WHERE TO OPEN A PIZZA PLACE IN ETOBICOKE

TABLE OF CONTENTS

1. INTRODUCTION / BUSINESS PLAN **2**

1.1 INTRODUCTION 2

1.2 BUSINESS PROBLEM 2

2. DATA TO DETERMINE BEST LOCATION FOR PIZZA PLACE IN ETOBICOKE **2**

3. METHODOLOGY **3**

4. RESULTS **10**

5. DISCUSSION **12**

6. CONCLUSION **14**

5. REFERENCES **14**

1. Introduction/Business Problem

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. Census Tracts include several city blocks and have on average about 4,000 people. Neighborhoods are comprised of from 2 to 5 Census Tracts.[3]

Location can make the difference between a successful restaurant and a good idea that never really got off the ground. It is all about finding the right ingredients – the right combination of a desirable location, key demographics, and available parking – when cooking up a recipe for a successful restaurant. [4] And further steps are required to analyze and be taken in consideration, including creating a business plan, having bank approval or finance available, recruiting a team and decide about equipment suppliers, etc.

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.

Another factor determining an optimum solution where to start a new Pizza Place is linked to the neighborhoods people's income. More people earning more money is linked proportionally to people spending more time and money into restaurants.

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.

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

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.

	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

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	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
45	M2L	96512.0	North York	York Mills, Silver Hills	11717.0
23	M4G	94853.0	East York	Leaside	19076.0
12	M1C	93943.0	Scarborough	Rouge Hill, Port Union, Highland Creek	35626.0
11	M9B	91110.0	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	32400.0
7	M3B	90841.0	North York	Don Mills	13324.0

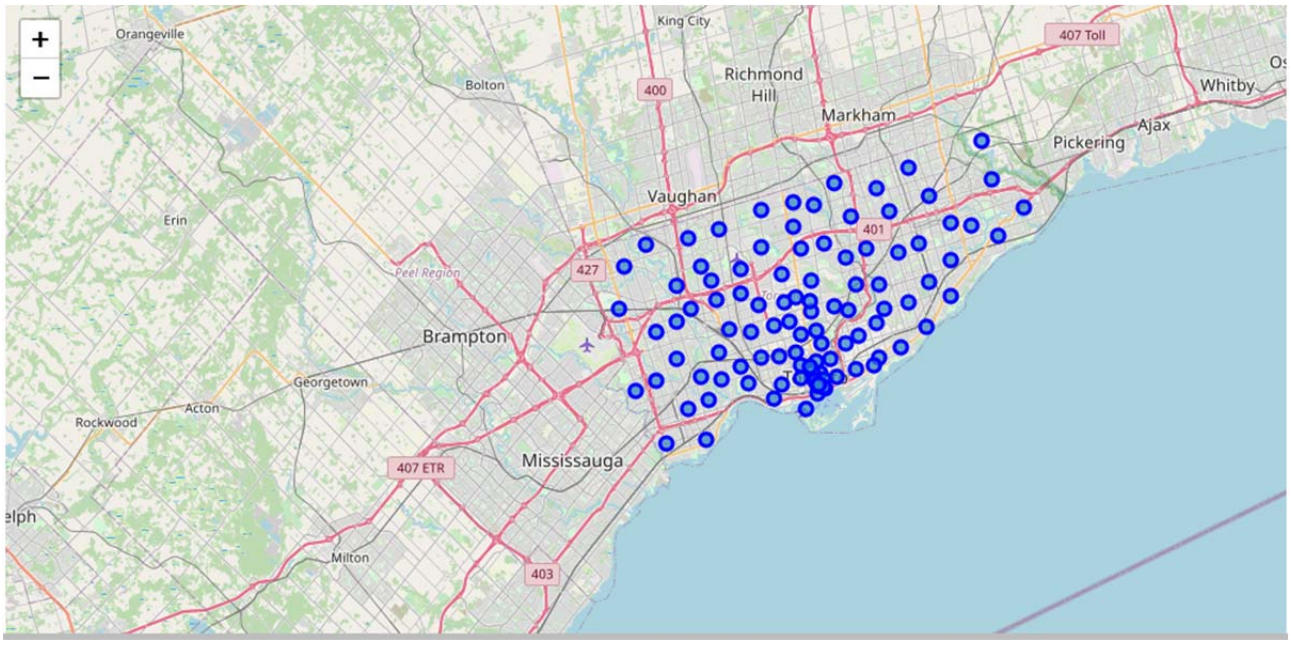
With the help of geopy library to get the latitude and the longitude coordinates of each neighborhood.

