdale	233832

3. Toronto borough-neighbourhood data -

This wiki page provides a list of neighbourhoods for each borough in Toronto City..

Website url - https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

CDN number	Neighborhood	Borough	Neighbourhoods covered	Map	Unnamed: 5
0	129	Agincourt North	Scarborough	Agincourt and Brimwood	NaN
1	128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern	NaN
2	20	Alderwood	Etobicoke	Alderwood	NaN
3	95	Annex	Old City of Toronto	The Annex and Seaton Village	NaN
4	42	Banbury-Don Mills	North York	Don Mills	NaN

Before we can use this data for our analysis, it must be cleansed. We will identify the neighborhoods that belong to Etobicoke and generate a consolidated dataset for our further analysis.

	Neighborhood	Borough	Longitude	Latitude	Home Prices
2	Alderwood	Etobicoke	-79.541611	43.604937	504233
32	Edenbridge-Humber Valley	Etobicoke	-79.522458	43.670886	873268
34	Elms-Old Rexdale	Etobicoke	-79.548983	43.721519	233832
36	Eringate-Centennial-West Deane	Etobicoke	-79.580445	43.658017	423034
37	Etobicoke West Mall	Etobicoke	-79.568939	43.645063	298426

Methodology

The data is now ready and we can proceed with the analysis. This step will include 2 steps:

Exploratory data analysis - We analyse and select our desired neighborhoods for testing our model. Using the prepared dataset we have identified the below listed 5 neighborhoods which fulfill our pricing criteria.

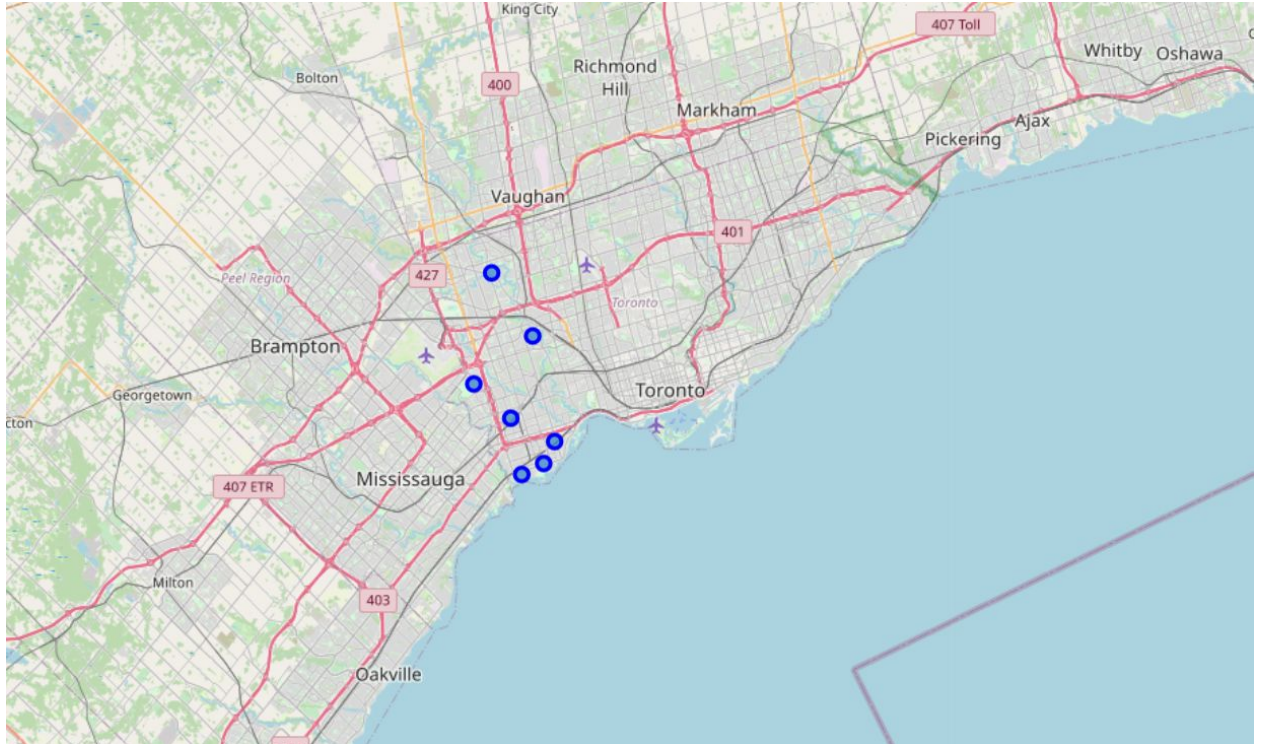
- Eringate-Centennial-West Deane
- Humber Heights-Westmount
- Islington-City Centre West
- Long Branch
- Mimico

