

Location prediction of the eastern European cuisine restaurant in the Toronto area



Capstone Project



Introduction

- The family of immigrants from an eastern European country that are the residents of Toronto for past 25 years have decided to pursue their own business and open a restaurant in Toronto area.
- The idea is to open an affordable restaurant for the middle-income customers that would especially cater to the customers of the similar eastern European background.



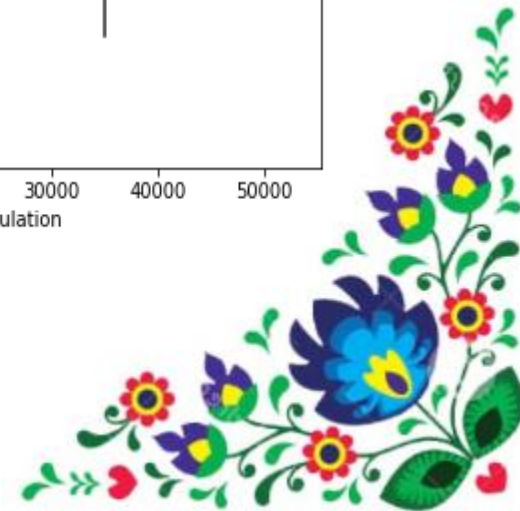
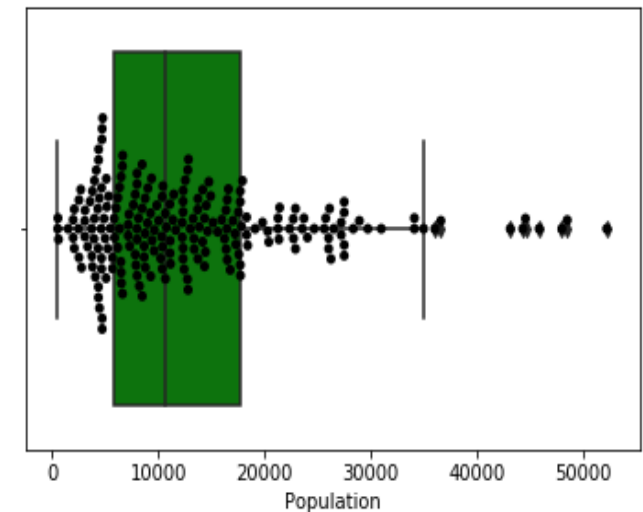
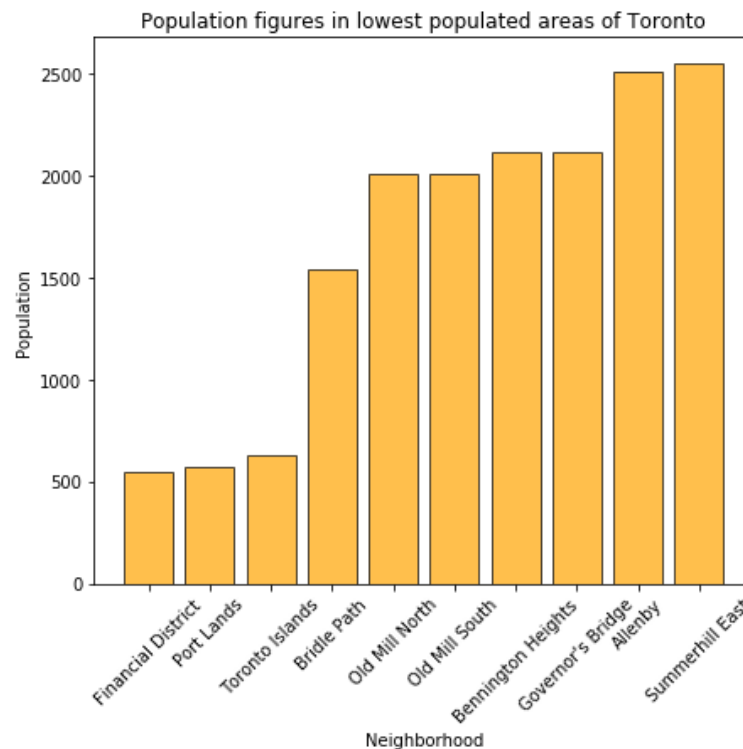
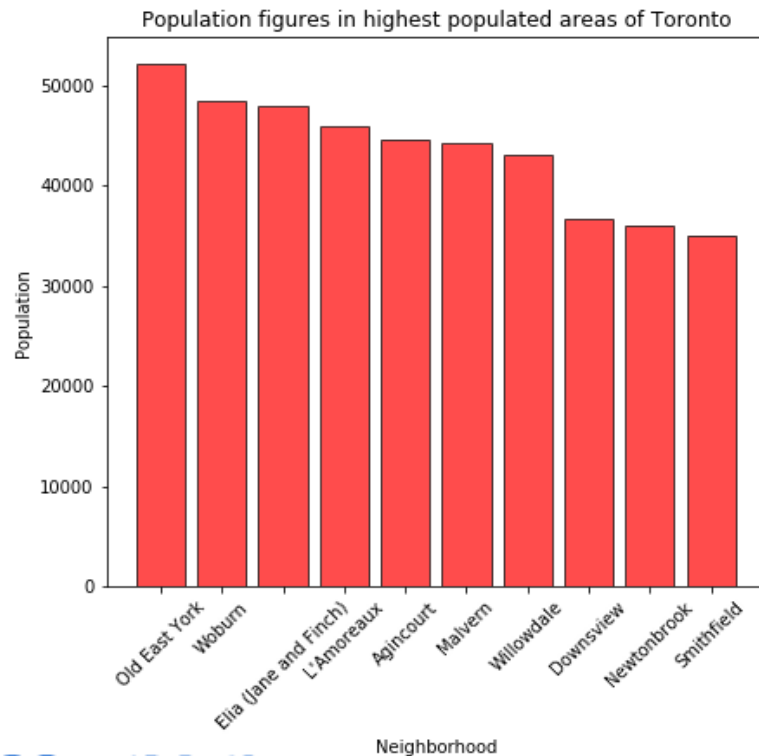
Data acquisition