	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

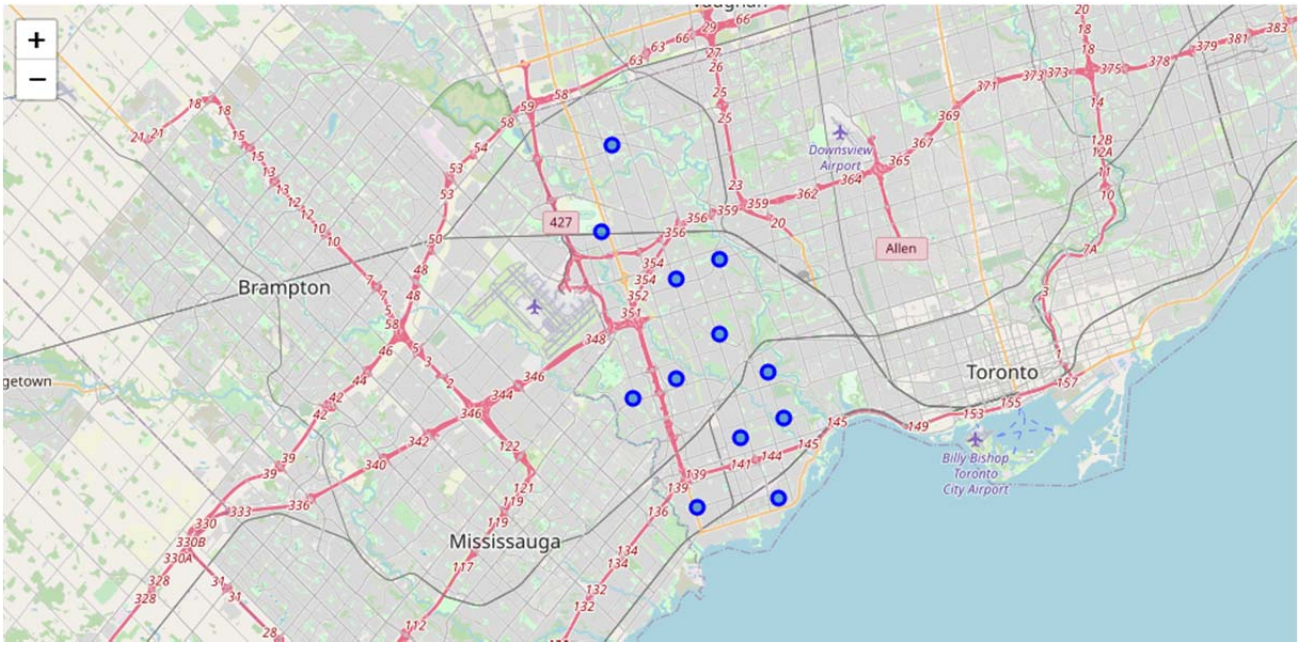
We are ready to scrape now the demographic ethnic table from Wikipedia. Some ethnic information missing was also added into a csv file, to allow us to further interpret the ethnic demographic data.

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.



Now, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. HTTP requests would be made to the Foursquare API server using Postal codes of the Etobicoke city neighborhoods to pull the location information (Latitude and Longitude).

Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Using Foursquare API we found that are 42 unique categories and 77 venues in Etobicoke.

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 venues by categories.

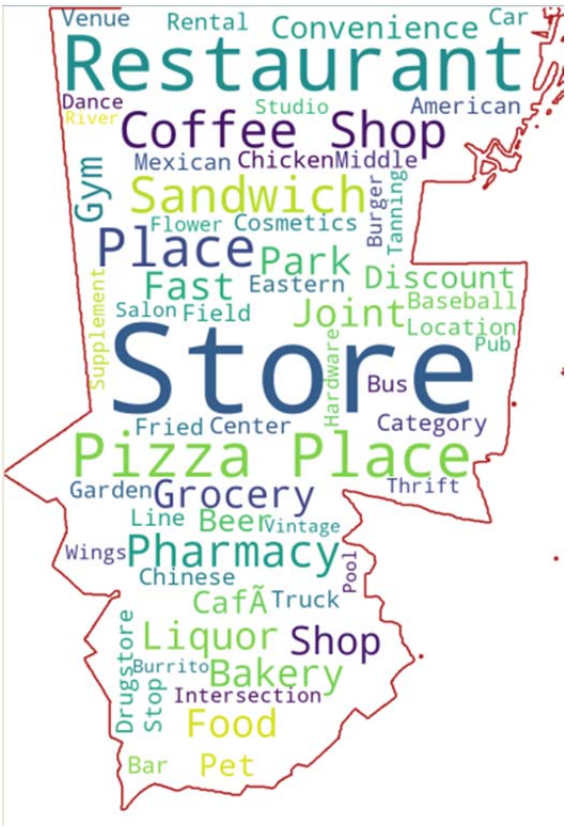
An easy way to visualize the top 10 most common venues is to put together into a data frame and used it to sort the venues in descending order.

Using WordCloud function for our category venues, we were able to use an Etobicoke mask image and transform it in the numpy array form.

Extensive comparative analysis of Etobicoke neighborhoods is used to identify the optimum location using python's scientific libraries Pandas, NumPy and Scikit-learn.

As can be seen in the picture on the right side, the most common venue in Etobicoke contains the word Store, which is part of many different venues categories. The most second found venue is Restaurant and on the third place can be seen the Pizza Place.

The Wordcloud function is not to be used here as a predictor, but rather as an indicator of the venues found in Etobicoke.



The mean of the frequency of occurrence of each category will help to group by categories.

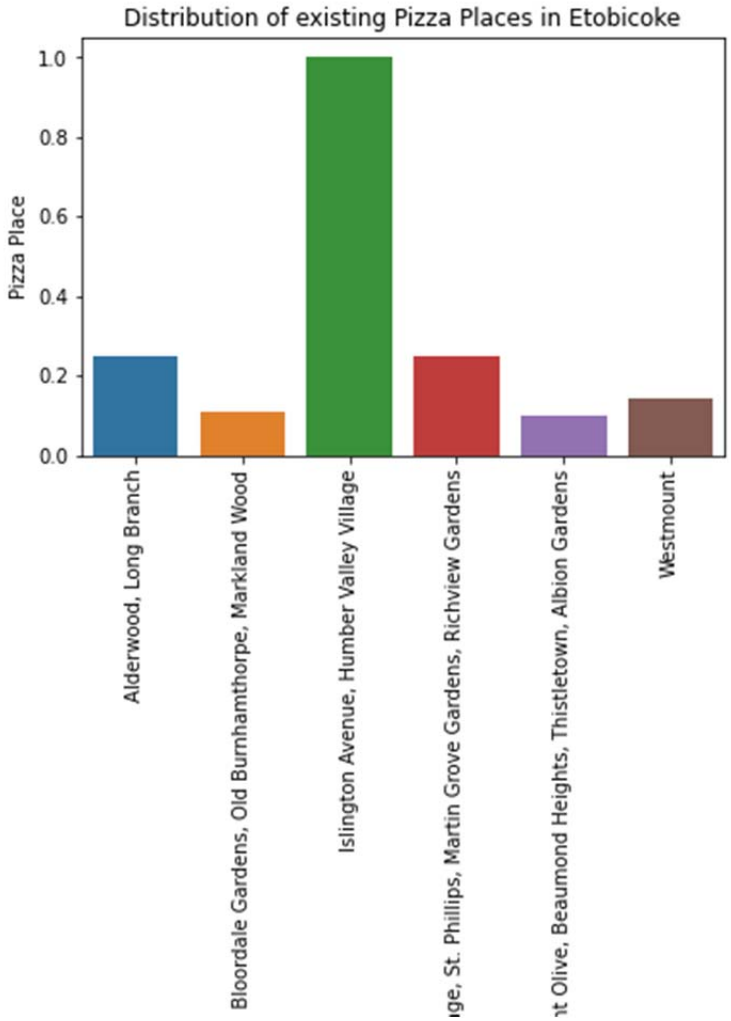
Our decision to advise which Etobicoke neighborhood is optimum to start a Pizza place business must be based primordially on the areas with no Pizza places.