Neighborhood	Borough	Longitude	Latitude	Home Prices
Eringate-Centennial-West Deane	Etobicoke	-79.580445	43.658017	423034
Humber Heights-Westmount	Etobicoke	-79.522416	43.692233	491396
Islington-City Centre West	Etobicoke	-79.543317	43.633463	491678
Long Branch	Etobicoke	-79.533345	43.592362	459088
Mimico	Etobicoke	-79.500137	43.615924	429941

Statistical descriptive data: With the help of describe function in python we can get the min, max, mean, and other statistical information about our neighborhoods' home prices.

	Longitude	Latitude	Home Prices
count	7.000000	7.000000	7.000000
mean	-79.536216	43.647225	452880.571429
std	0.028634	0.052811	31628.292755
min	-79.580445	43.592362	414216.000000
25%	-79.553404	43.608256	426487.500000
50%	-79.533345	43.633463	459088.000000
75%	-79.516387	43.675125	476103.500000
max	-79.500137	43.737988	491678.000000

The map below shows the selected 5 neighborhoods in the borough of Etobicoke.



Modelling - We will create a machine learning model using k-mean clustering mechanism and try to identify the most suitable neighbourhood amongst the selected ones. Clustering the neighborhood based on similar venues will let us compare and make better suggestions. We will be creating 5 clusters in our neighborhood.

We have used Foursquare api to get the nearby venues for each neighbourhood. Foursquare api provides JSON data which can be easily converted into Pandas dataframe and used to perform analysis.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Eringate-Centennial-West Deane	43.658017	-79.580445	Pizza Pizza	43.660392	-79.582686	Pizza Place
1	Eringate-Centennial-West Deane	43.658017	-79.580445	Tim Hortons	43.660425	-79.583034	Coffee Shop
2	Eringate-Centennial-West Deane	43.658017	-79.580445	Golden Wok Chinese Restaurant	43.660491	-79.582319	Chinese Restaurant
3	Eringate-Centennial-West Deane	43.658017	-79.580445	Eringate park	43.661668	-79.581093	Park
4	Eringate-Centennial-West Deane	43.658017	-79.580445	Mac's	43.661684	-79.582728	Convenience Store