- **Neighbourhoods, Borough and postal code data:**
 - The information is obtained from canadapost.ca website, the grouped data is taken from [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M]. This data is primarily needed to access the geographical data and Foursquare API. Tables were scraped using the BeautifulSoup toolbox
- **Neighbourhood geo data:**
 - The coordinates data for each neighbourhood was obtained using geocoder-arccgis method on postal code information from the Toronto neighbourhood data
- **Population, average income and demographic data:**
 - The demographics data was obtained from the [https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods]. This will be used to determine the best-suited neighbourhood in terms of number and income of target customer group (i.e. immigrants from Eastern Europe). Tables were scraped using the BeautifulSoup toolbox
- **Venue data:**
 - Foursquare API is used to obtain the venues in Toronto area. This will be used to determine the amount of competition in each neighbourhood



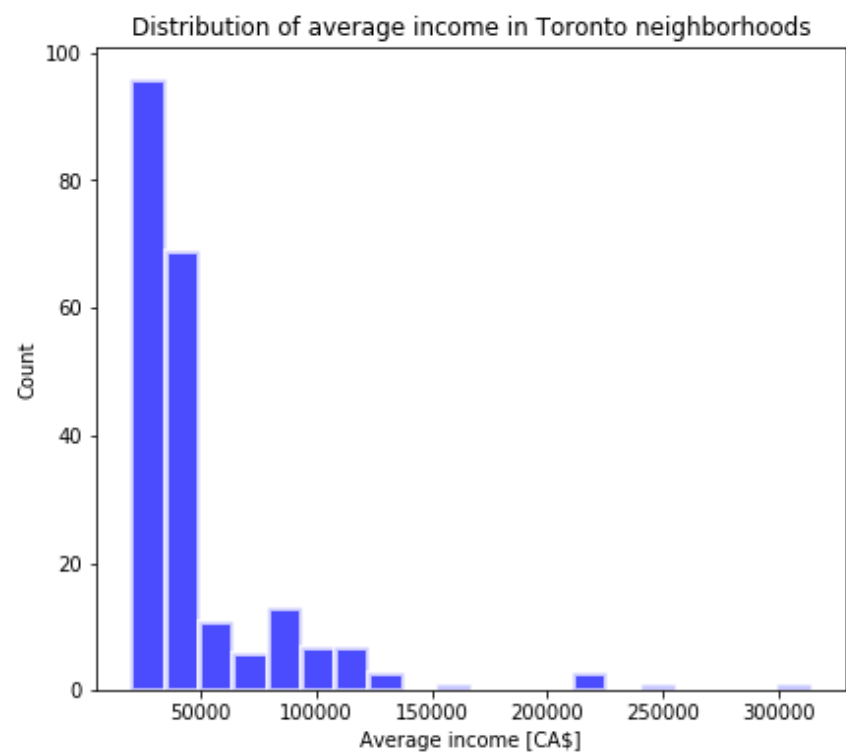
Data analysis

- The problem and the possible solutions need to be evaluated and understood through data, in order to utilize the clustering/ classification algorithms properly
- Population data in the Toronto neighbourhoods:



Data analysis

- Average income data in the Toronto neighbourhoods:
 - The table indicates that the mean average income in Toronto is around 50 000 Canadian dollars and the mean population per neighbourhood is around 14 000 residents.

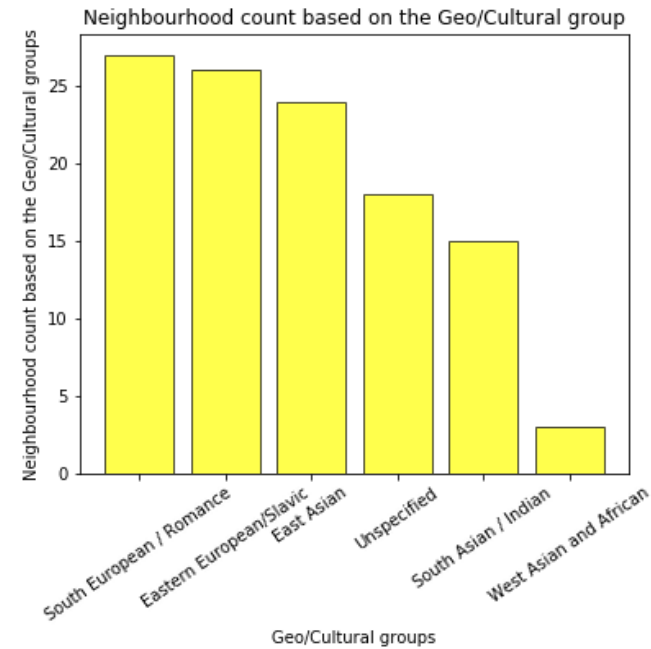
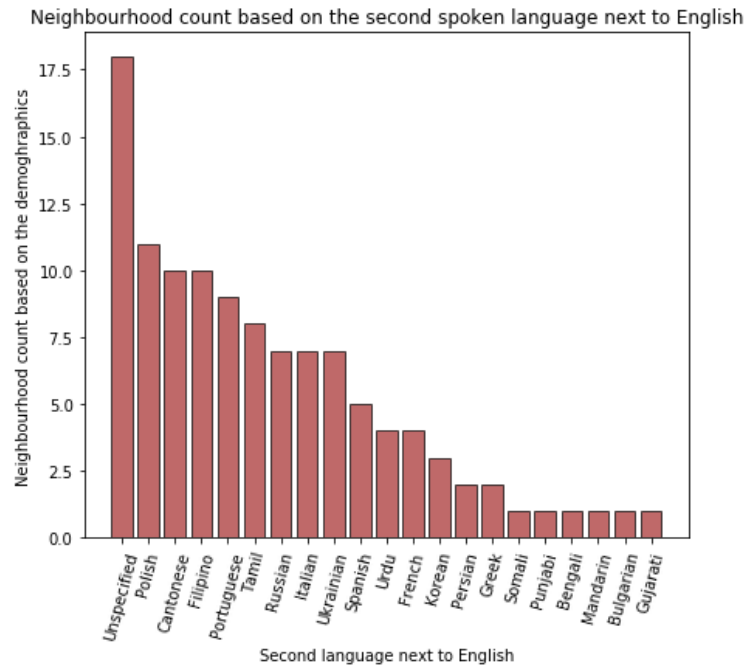


