Presentation

WHERE TO OPEN A PIZZA PLACE IN ETOBICOKE

1.1 Introduction

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016,[1] it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario,[2] while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

The 140 neighborhoods used by the City of Toronto were developed to help government and community organizations with their local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighborhoods do not change over time, allowing researchers to examine changes over time. In order to ensure high quality social data, the neighborhoods were defined based on Statistics Canada Census Tract boundaries.

1.2. Business Problem

The purpose of this project is to examine the most suitable and convenient for somebody who wants to open an Italian Pizza Place in Etobicoke borough, Toronto.

The location of the Pizza Place should be in one of the Eobicoke neighborhoods where no Pizza Place is present or a limited of such places, but there is enough population to sustain such business, including people with Italian descend.

2. Data to determine best location for Pizza Place in Etobicoke, Toronto

* For solving our business problem, the following data processing was required:
* Exploring and clustering the neighborhoods in Toronto, based on the following Wikipedia page, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, in order to obtain the data required to build our table of postal codes and to transform the data into a suitable format that we will use to interpret and provide a solution. This will be done using Beautiful Soup package.
* Importing the populations for each Toronto borough from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Tables/File.cfm?T=1201&SR=1&RPP=9999&PR=0&CMA=0&CSD=0&S=22&O=A&Lang=Eng&OFT=CSV.
* Importing average income after tax for each neighborhood from stat Canada: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/search-recherche/change-geo.cfm?Lang=E&Geo1=FSA
* Importing etnic information from https://www.toronto.ca/wp-content/uploads/2018/05/972c-City_Planning_2016_Census_Profile_2014_Wards_Ward01.pdf
* Using Foursquare location data to extract the information with the most common venues, fit them into clusters, and determine the best outcome cluster using k-means clustering.
* Using geopy library to get the latitude and longitude values of Toronto and Etobicoke.
* Using **Folium** for great visualization into the generated map, and click on each circle mark to reveal the name of the neighborhood and its respective borough.
* Simplifying the generated Etobicoke map and segment and cluster only the neighborhoods in ETOBICOKE.

3. Methodology

BeautifulSoup package is used to transform the data from the table on the Wikipedia page into a panda data frame. Only the cells that have an assigned borough will be processed and we ignore cells with a borough that is Not assigned. The rows with same postal code will be combined into one row with the neighborhoods separated with a comma.

We will manually download and clean the data with the census population from 2016 and joining the neighborhood postal code data with the population data.

From the Stats Canada Website we have obtained the information that Canadian families and unattached individuals had a median after-tax income of $57,000 in 2016.

	PostalCode	Borough	Neighborhood	Population_2016
0	M3A	North York	Parkwoods	34615.0
1	M4A	North York	Victoria Village	14443.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	41078.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	21048.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	10.0

	PostalCode	AfterTaxIncome2015	Borough	Neighborhood	Population_2016
66	M2P	115237.0	North York	York Mills West	7843.0
55	M5M	111821.0	North York	Bedford Park, Lawrence Manor East	25975.0
61	M4N	109841.0	Central Toronto	Lawrence Park	15330.0
74	M5R	108271.0	Central Toronto	The Annex, North Midtown, Yorkville	26496.0
97	M8X	97210.0	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	10787.0

