**Opening a Middle Eastern Restaurant in Toronto**

**IBM DATA SCIENCE PROJECT**

**December 4, 2019**



1. **INTRODUCTION**

**1.1 BUSINESS PROBLEM**

The city of Toronto is packed with restaurants, nightlife, and different ethnicities and types of cuisines. An entrepreneur wants to open a Middle Eastern Restaurant in Toronto but doesn’t know which neighborhood to open it in. The things that need to be taken into consideration are population of every neighborhood, the best neighborhoods where the habitants would likely eat in this type of cuisine, and where the competition is limited (such as the number of Middle Eastern restaurants in each neighborhood). As Middle Eastern food is popular among the Arab community, so this entrepreneur might think of open the restaurant in an area where the Arab community resides.

The objective of this project is about finding the best neighborhood in the city of Toronto to open a Middle Eastern Restaurant. By using data science methods and tools along with machine learning algorithms such as clustering, this project will aim to provide a solution to answer the business question: Where should the entrepreneur consider opening a Middle Eastern restaurant in Toronto?

**1.2 TARGET AUDIENCE**

Entrepreneurs or chefs who want to open a Middle Eastern restaurant in Toronto would find this project very interesting and informative in regards to competitive advantage and business values.

1. **DATA**

To solve this problem, we will need the following data:

1. List of Neighborhoods in Toronto, Canada using the FoursquareAPI.
2. Latitudes and Longitudes of these neighborhoods.
3. Venue data related to Middle Eastern Restaurants. This will help us find neighborhoods that are more suitable to open the restaurant in.
4. Look into the demographic data of a specific area. For example: areas with the majority of Arabs people would be good for opening a Middle Easter Restaurant.

***Extracting Data***

* The scrapping of Toronto Neighborhoods via Wikipedia:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

* Getting Latitude and Longitude Data of these neighborhoods via Geocoder package.
* Using Foursquare API to get venue data related to these neighborhoods.

**Demographic data from the City of Toronto's open data**

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv>

We will use the features Ethnic Origin and Neighborhood Information (from the above data file) for each neighborhood, in order to cluster the neighborhoods of Toronto.

Table 1: Ethnic Origin Data from [Middle-Eastern-Population.xlsx](https://github.com/zain1995-design/Coursera_Capstone/blob/master/Middle-Eastern-Population.xlsx)



We can see that:

* We have the name of each neighbourhood in each column name (starting at position 6)
* We have the name of each ethnic origin in the Characteristic column
* The number of people living in each neighbourhood, associated to each ethnic origin name.

**List of venues by neighbourhood using the FoursquareAPI**

In order to obtain the list of venues, and especially the list of Middle Eastern restaurants same as the one we want to open, we are going to request FoursquareAPI with an Explore query.  
The documentation for the Explore query can be found here:

* <https://developer.foursquare.com/docs/api/venues/explore>

We query FoursquareAPI suplying the neighbourhood's information (coordinates calculated with the **Geocoder** package), the radius of scan(500m radius), and the limit of number of venues we want to retrieve(limit of 100 venues).

1. **Methodology**

First I need to get the list of neighbourhoods in Toronto. This is done by extracting the list of neighbourhoods from Wikipedia page: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

I used web scraping by utilizing pandas html table scraping method as it’s much easier to pull tabular data directly from a web page into a data frame.

I then got the coordinates of the neighbourhoods and postal codes to utilize Foursquare to pull the list of venues near these neighbourhoods.

After gathering all the coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

I then used the Foursquare API to pull the list of top 100 venues within a 500 meters radius. I used my foursquare developer account in order to obtain account ID and API key to pull the data. We get the names, categories, latitude, and longitude of the venues. I can even check how many unique categories that I can get from the venues. I will then analyze each neighbourhood by grouping the rows by neighbourhood and taking the mean on the frequency of occurrence of each venue category. This will prepare me for clustering in the next section.

I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. I used this method as it is one of the simplest and most popular unsupervised machine learning algorithms. I divided the neighbourhoods of Toronto into 3 clusters based on their frequency of occurrence for “Middle Eastern Food”. Based on the results of the cluster, I will be able to recommend the idea location to open the restaurant in Toronto.

The table below represents the top 10 neighbourhoods with the highest population of Middle Eastern Origin.

