

Finding the Optimal Location for establishing a Mexican Restaurant in Toronto City, Canada.

AHMED HASSAN

29.06.2020

1. Introduction

1.1 Background

There are several factors to be considered to open a restaurant to achieve the maximum profits such as: Customers, Staffing, Competition, Menu etc. One of the most important factors is finding the best **Location** to open a Mexican Restaurant. So, in this project we are going to use Data Science to find the Optimal Location to establish a Mexican Restaurant.

1.2 Problem

The aim of this project is to help stakeholders, who are interested in opening a **Mexican Restaurant in Toronto, Canada**, to find the optimal location to establish the Restaurant. We need to consider places **aren't** crowded with restaurants. Specially we are looking for a place with **no** Mexican Restaurants in vicinity.

1.3 Interest

Obviously, Stakeholders who are looking for either establishing a new restaurant with a special cuisine in a certain place or have plans to expand their franchised restaurants would be very interested in competitive advantage and business values.

2. Data acquisition and cleaning

2.1 Data Sources

In this project we will get data about the neighborhoods in Toronto City from Wikipedia website, [here](#). We will find in this link a table consisting of three columns (Postal Codes, Boroughs, Neighborhoods). We need also Geographical coordinates for each neighborhood to explore the City using Foursquare API, [here](#). In addition to the population information in the city to calculate the percentage of Latin American people in each neighborhoods [here](#).

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values from earlier seasons, because of lack of record keeping. As we can see in the table we've scraped from Wikipedia webpage, there are a lot of values in 'Borough' and 'Neighborhoods' columns assigned as "Not assigned". So, after we read the table and convert it into a data frame, we remove these values from our data frame. We have also to do some adjustments and add the coordinates columns to make our data frame clean, simple and easy to use, without any confusion.

	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American
0	North York	Victoria Village	M4A	43.725882	-79.315572	17510.0	430.0
1	York	Humewood-Cedarvale	M6C	43.693781	-79.428191	14365.0	530.0
2	East Toronto	The Beaches	M4E	43.676357	-79.293031	21567.0	215.0
3	Scarborough	Woburn	M1G	43.770992	-79.216917	53485.0	745.0
4	North York	Hillcrest Village	M2H	43.803762	-79.363452	16934.0	220.0
5	East York	Thorncliffe Park	M4H	43.705369	-79.349372	21108.0	305.0
6	Scarborough	Scarborough Village	M1J	43.744734	-79.239476	16724.0	280.0
7	North York	Bayview Village	M2K	43.786947	-79.385975	21396.0	410.0
8	North York	Humber Summit	M9L	43.756303	-79.565963	12416.0	895.0
9	York	Weston	M9N	43.706876	-79.518188	17992.0	1390.0

3. Exploratory Data Analysis

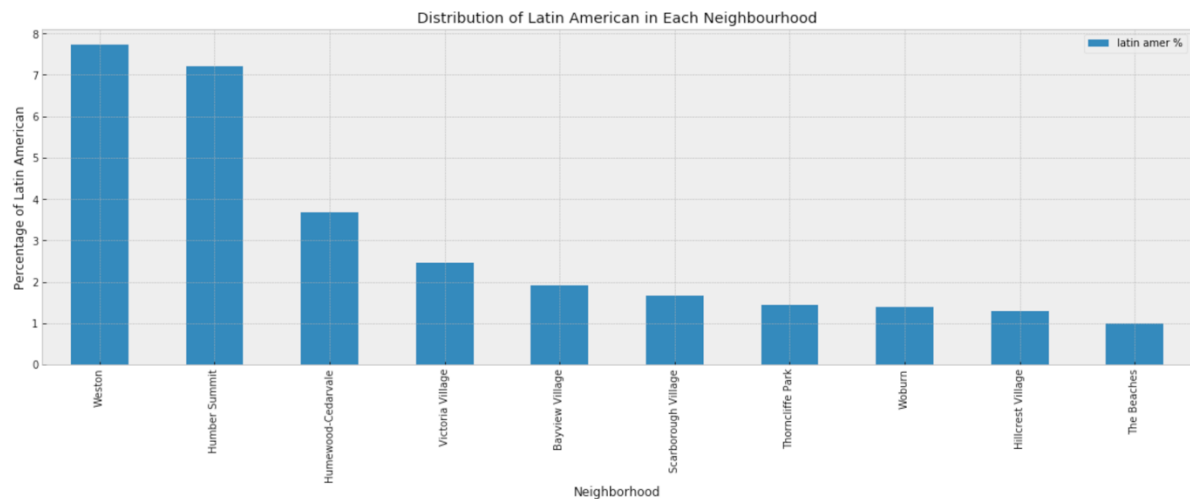
3.1 Calculation of Latin American Distribution