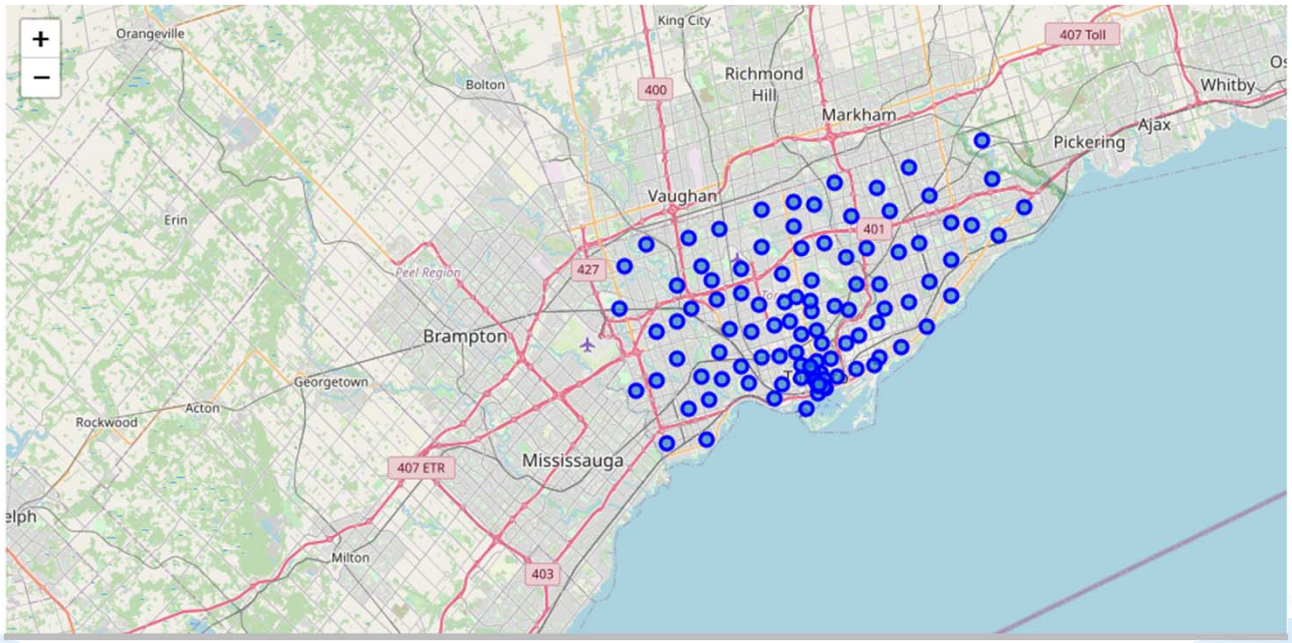
With the help of geopy library to get the latitude and the longitude coordinates of each neighborhood.

	Neighborhood	Population_2016	AfterTaxIncome2015	Latitude	Longitude
4	The Kingsway, Montgomery Road, Old Mill North	10787.0	97210.0	43.653654	-79.506944
5	West Deane Park, Princess Gardens, Martin Grov...	32400.0	91110.0	43.650943	-79.554724
1	Mimico NW, The Queensway West, South of Bloor,...	17038.0	78903.0	43.628841	-79.520999
0	Kingsview Village, St. Phillips, Martin Grove ...	33743.0	67497.0	43.688905	-79.554724
3	Old Mill South, King's Mill Park, Sunnylea, Hu...	21299.0	63142.0	43.636258	-79.498509
2	Northwest, West Humber - Clairville	40684.0	59873.0	43.706748	-79.594054