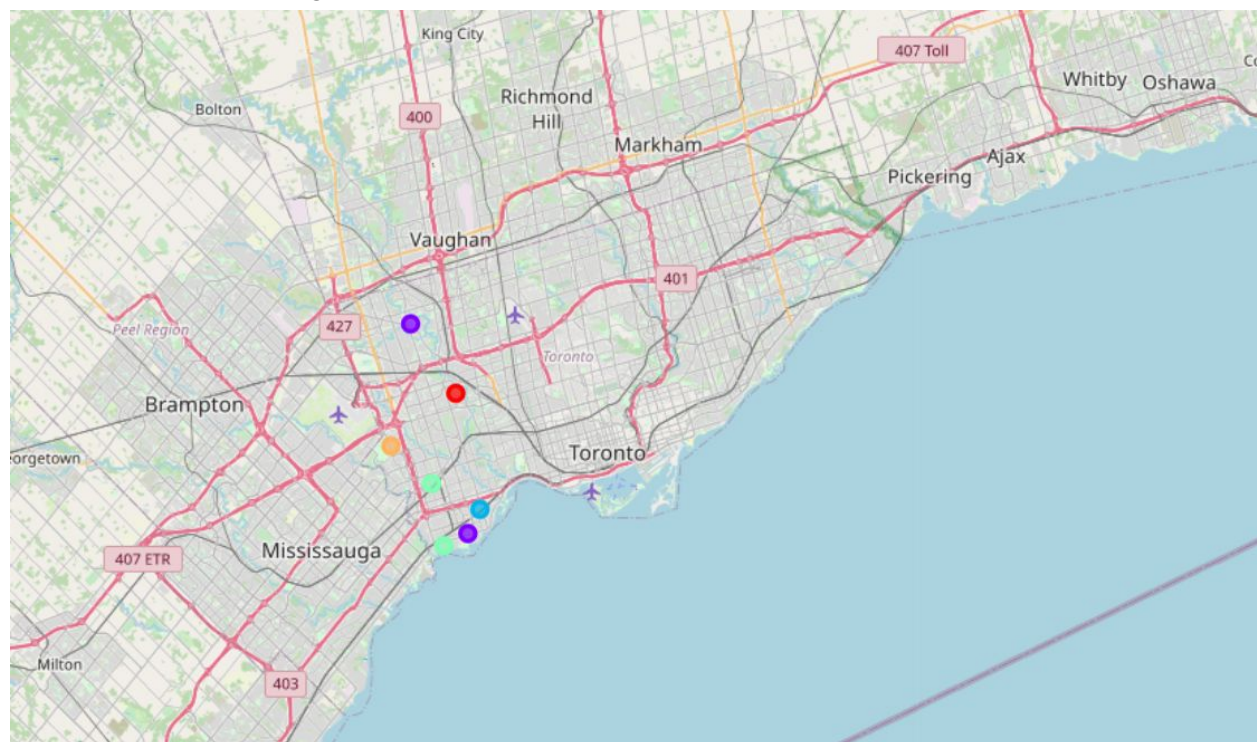
Once we have the venues, we will use one-hot encoding technique to convert them and use them in our machine learning model. The data is then grouped into neighborhoods and a list of 10 common venues is identified in the process.

	Neighborhood	American Restaurant	Asian Restaurant	Bakery	Bank	Bar	Beer Store	Café	Caribbean Restaurant	Chinese Restaurant	Coffee Shop	Convenience Store	Fast Food Restaurant	Fried Chicken Joint	Garden Center	Gas Station
0	Eringate-Centennial-West Deane	0.0	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.166667	0.166667	0.166667	0.000000	0.000000	0.000000	0.000000
1	Humber Heights-Westmount	0.0	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333
2	Islington-City Centre West	0.0	0.0	0.00	0.066667	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.133333	0.066667	0.066667	0.000000
3	Long Branch	0.0	0.0	0.00	0.133333	0.066667	0.066667	0.066667	0.0	0.000000	0.133333	0.000000	0.000000	0.000000	0.000000	0.000000
4	Mimico	0.0	0.0	0.25	0.000000	0.250000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Now that we have all the neighborhood data ready, we will use the k-means clustering technique to create 5 clusters for the neighborhoods. We will then analyse each neighborhood to provide our conclusion.

Neighborhood	Borough	Longitude	Latitude	Home Prices	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Eringate-Centennial-West Deane	Etobicoke	-79.580445	43.658017	423034	4	Chinese Restaurant	Hockey Arena	Convenience Store	Park	Coffee Shop	Pizza Place	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Worship Center
Humber Heights-Westmount	Etobicoke	-79.522416	43.692233	491396	0	Gas Station	Park	Pizza Place	Greek Restaurant	Garden Center	Fried Chicken Joint	Fast Food Restaurant	Convenience Store	Coffee Shop	Worship Center
Islington-City Centre West	Etobicoke	-79.543317	43.633463	491678	3	Restaurant	Fast Food Restaurant	Women's Store	Garden Center	Greek Restaurant	Ice Cream Shop	Pizza Place	Bank	Fried Chicken Joint	Sanc
Long Branch	Etobicoke	-79.533345	43.592362	459088	3	Grocery Store	Bank	Coffee Shop	Restaurant	Greek Restaurant	Wings Joint	Italian Restaurant	Pharmacy	Pizza Place	
Mimico	Etobicoke	-79.500137	43.615924	429941	2	Grocery Store	Bakery	Bar	Skating Rink	Chinese Restaurant	Garden Center	Fried Chicken Joint	Fast Food Restaurant	Convenience Store	C

The map shows the neighborhood clusters.