To calculate the percentage of Latin American people in each neighborhood, we need to divide number of Latin American people by the Total Population in each neighborhood and multiply by 100%. Then store the results of each neighborhood in a new column and append this column to our Data frame as shown below:

	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %
0	North York	Victoria Village	M4A	43.725882	-79.315572	17510.0	430.0	2.455740
1	York	Humewood-Cedarvale	M6C	43.693781	-79.428191	14365.0	530.0	3.689523
2	East Toronto	The Beaches	M4E	43.676357	-79.293031	21567.0	215.0	0.996893
3	Scarborough	Woburn	M1G	43.770992	-79.216917	53485.0	745.0	1.392914
4	North York	Hillcrest Village	M2H	43.803762	-79.363452	16934.0	220.0	1.299161
5	East York	Thorncliffe Park	M4H	43.705369	-79.349372	21108.0	305.0	1.444950
6	Scarborough	Scarborough Village	M1J	43.744734	-79.239476	16724.0	280.0	1.674241
7	North York	Bayview Village	M2K	43.786947	-79.385975	21396.0	410.0	1.916246
8	North York	Humber Summit	M9L	43.756303	-79.565963	12416.0	895.0	7.208441
9	York	Weston	M9N	43.706876	-79.518188	17992.0	1390.0	7.725656

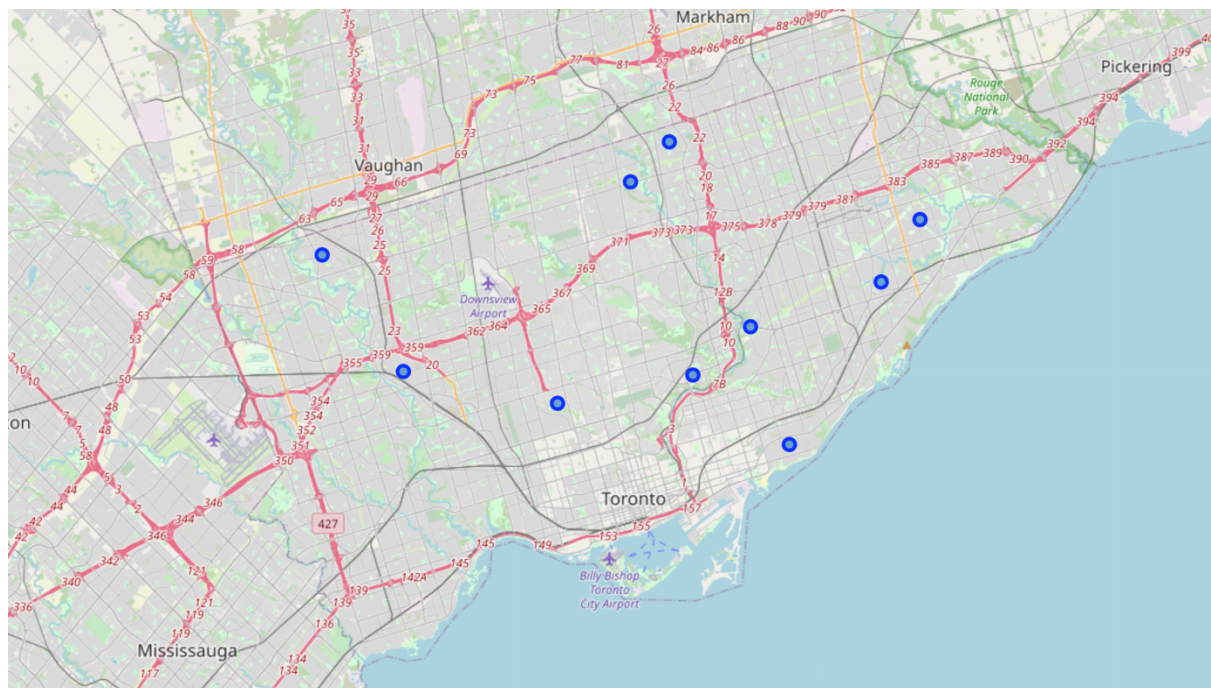
I think that there is a linear relationship between the population of Latin American and the demand for its respective cultural cuisine. So, it would make more sense to our clients to start their business projects in neighborhoods that are relatively more densely populated with Latin Americans.