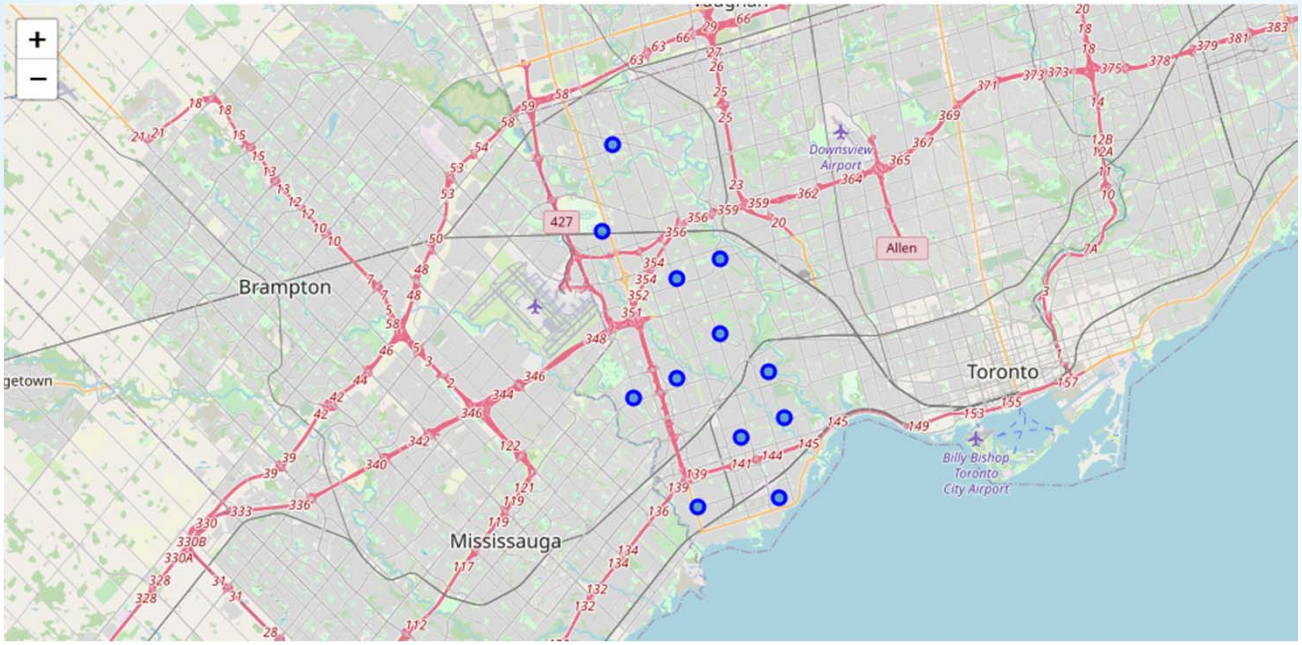
We are ready to scrape now the demographic ethnic table from Wikipedia. Some ethnic information missing was also added into a csv file, to allow us to further interpret the ethnic demographic data.

Neighborhood	Population	Ethnic Origin #1	Percentage #1	Ethnic Origin #2	Percentage #2	Ethnic Origin #3	Percentage #3	Ethnic Origin #4	Percentage #4	Ethnic Origin #5	Percentage #5	Ethnic Origin #6
Etobicoke-Lakeshore	127520.0	English	17.1	Canadian	15.9	Irish	14.4	Scottish	13.5	Polish	9.2	Italian
Etobicoke North	116960.0	East Indian	22.2	Jamaican	6.2	Canadian	5.7	Iraqi	4.8	Italian	3.9	Filipino
Etobicoke Centre	116055.0	Italian	15.1	English	14.3	Canadian	12.1	Irish	10.8	Scottish	10.4	Ukrainian

With the help of Folium- Python visualization library we will create a map of Toronto with neighborhoods superimposed on top that will be used to visualize the neighborhoods cluster distribution of Etobicoke over an interactive leaflet map.



However, for illustration purposes, let's simplify the above map and segment and cluster only the neighborhoods in ETOBICOKE, which is one of many Toronto's borough. So let's slice the original data frame and create a new data frame of the Etobicoke neighborhood.



Our work is now concentrated in analyzing the Etobicoke neighborhoods regarding venues and what type of categories are specific to which area. The mean of the frequency of occurrence of each category will help to group by categories.

For this step we will create a new data frame containing only these neighborhoods, as seen on the right picture.