	Population	Average income
count	113.000000	113.000000
mean	14135.557522	48619.380531
std	9794.219106	33966.522451
min	627.000000	21155.000000
25%	7672.000000	28403.000000
50%	12348.000000	36361.000000
75%	17602.000000	48965.000000
max	48507.000000	214110.000000

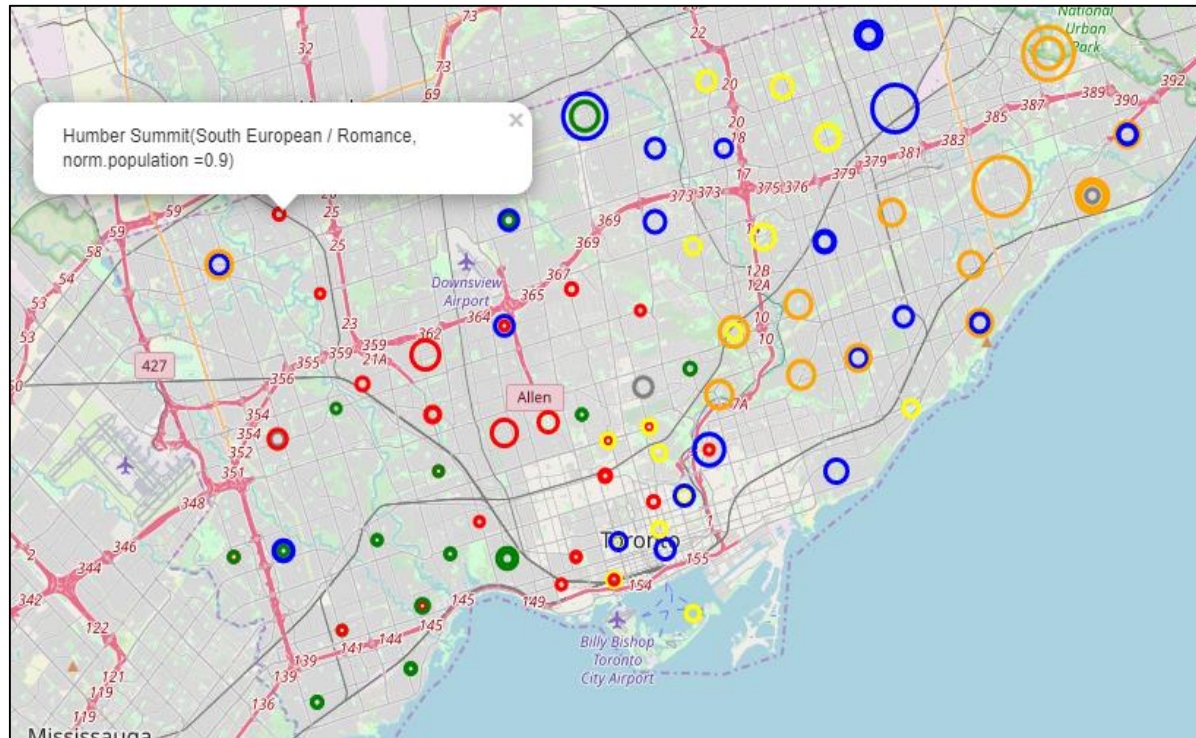


Data analysis

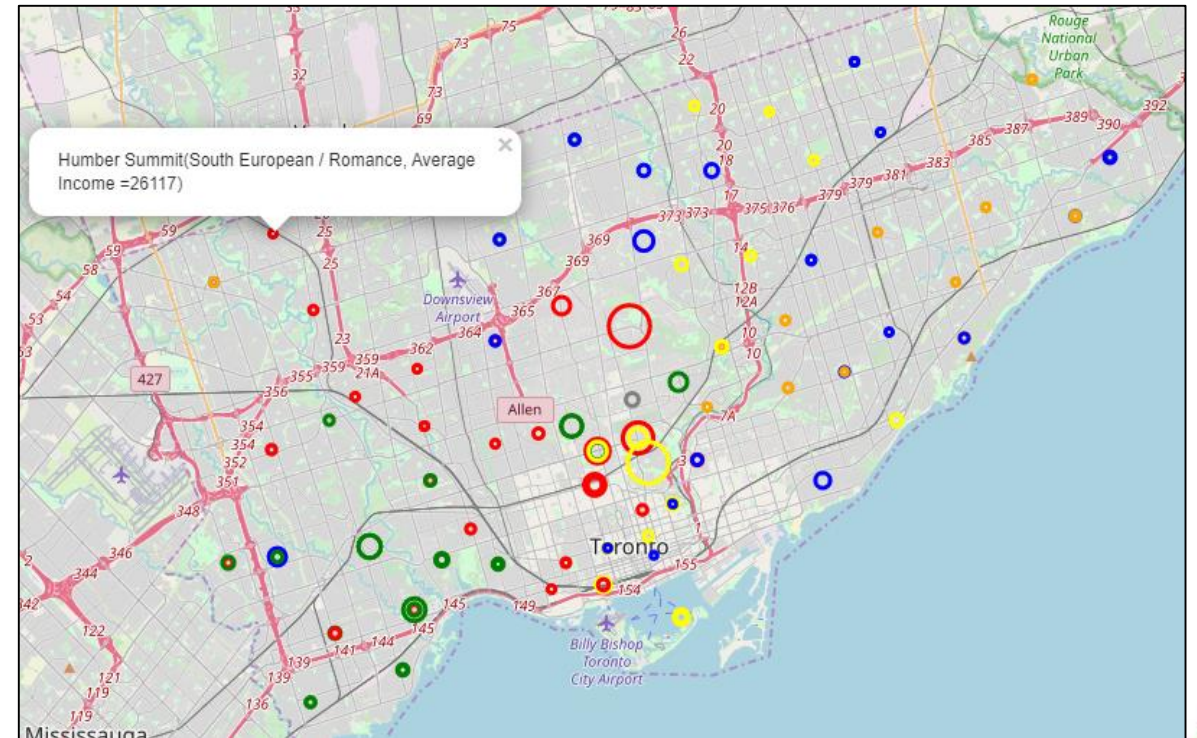
- Cultural / geographical groups of immigrants:
 - According to the obtained data eastern Europeans present the second largest group of Toronto residents



Data analysis through folium maps



1)Marker size is corresponding to the population size



2)Marker size is corresponding to the average income

- Red circle: Represents "South European / Romance" geo/cultural group
- Green circle: Represents "Eastern European / Slavic" geo/cultural group
- Orange circle: Represents "South Asian / Indian" geo/cultural group
- Blue circle: Represents "East Asian" geo/cultural group
- Yellow circle: Represents "Unspecified" geo/cultural group
- Grey circle: Represents "West Asian and African" geo/cultural group



Methodology – step 1

- Construction of the “ideal” neighborhood for classification

$$P_{ideal} = P_{mean} + \frac{(P_{max} - P_{mean})}{2} \quad - \text{Moderately high population}$$

