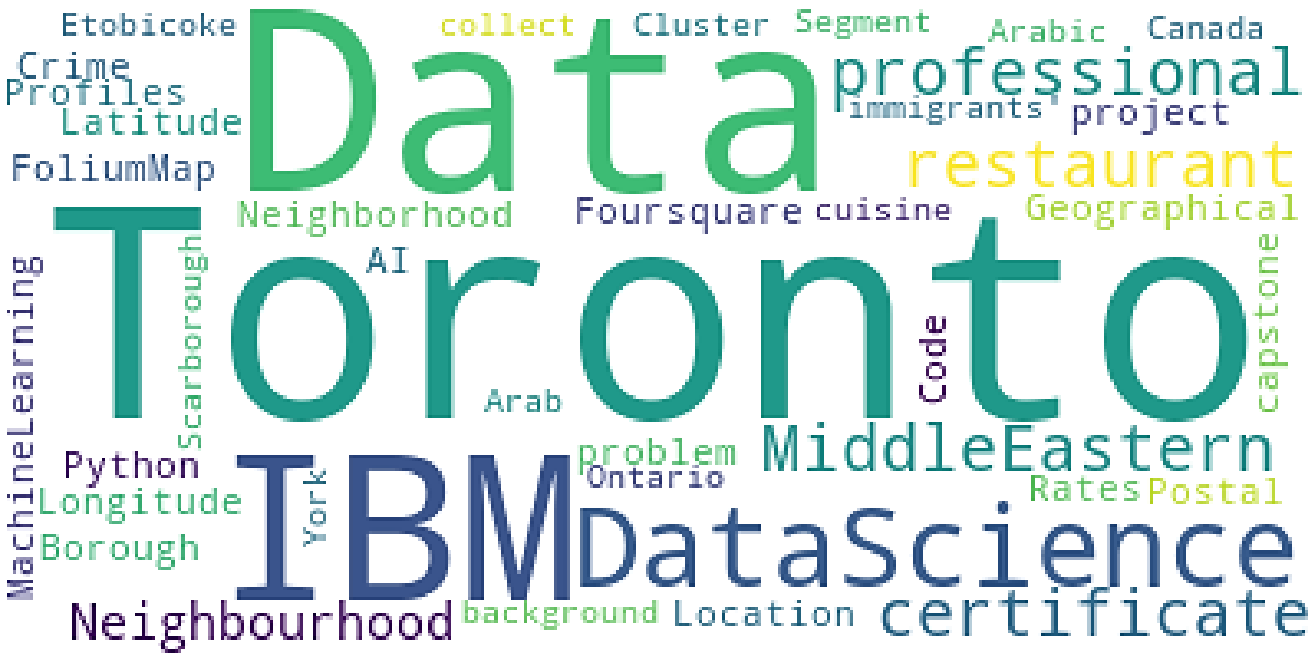
For the last few months, I have been working on earning the IBM Data Science professional certificate. IBM Data Science professional is a course that teaches you how to be a data scientist besides learning the widely used tools in the data science field. To complete the course, you need to create a portfolio of data science projects, the final one is a Capstone Project. As a Capstone for the professional certification, we are asked to describe a problem, discuss its background, and collect data from online resources and location data to create a solution for an exciting problem. This blog explains my process for determining the perfect spot in Toronto, Canada to open a middle eastern restaurant. In this project, we explain all the processing steps starting by problem formulation, data preparation, methodology, and finally the analysis.



1. Introduction

1.1. Background

Toronto, Ontario, was defined as a city in 1834, with a population of 6,197,000 (2020). Toronto is Ontario's capital city, Canada's largest city hall, and ranked as the fourth largest city in North America. It includes the former cities of Toronto, York, Scarborough and Etobicoke, York, and East York. The city accommodates a huge number of the immigrants, also it's a national and international hub for communications, cultural lives, and businesses. Canadians from the Arabian regions are one of the largest ethnic groups in Canada. In 2016, almost 523,235 people from the middle east lived in Canada, representing 1.5% of the total number of the Canadian population. In 2016, Ontario was home to approximately 43% of the overall Arabs in Canada.