From our newly created panda data frame we need to sort out the neighborhoods missing a Pizza place.

For this step we will create a new data frame containing only these neighborhoods, as seen below:

	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...

It is also important to have a look at neighborhoods that already have Pizza places present. To have a better distribution picture of Etobicoke neighborhoods with Pizza Places that are already operated, we will create a histogram.



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.

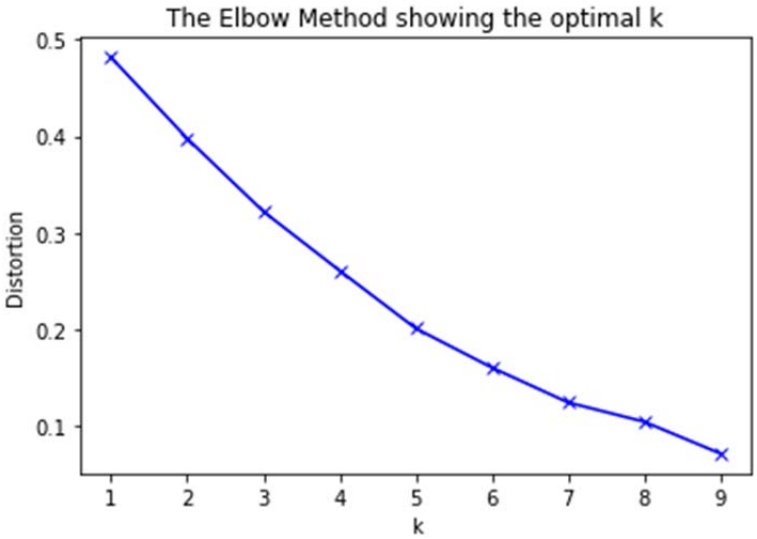
Next we will apply unsupervised machine learning algorithm K-mean to cluster the different categories of Etobicoke neighborhoods. These clusters will be analyzed to allow us drawing conclusions with the help of other characteristics as well, related to ethnic demographic, neighborhood population numbers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

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

First, let's determine the optimal value of K for our dataset using the Silhouette Coefficient Method

A higher Silhouette Coefficient score relates to a model with better-defined clusters. The Silhouette Coefficient is defined for each sample and is composed of two scores: `

a: The mean distance between a sample and all other points in the same class.

b: The mean distance between a sample and all other points in the next nearest cluster.

The Silhouette Coefficient is for a single sample is then given as:

s=b-a/max(a,b)

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.