Let's visualize these percentages using barplot from matplotlib



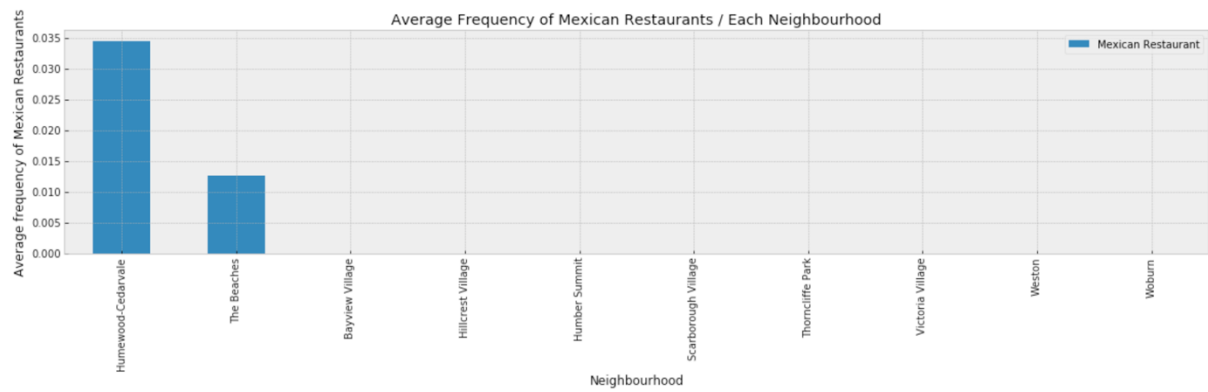
3.2 Visualization of the 10's Neighborhoods in Toronto where Latin American live in

Folium.Map



3.3 Frequency Distribution of Mexican Restaurants in each neighborhood

We can calculate the mean value of the number of Mexican restaurants located in each neighborhood with the help Foursquare API Explore function. So, it will give us a good understanding of the frequency distribution of Mexican restaurants in each neighborhood.



4. Predictive Modeling

4.1 Preprocessing and Normalization

We need to adjust values measured on different scales to a notionally common scale, so we normalize the percentage of Latin American and the percentage of Mexican restaurants columns to a common scale as follow:

	% Latin American	% Mexican Restaurants
0	-0.316726	-0.444042
1	-0.033935	2.804045
2	-0.924726	0.748294
3	0.574065	-0.444042
4	-0.910586	-0.444042
5	-0.670214	-0.444042
6	-0.740912	-0.444042
7	-0.373284	-0.444042
8	0.998251	-0.444042
9	2.398065	-0.444042

4.2 Kmeans

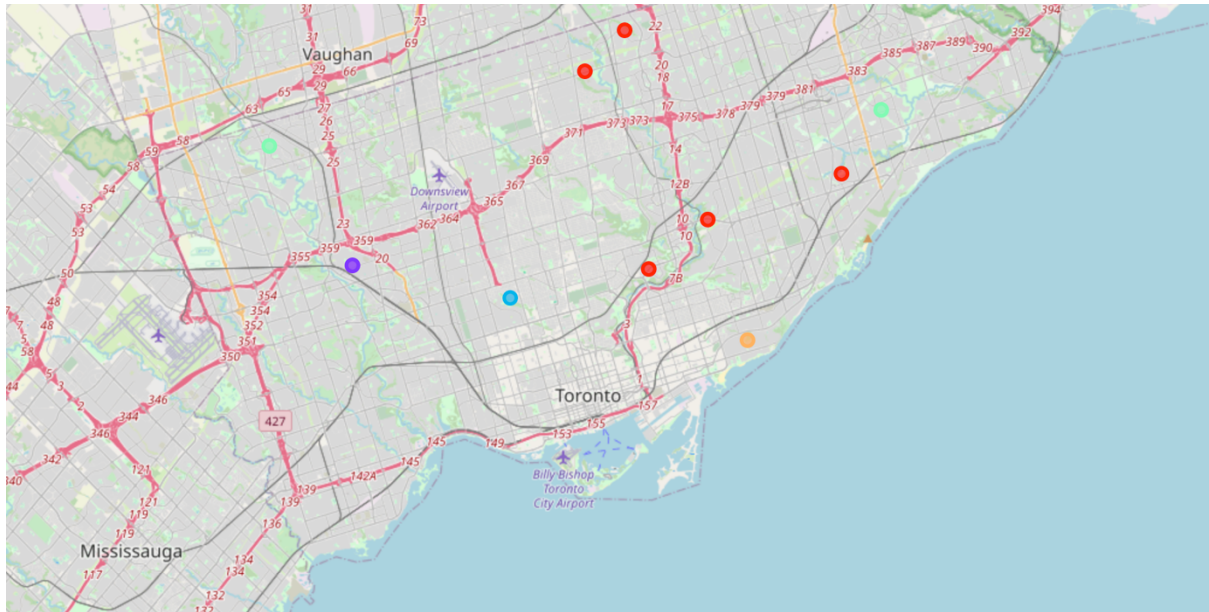
As we know in Kmeans algorithm we need to define the value of K which is the number of Clusters we will create from the kmeans algorithm. In this case I found K = 5 is the best value.