1.2. Business Problem

Being one of the most immigrant-friendly places in Canada with a vast Arab Canadian population community, specifically in Toronto, we can clearly conclude that Toronto is one of the best places to start a middle eastern restaurant. As a personal experience, I moved from Egypt to Toronto few months ago. Like most Arab immigrants and long-term visitors, we have a soft spot when it comes to Eastern cuisine. From the moment I settled in, I started exploring nearby places to find middle eastern food and my search resulted in a lot of restaurants from different middle eastern cultures such as Egyptian, Moroccan, Lebanese, etc. However, I faced a big challenge while searching, specially that the majority of these restaurants are far from downtown Toronto or located in a rural neighborhood.

In consequence, I spent a lot of thought considering the great idea of starting a middle eastern restaurant in Toronto, so there will be no need to travel hundreds of kilometers to enjoy the flavorsome Middle Eastern food. However, it will be competitive to open a restaurant in such a place, we need to make sure it's a good idea from different aspects. The main aspects to be considered to determine the most suitable spot to start the dream restaurant is the neighborhood population, place that people tend to gather at especially the Arab community. Besides, we need to locate the most nearby middle eastern restaurants to that spot, and finally, it should provide a high profit to the owner.

1.3. Target Audience

This project target peoples who are interested in:

1. Starting a new personal business by opening a middle eastern restaurant in Toronto. This report will help any investor to make the right decision regarding opening a restaurant that targets the middle eastern community.
2. The middle eastern community in Toronto who wants to find the neighborhoods that have middle eastern food.
3. Any researcher specifically data Scientists, who want to explore Toronto neighborhoods and analysis the data using machine learning techniques.

2. Data

After surveying available datasets for Toronto neighborhoods, I chose to settle down with the most four adequate datasets to start my project. As shown below each dataset contains various unique information about Toronto that will help discover the perfect spot to open a middle eastern restaurant:

1- Foursquare Location Data: To get the location and the venue information. (https://developer.foursquare.com/docs)

2- Neighbourhoods Dataset: To get information about Toronto neighborhoods such as location, and the name of all the neighborhoods present in Toronto. (https://open.toronto.ca/dataset/neighbourhoods/)

3- Neighbourhood Crime Rates (http://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-)

4- Neighbourhood Profiles Data: To get detailed information about each neighborhood in Toronto. (https://open.toronto.ca/dataset/neighbourhood-profiles/)

3. Methodology

This report aims to analyze the neighborhoods of Toronto city from different data sets and find the perfect spot to open a middle eastern restaurant. The following datasets will be utilized in the project:

1) Neighbourhoods Dataset:

This dataset contains neighborhood names as well as the geographic coordinates (latitude and longitude). The geographic coordinates will be used for two purposes; visualizing Toronto map using choropleth map with Folium and call Foursquare API to determine up to 10 food vendors within the center of each neighbourhood.

[7]:

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

2) FourSquare:

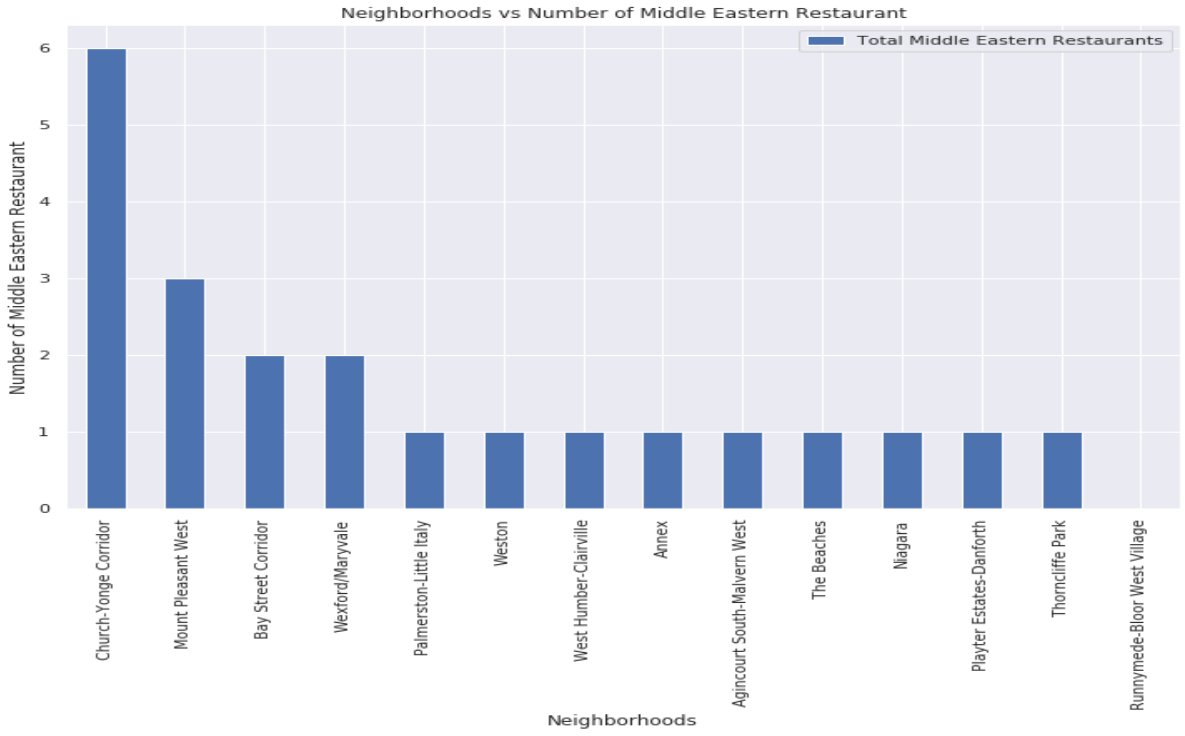
Foursquare API is used to collect data to find the most common venues within a specific radius of a given geographic coordinate. In addition, we will find the total number of restaurants and middle eastern restaurants in each neighborhood.