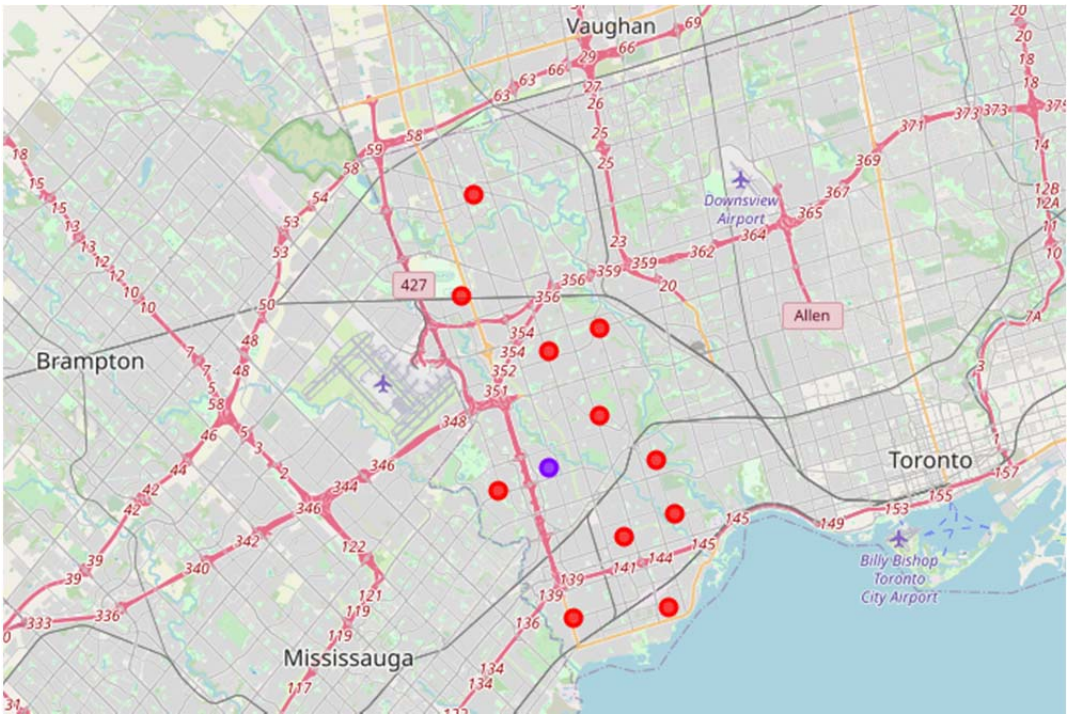
```
For n_clusters=2, The Silhouette Coefficient is 0.3778162740186173
For n_clusters=3, The Silhouette Coefficient is 0.34512282272962275
For n_clusters=4, The Silhouette Coefficient is 0.27162193123502054
For n_clusters=5, The Silhouette Coefficient is 0.2690004104118124
For n_clusters=6, The Silhouette Coefficient is 0.15628831768394813
For n_clusters=7, The Silhouette Coefficient is 0.11840708835144625
For n_clusters=8, The Silhouette Coefficient is 0.05613428672359816
For n_clusters=9, The Silhouette Coefficient is 0.03725646919365794
```

```
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.



Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

```
etobicoke_cluster1=etobicoke_merged.loc[etobicoke_merged['Cluster Labels'] == 0, etobicoke_merged.columns[[0] + list(range(1, etobicoke_merged.sh
```