$$I_{ideal} = I_{mean} \quad - \text{Mean income}$$

$$N_{ideal} = 1 \quad - \text{Corresponds to the eastern European neighborhood label}$$

	Geo/Cultural group	Language group	Label
0	South Asian / Indian	[Tamil, Gujarati, Bengali, Urdu, Punjabi]	0
1	East Asian	[Filipino, Cantonese, Mandarin, Korean]	0
2	West Asian and African	[Persian, Somali]	0
3	Unspecified	[Unspecified]	0
4	Eastern European/Slavic	[Russian, Bulgarian, Ukrainian, Polish]	1
5	South European / Romance	[Greek, French, Spanish, Portuguese, Italian]	0



Methodology – step 1

- Selection of the Clustering and Classification algorithm
- Clustering:

Algorithms
KMeans
DBSCAN

- Similar performance Init = k – means + +
- Kmeans chosen $n_{clusters} = 10$

- Classification

Algorithms
KN-neighbours
Logistic regression
Support-vector machine

- Similar performance KNN and LR
- SVM low accuracy with rbf, high accuracy with sigmoid kernel
- KNN highly dependent on the K value, with n = 8+ accuracy is higher
- Logistic regression chosen for convenience



Results – step 1

- Neighborhoods were filtered based on the obtained cluster where the “ideal” neighborhood would belong to. This leaves the neighborhoods with the similar characteristics to the ideal one.