[50]:

	Neighborhood	Longitude	Latitude	Total Middle Eastern Restaurants	Total Restaurants
72	Church-Yonge Corridor	-79.379017	43.659649	6	28
22	Mount Pleasant West	-79.393360	43.704435	3	27
82	Bay Street Corridor	-79.385721	43.657511	2	18
64	Wexford/Maryvale	-79.298637	43.748572	2	4
35	Playter Estates-Danforth	-79.354887	43.679700	1	18

[22]:

Neighborhood	1st Most Common cuisine	2nd Most Common cuisine	3rd Most Common cuisine	4th Most Common cuisine	5th Most Common cuisine	6th Most Common cuisine	7th Most Common cuisine	8th Most Common cuisine	9th Most Common cuisine	10th Most Common cuisine
Agincourt North	Total Restaurants	Chinese Restaurant	Japanese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Cantonese Restaurant	Caribbean Restaurant	Korean Restaurant	Argentinian Restaurant	Italian Restaurant
Agincourt South-Malvern West	Total Restaurants	Chinese Restaurant	Seafood Restaurant	Asian Restaurant	Cantonese Restaurant	Filipino Restaurant	Mediterranean Restaurant	Restaurant	Malay Restaurant	American Restaurant
Alderwood	Total Restaurants	Latin American Restaurant	Japanese Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Filipino Restaurant
Annex	Total Restaurants	Middle Eastern Restaurant	French Restaurant	Indian Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Tapas Restaurant	Indonesian Restaurant	Turkish Restaurant
Banbury-Don Mills	Total Restaurants	Italian Restaurant	Cantonese Restaurant	Falafel Restaurant	Japanese Restaurant	Indonesian Restaurant	Indian Restaurant	Greek Restaurant	German Restaurant	French Restaurant



3) Neighborhood Profile Toronto:

This dataset contains the data for each of City of Toronto's neighbourhoods, they are collected during the latest Census of Population in 2016. Data collected contains age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labour. These social planning neighbourhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area.

Dataset are cleaned and filtered to only contain the data needed to spot the perfect place to open a restaurant. Particularly, we are interested in the following data points for each neighborhood:

- Population in the neighborhood.
- Population of Persons living alone.
- Population percentage of an immigrant from Middle East countries.
- Population percentage of an immigrant who speaks Arabic as: Mother tongue, Knowledge of language, Language spoken most often at home, Language used most often at work, and First official language spoken.
- Target age youth between 15-24 years and working age between 25-54 years.
- High Average Total income +40,000$. (Neighborhood Profile Toronto)

[48]:

	Neighbourhood	Total Population	Persons living alone (total)	Ethnic origin(Middle Eastern)	Language (Arabic)	Target Age(youth and working age)	Target income (40,000$ +)
0	Agincourt North	29113	1355	475	370	15010	5215
1	Agincourt South-Malvern West	23757	1625	805	325	13325	4720
2	Alderwood	12054	1105	205	50	6455	4065
3	Annex	30526	7880	1420	325	18790	12345
4	Banbury-Don Mills	27695	4360	2305	585	13540	10625

4) Neighbourhood Crime Rates:

This dataset contains Crime Data by Neighbourhood. Data includes four-year averages and crime rates per 100,000 people by neighbourhood based on 2016 Census Population. The crime rates will be added to the final table to find the safest place to open a restaurant. We only choose the rates from 2019 for the following crimes: Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide. Then we calculated the total rate for each neighborhood for all crimes combined.