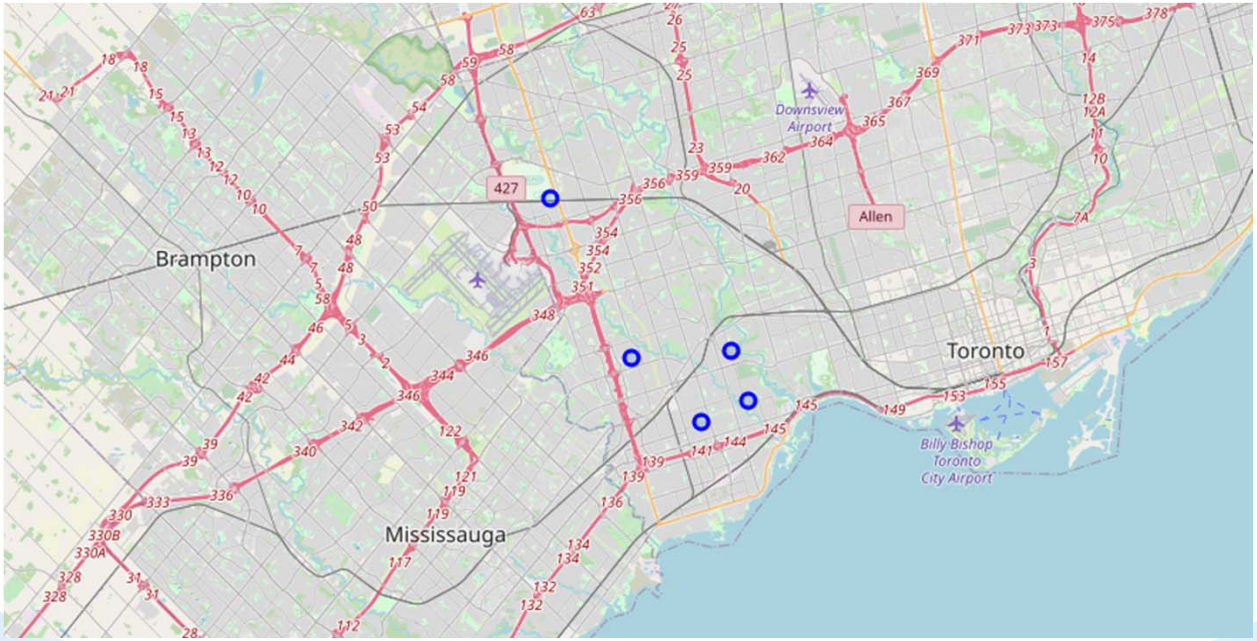
	Neighborhood
4	Mimico NW, The Queensway West, South of Bloor,...
6	Northwest, West Humber - Clairville
7	Old Mill South, King's Mill Park, Sunnylea, Hu...
9	The Kingsway, Montgomery Road, Old Mill North
10	West Deane Park, Princess Gardens, Martin Grov...

As we can see there are quite few neighborhoods in Etobicoke with Pizza Places and that make them very desirable for people to spend time in these areas.

With a new panda data frame created for neighborhoods with no Pizza places created, we need to merge it with our data frame containing the latitude and longitude.

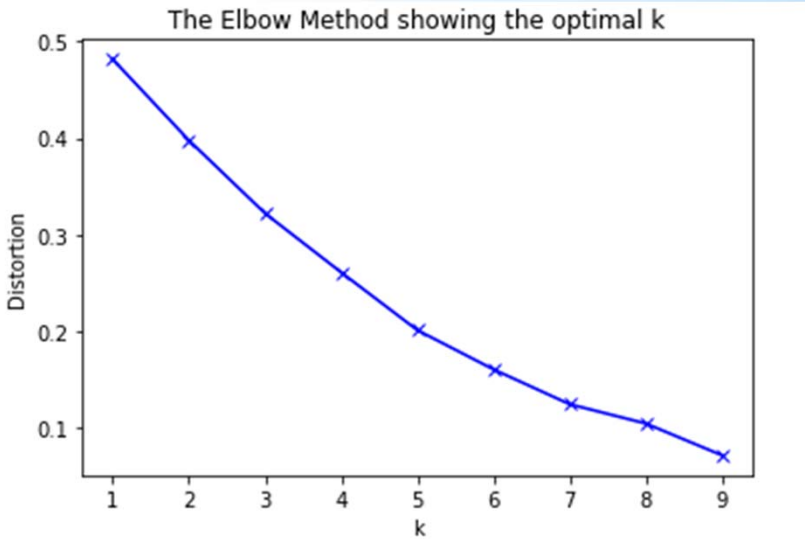
	Neighborhood	Population_2016	AfterTaxIncome2015	Latitude	Longitude
3	The Kingsway, Montgomery Road, Old Mill North	10787.0	97210.0	43.653654	-79.506944
4	West Deane Park, Princess Gardens, Martin Grov...	32400.0	91110.0	43.650943	-79.554724
0	Mimico NW, The Queensway West, South of Bloor,...	17038.0	78903.0	43.628841	-79.520999
2	Old Mill South, King's Mill Park, Sunnylea, Hu...	21299.0	63142.0	43.636258	-79.498509
1	Northwest, West Humber - Clairville	40684.0	59873.0	43.706748	-79.594054