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv>

Table 2: Top 10 Middle Eastern populations in Toronto neighbourhoods and low 3.

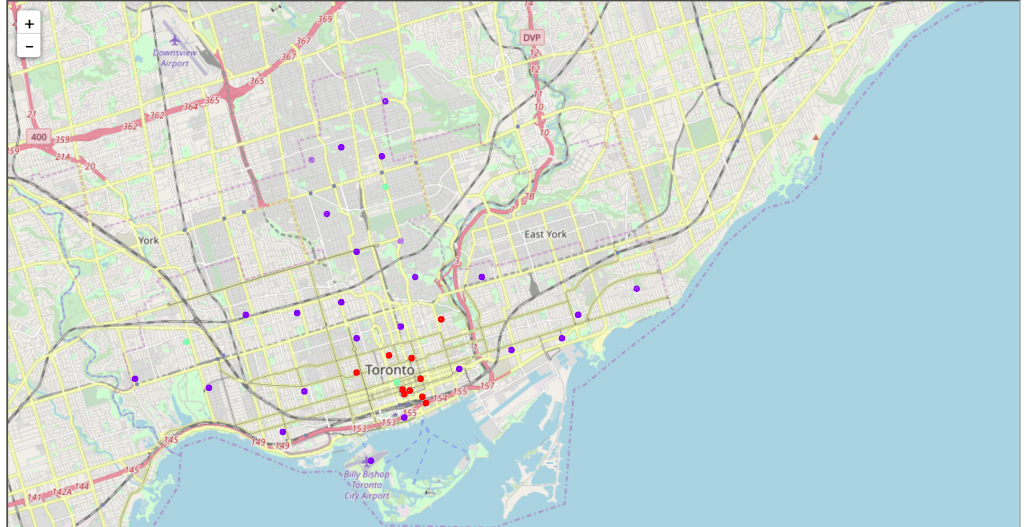
|  |  |
| --- | --- |
| **Toronto Neighbourhood Name** | **Middle Eastern Origin Population** |
| Willowdale East | 10925 |
| Mount Olive-Silverstone-Jamestown | 8465 |
| Parkwoods-Donalda | 6620 |
| Waterfront Communities-The Island | 6465 |
| Don Valley Village | 5715 |
| L'Amoreaux | 5710 |
| Bay Street Corridor | 4310 |
| Wexford/Maryvale | 4115 |
| Bayview Village | 4065 |
| Thorncliffe Park | 3915 |
| Humber Heights-Westmount | 305 |
| Guildwood | 260 |
| Keelesdale | 110 |

1. **Results**

We divided Toronto Neighbourhoods into 3 clusters based on how many Middle Eastern restaurants as in each neighbourhood:

* Cluster 1: Neighbourhoods with little Middle Eastern restaurants.
* Cluster 2: Neighbourhoods with no Middle Eastern restaurants.
* Cluster 3: Neighbourhoods with a high number of Middle Eastern restaurants.

The results are visualized in the below map with cluster 1 in red color, cluster 1 in purple color, and cluster 3 in light green color.

.

1. **Discussions**

Most of the Middle Eastern restaurants are in cluster 3 which is around Willowdale, Parkwoods-Donaldo, Don Valley Village, L’amouraux, and bayview Village neighbourhoods. In cluster 1, there are little Middle Eastern restaurants which include Thorncliffe Park and Wexford neighbourhoods. In Cluster 2, there are no Middle Eastern restaurants which include Guildwood, Humber Heights-Westmount, and Keelesdale neighbourhoods in Toronto.

Therefore, I recommend that the entrepreneur opens the restaurant in Wexford or Thorncliffe Park with little nearby Middle Eastern restaurants(less competition) and a high number of Middle Eastern populations in them according to Table 2 in the report.

**Limitations:**

In this report, I only took 2 factors into consideration: the number of Middle Eastern restaurants in each neighbourhood in Toronto and the population number with Middle Eastern Origins. There are other factors that can be taken into consideration such as the residents’ mean income in each of the neighbourhoods. This factor can also influence the final result of this report.

1. **Conclusion**

In this report, I have identified the business problem, the data required to solve the problem, extracting and preparing the data for analysis, performing machine learning technique by utilizing k-means clustering, and providing the results and recommendations to the entrepreneur.