[3]:

	Neighbourhood	Assault_Rate_2019	AutoTheft_Rate_2019	BreakandEnter_Rate_2019	Robbery_Rate_2019	Homicide_Rate_2019	TheftOver_Rate_2019	Neighbourhood Crimes rate
0	Yonge-St.Clair	295.3	47.9	223.5	31.9	0.0	47.9	646.5
1	York University Heights	1340.9	521.9	391.4	286.3	0.0	101.5	2642.0
2	Lansing-Westgate	445.4	198.0	241.3	68.1	0.0	68.1	1020.9
3	Yorkdale-Glen Park	1411.8	412.1	567.4	283.7	6.8	195.9	2877.7
4	Stonegate-Queensway	327.3	135.7	255.5	87.8	0.0	16.0	822.3

4. Results and Discussion

4.1. Find the total score for each neighborhood

A standardized score will be calculated for each neighborhood by using MinMaxScaler in python with a range of 0-100. To be accurate, all values will be added except the crime rate, we convert it to a negative value so it will be subtracted from the total score of each neighborhood. The scores are merged with the table in the last column as shown in figure. Furthermore, we

will sort values in descending sorting so we can see the highest score, as shown in figure Waterfront Communities- The Island has the best score with 100%.

[4]:

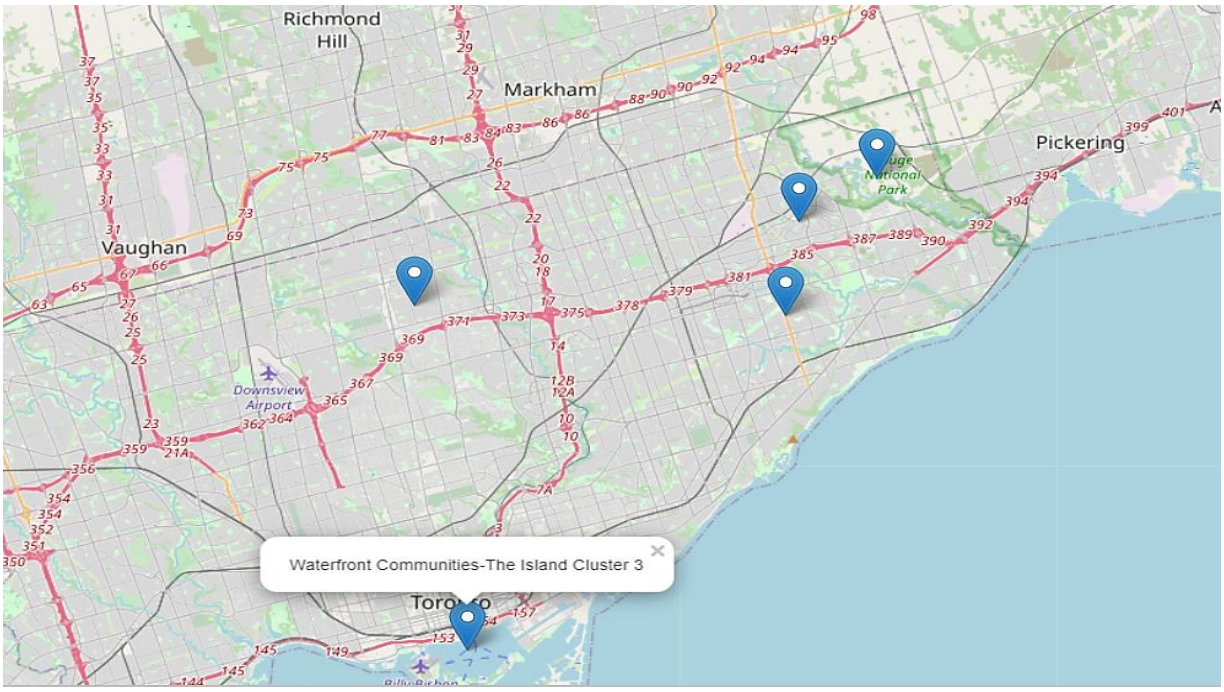
	Neighborhood	Total Population	Persons living alone (total)	Ethnic origin(Middle Eastern)	Language(Arabic)	Target Age(youth and working age)	Target income (40,000$ +)	Neighbourhood Crimes rate	Total score(%)
73	Waterfront Communities-The Island	65913	22035	4210	1535	52945	34230	-2139.2	100.0
79	Willowdale East	50434	7380	8805	1755	32790	15185	-747.5	62.4
81	Woburn	53485	4220	2655	735	29605	9535	-1237.9	52.5
57	Rouge	46496	1640	1280	340	25210	12060	-802.2	44.9
40	Niagara	31180	10075	1440	415	25735	16235	-1581.1	43.3