Using folium library we will plot the neighborhoods with no Pizza Places.



Next we will apply unsupervised machine learning algorithr mean to cluster the different categories of Etobicoke neighborhoods. These clusters will be analyzed to allow us drawing conclusions with the help of other characteristics well, related to ethnic demographic, neighborhood popula numbers and also the presence or not of Pizza Places.

Unsupervised machine learning algorithm K-mean clusterin would be applied to form the clusters of different categori places residing in and around the neighborhoods. These clu from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to der the conclusions.



The Elbow Method is more of a decision rule, while the Silhouette is a metric used for validation while clustering. Thus, it can be used in combination with the Elbow Method.

The Elbow Method and the Silhouette Method are not alternatives to each other for finding the optimal K. Rather they are tools to be used together for a more confident decision.

We are using now the Elbow method of determining the number cluster that will use to cluster the data in.

We can observe that the “elbow” is not clearly defined. In the next step, we will use the Silhoute method to look for an optimum k-mean.

Now, to find the optimal value of k for KMeans, loop through 1..n for n_clusters in KMeans and calculate Silhouette Coefficient for each sample.

A higher Silhouette Coefficient indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.

After running the python sequence for determining the Silhouette coefficient, the below result was generated.

Below is shown a table with the most 10 venues for each neighborhood, including the cluster column as well.

Below is shown a table with the most 10 venues for each neighborhood, including the cluster column as well.