Results

The clusters are analysed to get a better understanding of the neighborhood.

Cluster 1:

Cluster 1 satisfies all the criteria laid by Mr. X. Although parks and eateries are sufficient, it only provides one store for Women. It doesn't provide a full shopping experience to its residents.

Cluster 1

```
etobioke_merged.loc[etobioke_merged['Cluster Labels'] == 0, etobioke_merged.columns[[1] + list(range(5, etobioke_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
49	Etobicoke	0	Gas Station	Park	Pizza Place	Greek Restaurant	Garden Center	Fried Chicken Joint	Fast Food Restaurant	Convenience Store	Coffee Shop	Women's Store

Cluster 2:

Cluster 2 looks like a suitable match for Mr. X. It has a Park and also provides a lot of different restaurant options. It has supermarkets and convenience stores to meet most of the daily needs. Its also supported by availability of banks and pharmacies.

Cluster 2

```
etobioke_merged.loc[etobioke_merged['Cluster Labels'] == 1, etobioke_merged.columns[[1] + list(range(5, etobioke_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
80	Etobicoke	1	Pub	Coffee Shop	Supermarket	Indian Restaurant	Italian Restaurant	Park	Caribbean Restaurant	Fast Food Restaurant	Convenience Store	Chinese Restaurant
110	Etobicoke	1	Indian Restaurant	Caribbean Restaurant	American Restaurant	Pharmacy	Coffee Shop	Ice Cream Shop	Pizza Place	Bank	Supermarket	Thai Restaurant

Cluster 3:

Cluster 3 doesn't have a park and is thus not considered as an option for Mr. X

Cluster 3

```
etobioke_merged.loc[etobioke_merged['Cluster Labels'] == 2, etobioke_merged.columns[[1] + list(range(5, etobioke_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
73	Etobicoke	2	Grocery Store	Bakery	Bar	Skating Rink	Chinese Restaurant	Garden Center	Fried Chicken Joint	Fast Food Restaurant	Convenience Store	Coffee Shop

Cluster 4:

Cluster 4 also provides a wide variety of facilities to the residents. It has a lot to offer on eateries but doesn't have a park which is on Mr. X's priority criteria and hence fails our recommendation

Cluster 4

```
etobioke_merged.loc[etobioke_merged['Cluster Labels'] == 3, etobioke_merged.columns[[1] + list(range(5, etobioke_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
54	Etobicoke	3	Restaurant	Fast Food Restaurant	Women's Store	Garden Center	Greek Restaurant	Ice Cream Shop	Pizza Place	Bank	Fried Chicken Joint	Sandwich Place
68	Etobicoke	3	Grocery Store	Bank	Coffee Shop	Restaurant	Greek Restaurant	Wings Joint	Italian Restaurant	Pharmacy	Pizza Place	Café

Cluster 5:

Cluster 5 includes all the criteria laid by Mr. X. It also offers gas stations as an added advantage.

Cluster 5

```
etobioke_merged.loc[etobioke_merged['Cluster Labels'] == 4, etobioke_merged.columns[[1] + list(range(5, etobioke_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
36	Etobicoke	4	Chinese Restaurant	Hockey Arena	Convenience Store	Park	Coffee Shop	Pizza Place	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Women's Store

Discussion

The aim of this project was to identify and suggest a neighborhood area where Mr.X could buy a home as per his budget. In this project we tried to identify the neighborhoods where houses were affordable enough. We created neighborhood clusters to compare them and get a better insight. We also used foursquare data to help us understand the surrounding of those neighborhood clusters. In this process we realised we have cluster 1, 2 and 5 which look suitable for Mr. X preference, each having their own advantages.

Conclusion

Although Cluster 1 and 5 satisfy all the criterias from Mr. X, as compared to Cluster 2 they provide limited options to the residents. Cluster 2 has a variety of restaurants to choose from, has parks and stores. It also has an added advantage of having banks and pharmacies in the area.

Our suggestion would be to buy a home in cluster 2 neighborhood. Cluster 1 and 5 can also be considered if needed.