PostalCode	Borough	Neighborhood	Population	Income	Second Language	Geo/Cultural group
M8Y	Etobicoke	Sunnylea	17602	51398	Polish	Eastern European/Slavic
M6R	West Toronto	Roncesvalles	15996	46820	Polish	Eastern European/Slavic
M3H	North York	Bathurst Manor	14945	34169	Russian	Eastern European/Slavic
M8W	Etobicoke	Alderwood	11656	35239	Polish	Eastern European/Slavic
M6S	West Toronto	Swansea	11133	58681	Polish	Eastern European/Slavic
M8V	Etobicoke	Humber Bay Shores	10775	39186	Russian	Eastern European/Slavic
M8V	Etobicoke	New Toronto	10455	33415	Polish	Eastern European/Slavic
M9C	Etobicoke	Markland Wood	10240	51695	Polish	Eastern European/Slavic
M8W	Etobicoke	Long Branch	9625	37288	Polish	Eastern European/Slavic
M9P	Etobicoke	Westmount	5857	35183	Ukrainian	Eastern European/Slavic
M8V	Etobicoke	Mimico South	4732	47011	Polish	Eastern European/Slavic
M8Y	Etobicoke	Mimico NE	4732	47011	Polish	Eastern European/Slavic
M8Z	Etobicoke	Mimico NW	4732	47011	Polish	Eastern European/Slavic
M9B	Etobicoke	West Deane Park	4395	41582	Ukrainian	Eastern European/Slavic
M6N	York	Runnymede	4382	42635	Ukrainian	Eastern European/Slavic
M6S	West Toronto	Runnymede	4382	42635	Ukrainian	Eastern European/Slavic



Methodology – step 2

- **Foursquare API** was used to determine the competition in the filtered neighborhoods, high competition areas were further filtered out

	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
56	Roncesvalles	Inter Steer	43.649796	-79.450310	Eastern European Restaurant
59	Roncesvalles	Café Polonez	43.645113	-79.448517	Eastern European Restaurant
88	Roncesvalles	Chopin Restaurant	43.644165	-79.448162	Eastern European Restaurant
196	Mimico NW	Zam	43.620798	-79.528265	Eastern European Restaurant

- Neighborhoods were filtered and assigned unique labels (unique clusters) for the final classification step

	PostalCode	Borough	Neighborhood	Population	Income	Second Language	Geo/Cultural group	latitude	longitude	Labels
0	M3H	North York	Bathurst Manor	14945	34169	Russian	Eastern European/Slavic	43.757875	-79.448688	0
1	M6N	York	Runnymede	4382	42635	Ukrainian	Eastern European/Slavic	43.676125	-79.481932	1
2	M6S	West Toronto	Runnymede	4382	42635	Ukrainian	Eastern European/Slavic	43.649620	-79.476141	2
3	M6S	West Toronto	Swansea	11133	58681	Polish	Eastern European/Slavic	43.649620	-79.476141	3
4	M8V	Etobicoke	Humber Bay Shores	10775	39186	Russian	Eastern European/Slavic	43.612200	-79.495146	4
5	M8V	Etobicoke	Mimico South	4732	47011	Polish	Eastern European/Slavic	43.612200	-79.495146	5
6	M8V	Etobicoke	New Toronto	10455	33415	Polish	Eastern European/Slavic	43.612200	-79.495146	6
7	M8W	Etobicoke	Alderwood	11656	35239	Polish	Eastern European/Slavic	43.601131	-79.538785	7
8	M8W	Etobicoke	Long Branch	9625	37288	Polish	Eastern European/Slavic	43.601131	-79.538785	8
9	M8Y	Etobicoke	Mimico NE	4732	47011	Polish	Eastern European/Slavic	43.632835	-79.489550	9
10	M8Y	Etobicoke	Sunnylea	17602	51398	Polish	Eastern European/Slavic	43.632835	-79.489550	10
11	M9B	Etobicoke	West Deane Park	4395	41582	Ukrainian	Eastern European/Slavic	43.650347	-79.555040	11
12	M9C	Etobicoke	Markland Wood	10240	51695	Polish	Eastern European/Slavic	43.648573	-79.578250	12
13	M9P	Etobicoke	Westmount	5857	35183	Ukrainian	Eastern European/Slavic	43.696505	-79.530252	13