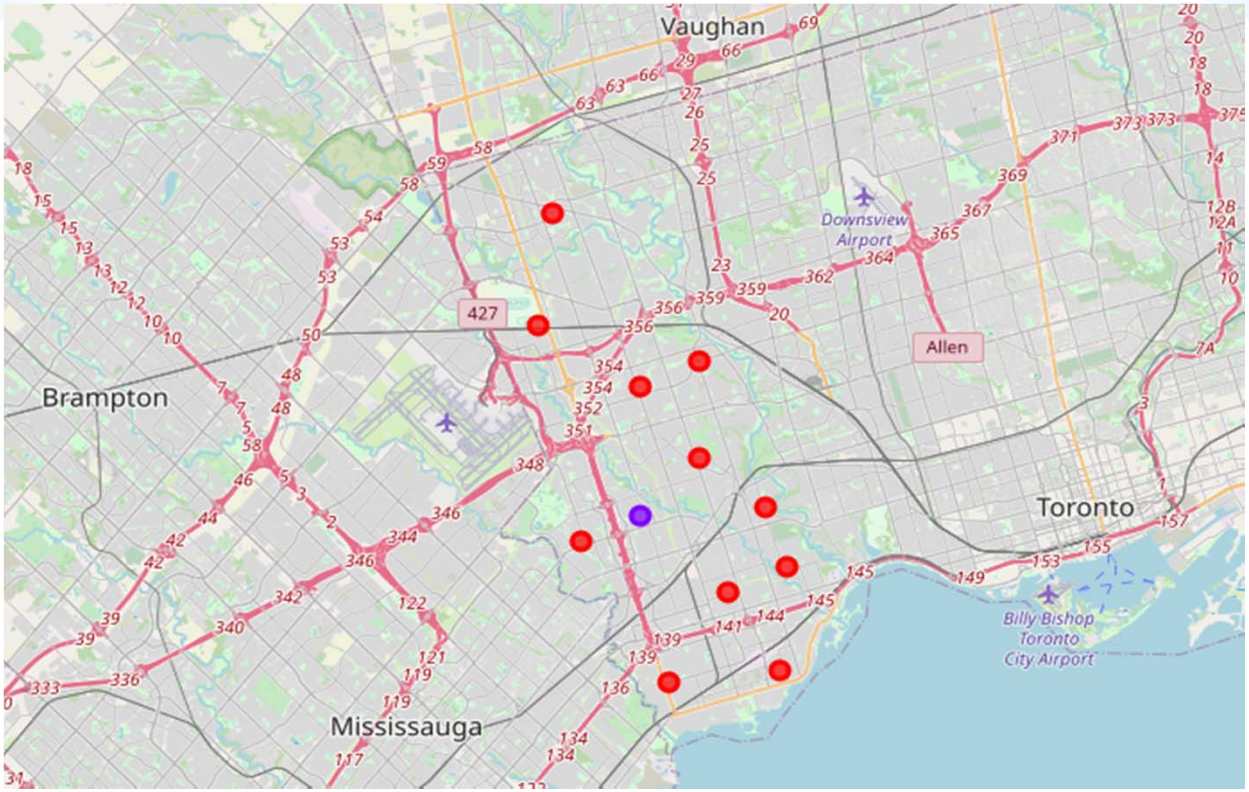
Neighborhood	Population_2016	AfterTaxIncome2015	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
Alderwood, Long Branch	20674.0	63602.0	43.602414	-79.543484	0	Pizza Place	Gym	Sandwich Place	Dance Studio	Pub	Coffee Shop	Pharmacy
Eringate, Bloordale Gardens, Old Burnhamthorpe...	38291.0	67878.0	43.643515	-79.577201	0	Pizza Place	Pet Store	Beer Store	Liquor Store	Convenience Store	Café	Coffee Shop
Islington Avenue, Humber Valley Village	35594.0	65760.0	43.667856	-79.532242	0	Pizza Place	Wings Joint	Flower Shop	Fast Food Restaurant	Drugstore	Discount Store	Dance Studio
Kingsview Village, St. Phillips, Martin Grove ...	33743.0	67497.0	43.688905	-79.554724	0	Pizza Place	Sandwich Place	Bus Line	Park	Chinese Restaurant	Fast Food Restaurant	Drugstore
Mimico NW, The Queensway West, South of Bloor,...	17038.0	78903.0	43.628841	-79.520999	0	Wings Joint	Hardware Store	Bakery	Burger Joint	Burrito Place	Convenience Store	Discount Store
New Toronto, Mimico South, Humber Bay Shores	37975.0	53099.0	43.605647	-79.501321	0	American Restaurant	Café	Gym	Cosmetics Shop	Liquor Store	Mexican Restaurant	Coffee Shop
Northwest, West Humber - Clairville	40684.0	59873.0	43.706748	-79.594054	0	Truck Stop	Bar	Drugstore	Rental Car Location	Garden Center	Convenience Store	Fried Chicken Joint

4. Results

Now is the time to visualize the clusters, using folium library. We can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Based on clusters information we have provided the Clustering of neighborhoods bar plot, showing the neighborhoods with the highest population.

Based on clusters information we have provided the Clustering of neighborhoods bar plot, showing the neighborhoods with the highest income.