4.2. Clusters the Neighbourhood Profiles Dataset

With the data now ready to be clustered, we run k-means to cluster the neighborhoods into five clusters. After observation, we found that cluster 3 includes the best neighborhood (Waterfront Communities- The Island) only. Secondly, cluster 3 with four neighborhoods with the following scores (62.4%, 52.4%, 44.9%, and 40.5%). As shown in the below table, we merge our selected clusters with neighborhood profile and the total scores data and that gives us a total of five neighborhoods. The folium map shown below marks the best 5 places to open a middle eastern restaurant from cluster 2 and 3.

[25]:

	Cluster Labels	Neighborhood	Total Population	Persons living alone (total)	Ethnic origin(Middle Eastern)	Language(Arabic)	Target Age(youth and working age)	Target income (40,000$ +)	Neighbourhood Crimes rate	Total score(%)
0	3	Waterfront Communities-The Island	65913	22035	4210	1535	52945	34230	-2139.2	100.0
1	2	Willowdale East	50434	7380	8805	1755	32790	15185	-747.5	62.4
2	2	Woburn	53485	4220	2655	735	29605	9535	-1237.9	52.5
3	2	Rouge	46496	1640	1280	340	25210	12060	-802.2	44.9
5	2	Malvern	43794	2110	1430	600	24485	7595	-1153.1	40.5



4.3. Merge neighborhood profile dataset and top 10 most common venues

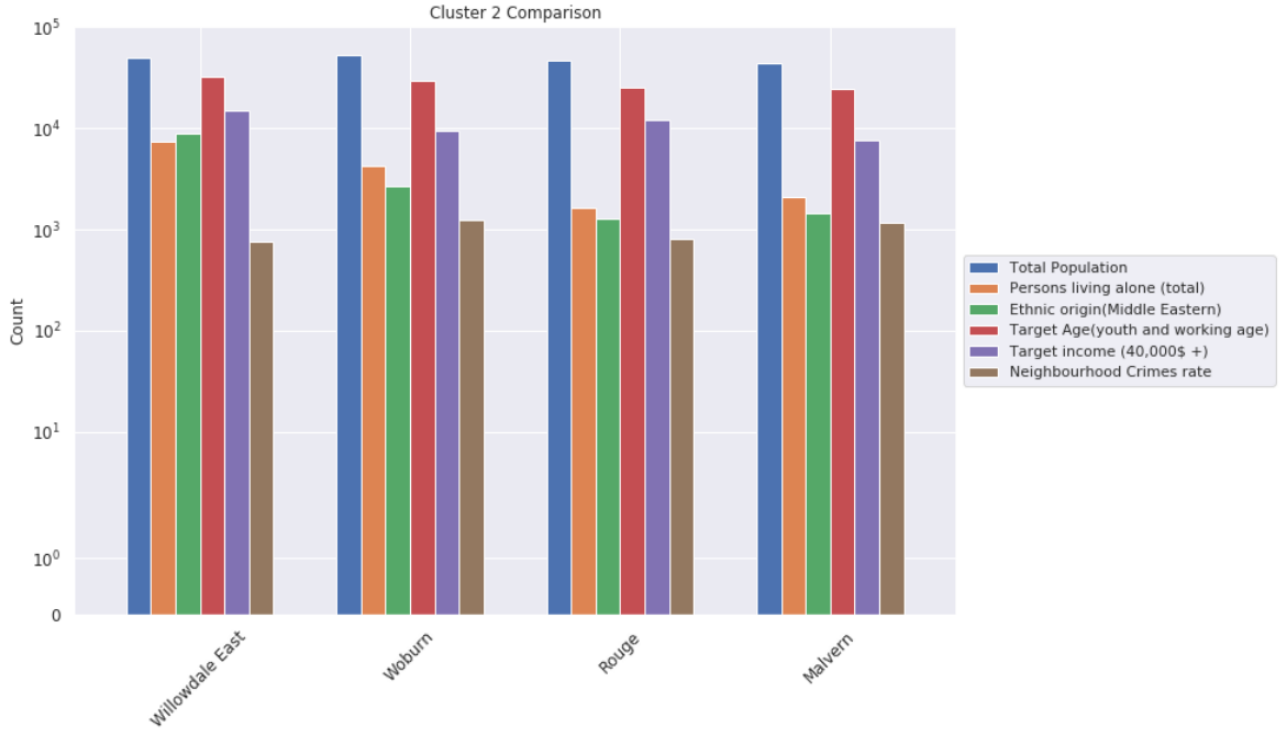
To add our final pieces of criteria in selecting the appropriate neighborhood(s), we merge the above table with the top 10 common venues in each neighborhood to get sure that there won't be any middle eastern restaurant in the same neighborhood we are selecting.