Results – step 2

- Logistic regression was chosen as a classification tool due to the ability to predict the probabilities. Doing so, the classification of a single item can give information on good class alternatives (high probability clusters that were not chosen)
- Due to inherent similarity between the neighborhoods, the values of probabilities were very similar

	PostalCode	Borough	Neighborhood	Population	Income	Geo/Cultural group	Labels	Probabilities %
10	M8Y	Etobicoke	Sunnylea	17602	51398	Eastern European/Slavic	10	7.311821
3	M6S	West Toronto	Swansea	11133	58681	Eastern European/Slavic	3	7.236099
0	M3H	North York	Bathurst Manor	14945	34169	Eastern European/Slavic	0	7.207163
12	M9C	Etobicoke	Markland Wood	10240	51695	Eastern European/Slavic	12	7.196573
4	M8V	Etobicoke	Humber Bay Shores	10775	39186	Eastern European/Slavic	4	7.159521
7	M8W	Etobicoke	Alderwood	11656	35239	Eastern European/Slavic	7	7.159083
8	M8W	Etobicoke	Long Branch	9625	37288	Eastern European/Slavic	8	7.134445
6	M8V	Etobicoke	New Toronto	10455	33415	Eastern European/Slavic	6	7.133471
5	M8V	Etobicoke	Mimico South	4732	47011	Eastern European/Slavic	5	7.092504
9	M8Y	Etobicoke	Mimico NE	4732	47011	Eastern European/Slavic	9	7.092504
1	M6N	York	Runnymede	4382	42635	Eastern European/Slavic	1	7.071060
2	M6S	West Toronto	Runnymede	4382	42635	Eastern European/Slavic	2	7.071060
11	M9B	Etobicoke	West Deane Park	4395	41582	Eastern European/Slavic	11	7.067436
13	M9P	Etobicoke	Westmount	5857	35183	Eastern European/Slavic	13	7.067259



Results (Top 3 locations) – step 2

	PostalCode	Borough	Neighborhood	Population	Income	Second Language	Geo/Cultural group	latitude	longitude	Labels	Probabilities %
0	M8Y	Etobicoke	Sunnylea	17602	51398	Polish	Eastern European/Slavic	43.632835	-79.489550	10	7.311821
1	M6S	West Toronto	Swansea	11133	58681	Polish	Eastern European/Slavic	43.649620	-79.476141	3	7.236099
2	M3H	North York	Bathurst Manor	14945	34169	Russian	Eastern European/Slavic	43.757875	-79.448688	0	7.207163



Sunnylea neighborhood in the Etobicoke area proved to be the best location



Conclusions and recommendations

- This study included the demographic analysis of the Toronto neighbourhoods, including the population data, income information and cultural makeup in order to best utilize clustering and classification algorithms for the prediction of the best location of eastern European restaurant
- Three best locations were obtained: 1) Sunnylea in the Etobicoke area, 2) Swansea in the West Toronto area and 3) Bathurst Manor in the North York area.
- With the Sunnylea being the best location according to this study
- Some directions for the future work should include:
 - Accounting for the proximity of the neighbourhoods, as this study looked in to neighbourhoods as separate entities
 - More detailed grouping should be performed in order to determine the similarity of cuisines between different nationalities (as Bulgarian cuisine can be more similar to Greek than eastern European, and some other examples).
 - Large portion of Neighbourhoods was not included in the study, due to faulty and unavailable data. This can be overcome with the higher workload and examination of different databases.