Population_2016
0
10000
20000
30000
40000
50000
Alderwood, Long Branch
Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens
ıy West, South of Bloor, Kingsway Park South West, Royal York South West
New Toronto, Mimico South, Humber Bay Shores
Northwest, West Humber - Clairville
IE, The Queensway East, Royal York South East, Kingsway Park South East
Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Gardens
The Kingsway, Montgomery Road, Old Mill North
Westmount
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Cluster 1
Cluster 2
Clustering of neighborhoods


AfterTaxIncome2015
0
20000
40000
60000
80000
100000
Alderwood, Long Branch
Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens
ay West, South of Bloor, Kingsway Park South West, Royal York South West
New Toronto, Mimico South, Humber Bay Shores
Northwest, West Humber - Clairville
JE, The Queensway East, Royal York South East, Kingsway Park South East
Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Gardens
The Kingsway, Montgomery Road, Old Mill North
Westmount
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Cluster 1
Cluster 2
Clustering of neighborhoods

4. Discussions

By examining each neighborhood for the presence of other Pizza Places and counting their numbers, we determine the neighborhoods lacking these places.

Other critical factors that we accounted for in our analysis are the number of people in the neighborhood, neighborhood income and the number of people with Italian descent.

Cluster 2 has only one neighborhood with the optimum population/income relative to other neighborhoods with no Pizza Places: *West Dean Park, Princess Gardens, Martin Groves, Islington, Cloverdale.*

From our our scraped data frame from Wikipedia, we see that Etobicoke Centre has the most Italian ethnic population from Etobicoke.

Neighborhood	Population	Ethnic Origin #1	Percentage #1	Ethnic Origin #2	Percentage #2	Ethnic Origin #3	Percentage #3	Ethnic Origin #4	Percentage #4	Ethnic Origin #5	Percentage #5	Ethnic Origin #6
Etobicoke-Lakeshore	127520.0	English	17.1	Canadian	15.9	Irish	14.4	Scottish	13.5	Polish	9.2	Italian
Etobicoke North	116960.0	East Indian	22.2	Jamaican	6.2	Canadian	5.7	Iraqi	4.8	Italian	3.9	Filipino
Etobicoke Centre	116055.0	Italian	15.1	English	14.3	Canadian	12.1	Irish	10.8	Scottish	10.4	Ukrainian

Another factor to differentiate between the three chosen neighborhoods with no Pizza Place is the Italian ethnic community numbers.

The only neighborhood situated on Etobicoke Centre is the one from Cluster 2: *West Dean Park, Princess Gardens, Martin Groves, Islington, Cloverdale.*

Also *West Dean Park, Princess Gardens, Martin Groves, Islington, Cloverdale* has a high number of population with high income after tax, which makes it a good candidate to open a Pizza Place there.

Based on our analysis considering the number of people in the neighborhood, neighborhood income, the number of people with Italian descent and no Pizza Place existent in the neighborhood, is a good idea to open a Pizza Place on *West Dean Park, Princess Gardens, Martin Groves, Islington, Cloverdale.*

	Neighborhood	Population_2016	AfterTaxIncome2015	Latitude	Longitude
3	The Kingsway, Montgomery Road, Old Mill North	10787.0	97210.0	43.653654	-79.506944
4	West Deane Park, Princess Gardens, Martin Grov...	32400.0	91110.0	43.650943	-79.554724
0	Mimico NW, The Queensway West, South of Bloor,...	17038.0	78903.0	43.628841	-79.520999
2	Old Mill South, King's Mill Park, Sunnylea, Hu...	21299.0	63142.0	43.636258	-79.498509
1	Northwest, West Humber - Clairville	40684.0	59873.0	43.706748	-79.594054

5. Conclusion

During this project we have identified the business problem, looking into which data will be beneficial for our analysis and processing the gathered data from various web sources. By using unsupervised machine learning algorithm using k-means clustering we were able to cluster Etobicoke neighborhoods to provide an optimum location for opening a Pizza Place.

Final decision should take into account other factors as parking, major roads close by, proximity to other various venues, and real estate rent values and also future business area developments.