[50]:

	Neighborhood	Total score(%)	1st Most Common cuisine	2nd Most Common cuisine	3rd Most Common cuisine	4th Most Common cuisine	5th Most Common cuisine	6th Most Common cuisine	7th Most Common cuisine	8th Most Common cuisine	9th Most Common cuisine	10th Most Common cuisine
1	Willowdale East	62.4	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Filipino Restaurant
2	Woburn	52.5	Indian Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Italian Restaurant	Indonesian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Filipino Restaurant
3	Rouge	44.9	Restaurant	Fast Food Restaurant	Korean Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Filipino Restaurant
5	Malvern	40.5	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Greek Restaurant	German Restaurant	French Restaurant

4.3. Examine Cluster 2 further

In this section, we focus on cluster 2 to compare the neighborhoods from different aspects such as population, persons living alone, Ethnic origin (Middle Eastern), Target Age(youth and working age), Target income (40,000$ +) and Neighbourhood Crimes rate'. As shown in the below figure, we can notice that Willowdale East comes in second place for the best place to open a middle eastern restaurant. However, choosing the best neighborhood from this group will depend on the investor preference criteria from the neighborhood profile criteria provided in this research.



5. Discussion

The aim of this study was to explore different neighborhoods in Toronto based on some criteria collected from different datasets. The following points summarize the datasets and features collected from each one:

1. The first criteria are from neighborhood profile dataset which includes Total Population, Persons living alone, Ethnic origin (Middle Eastern), Language (Arabic), Target Age (youth and working age), Target income (40,000$ +).

2. The second criteria are the crime rate in each neighborhood which are calculated from Neighbourhood Crime Rates dataset and includes different types of crime.
3. Finally, FourSquare API is used to find the top 10 common venues from each neighborhood to assure that the selected neighborhood won't have any other middle eastern restaurant.

6. Conclusion

Opening a restaurant is a risky investment that can lead to a large loss if not well studied and analyzed. Thus, in this project, I demonstrated the workflow necessary for an investor to determine what is the perfect spot to open a restaurant. To be more specific, I determined the optimal location to open a middle eastern restaurant in Toronto according to different aspect gathered from different data sources. From the results discovered in this research, the following observations and recommendations can be made:

- Based on this research results and the criteria given by the datasets and the clustering, the best neighborhood recommended would be Waterfront Communities — The Island neighborhood. With a total score of 100/100 and grouped in cluster 3 alone for being better than all other neighborhoods with a big difference. We can conclude that this location would be a hot spot to open a middle eastern restaurant.

- A secondary recommendation is made for the neighborhood of Willowdale East. This neighborhood is ranked the second in the top five total score assessment, with a score of 62.4/100. We can consider it as a second option to open our restaurant.

- Lastly, all selected neighborhood doesn't have any middle eastern restaurants in the top 10 common venues, so opening a middle eastern restaurant in those neighborhoods would be a great investment.

Hopefully, this kind of research and analysis will provide the initial guidance to decide where to open a middle eastern restaurant in Toronto, also it would be a great start for anyone who wants to have a better understanding of real- life examples using data-science.