```
etobicoke_cluster1
```

	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 M Com Ve
0	Alderwood, Long Branch	20674.0	63602.0	43.602414	-79.543484	0	Pizza Place	Gym	Sandwich Place	Dance Studio	Pub	Coffee Shop	Pharr
1	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	C
2	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	D St
3	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	Drug
4	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	Disc
5	New Toronto, Mimico South, Humber Bay	37975.0	53099.0	43.605647	-79.501321	0	American Restaurant	Café	Gym	Cosmetics Shop	Liquor Store	Mexican Restaurant	C

Cluster 2

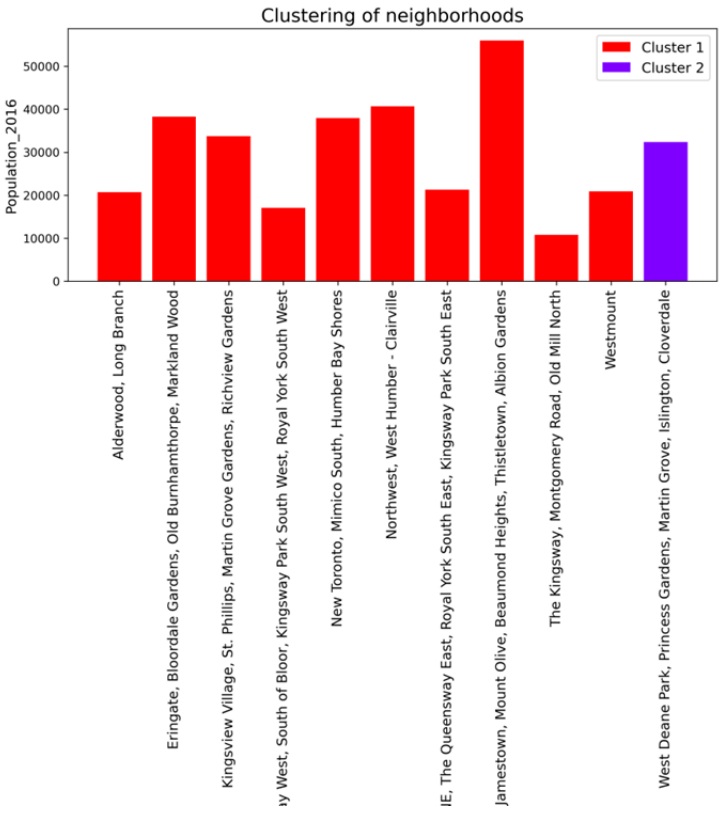
```
etobicoke_cluster2=etobicoke_merged.loc[etobicoke_merged['Cluster Labels'] == 1, etobicoke_merged.columns[[0] + list(range(1, etobicoke_merged.sh
```

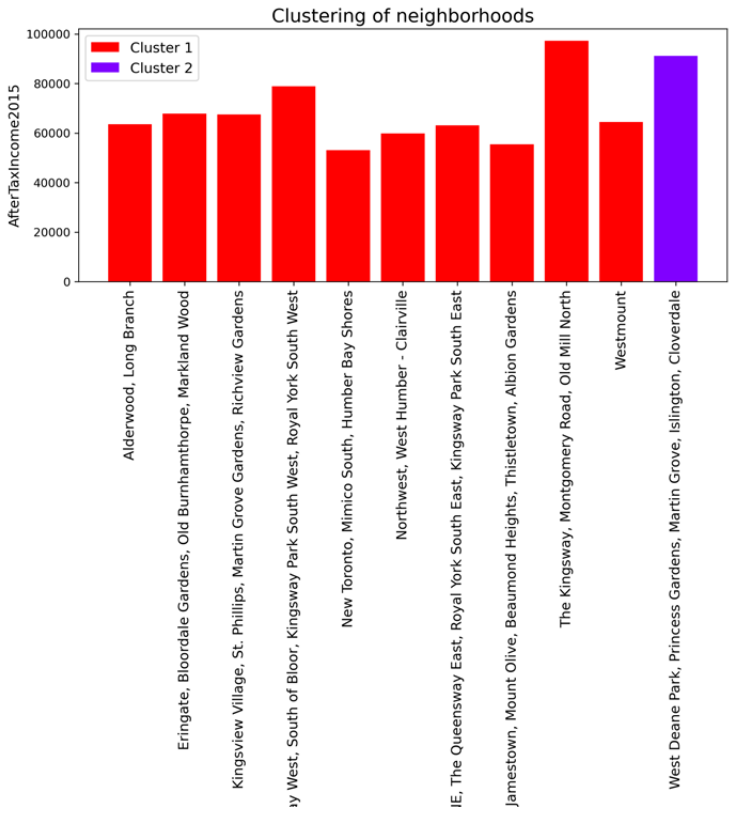
```
etobicoke_cluster2
```

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	8th Most Common Venue
West Deane Park, Princess Gardens, Martin Grov...	32400.0	91110.0	43.650943	-79.554724	1	Bakery	Wings Joint	Coffee Shop	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Drugstore	Discount Store

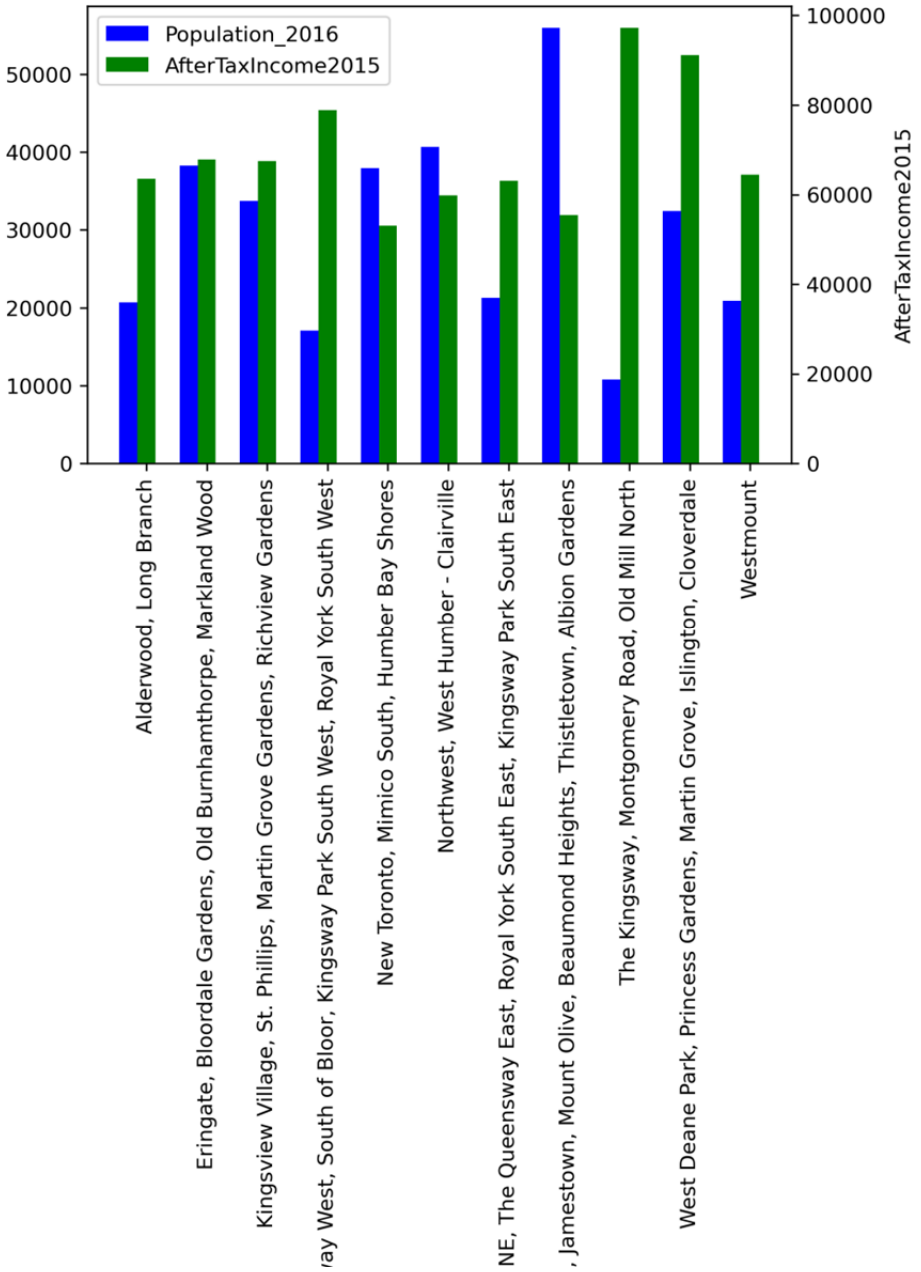
Based on clusters information we have provided the Clustering of neighborhoods bar plot, showing the neighborhoods with the highest population.

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





Very important to help to take a decision is to show the population and income for each neighborhood.



5. Discussions

The first step towards choosing an optimum location for opening a new Pizza Place in Etobicoke borough of Toronto is exploring carefully all its neighborhoods. 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.*

There are other neighborhoods on Cluster 1 that is worth our attention:

- *South Steeles, Silverstone, Humbergate, Jamestone*
- *Northwest, West Humber – Clairville*

There is a correlation of the neighborhoods with the higher population and higher income having Pizza Place as the most venues in the area.

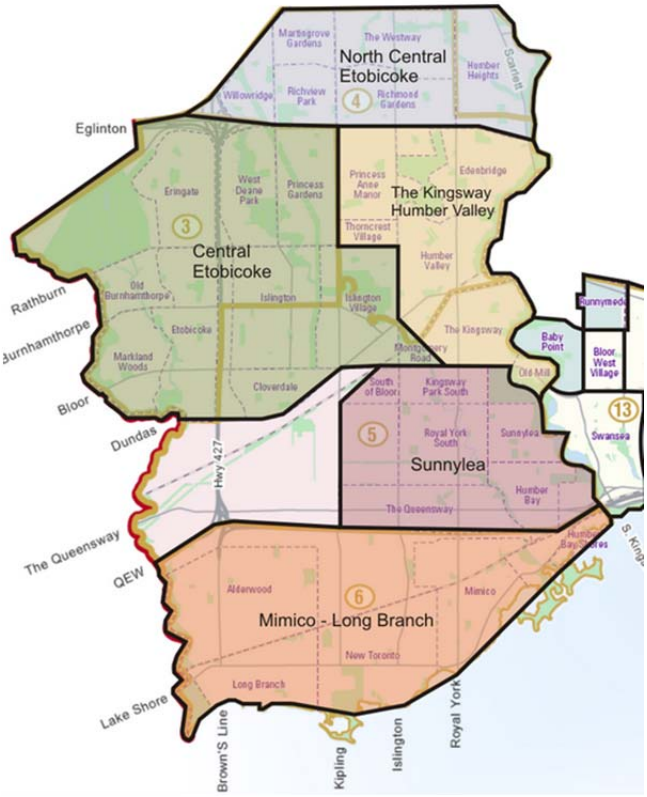
Below is presented our scraped data frame from Wikipedia and city of Toronto website regarding Etobicoke ethnic demographic distribution. We see that Etobicoke Centre has the most Italian ethnic population from Etobicoke.

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.

We will factor in the four parameters to determine the optimum neighborhood to open a Pizza Place: no existing Pizza Place in the neighborhood, population numbers, income after tax and the Italian ethnic group 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.

	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

We already found out that the optimum neighborhood to choose a Piza Place location relative to population/income is ***West Dean Park, Princess Gardens, Martin Groves, Islington, Cloverdale.***

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.*

6. 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.

7. References:

[1]*"Census Profile, 2016 Census". www12.statcan.gc.ca. Statistics Canada.

[2]*"Portrait of the Canadian Population in 2006: Subprovincial population dynamics,

[3]*Toronto Neighbourhood Profiles. Retrieved from https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/

[4]*Opening a Restaurant – Why Location is Important https://idealsoftware.co.za/opening-a-restaurant-why-location-is-important/