Then we add a new column to our data frame contains the labels of the clusters as follow:

Cluster Label		Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants
0	0	North York	Victoria Village	M4A	43.725882	-79.315572	17510.0	430.0	2.455740	-0.444042
1	2	York	Humewood-Cedarvale	M6C	43.693781	-79.428191	14365.0	530.0	3.689523	2.804045
2	4	East Toronto	The Beaches	M4E	43.676357	-79.293031	21567.0	215.0	0.996893	0.748294
3	3	Scarborough	Woburn	M1G	43.770992	-79.216917	53485.0	745.0	1.392914	-0.444042
4	0	North York	Hillcrest Village	M2H	43.803762	-79.363452	16934.0	220.0	1.299161	-0.444042
5	0	East York	Thornccliffe Park	M4H	43.705369	-79.349372	21108.0	305.0	1.444950	-0.444042
6	0	Scarborough	Scarborough Village	M1J	43.744734	-79.239476	16724.0	280.0	1.674241	-0.444042
7	0	North York	Bayview Village	M2K	43.786947	-79.385975	21396.0	410.0	1.916246	-0.444042
8	3	North York	Humber Summit	M9L	43.756303	-79.565963	12416.0	895.0	7.208441	-0.444042
9	1	York	Weston	M9N	43.706876	-79.518188	17992.0	1390.0	7.725656	-0.444042

4.2 Examine the Clusters

Let's visualize the Clusters on the Map



4.2.1 Cluster Label 0: (Red Circle)

Cluster Label		Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants
0	0	North York	Victoria Village	M4A	43.725882	-79.315572	17510.0	430.0	2.455740	-0.444042
4	0	North York	Hillcrest Village	M2H	43.803762	-79.363452	16934.0	220.0	1.299161	-0.444042
5	0	East York	Thornccliffe Park	M4H	43.705369	-79.349372	21108.0	305.0	1.444950	-0.444042
6	0	Scarborough	Scarborough Village	M1J	43.744734	-79.239476	16724.0	280.0	1.674241	-0.444042
7	0	North York	Bayview Village	M2K	43.786947	-79.385975	21396.0	410.0	1.916246	-0.444042

- MID percentage of Latin American
- LOW percentage of Mexican Restaurants

In this Cluster, Although the percentage of Latin American people is middle and the percentage of number of Mexican restaurants in these neighborhoods is very low, choosing this cluster wouldn't be the best option.

4.2.2 Cluster Label 1: (Purple Circle)

Cluster Label	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants	
9	1	York	Weston	M9N	43.706876	-79.518188	17992.0	1390.0	7.725656	-0.444042

- HIGH percentage of Latin American
- LOW percentage of Mexican Restaurants

In this Cluster, the percentage of Latin American people is high and the percentage of number of Mexican restaurants in these neighborhoods is very low, choosing this cluster would be one of the best options.

4.2.3 Cluster Label 2: (Blue Circle)

Cluster Label	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants	
1	2	York	Humewood-Cedarvale	M6C	43.693781	-79.428191	14365.0	530.0	3.689523	2.804045

- MID percentage of Latin American
- HIGH percentage of Mexican Restaurants

In this Cluster, the percentage of Latin American people is middle and the percentage of number of Mexican restaurants in these neighborhoods is high, choosing this cluster wouldn't be a right decision, because of the number of Mexican restaurants as competitors in these areas may lead to failure.

4.2.4 Cluster Label 3: (Green Circle)

Cluster Label	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants	
3	3	Scarborough	Woburn	M1G	43.770992	-79.216917	53485.0	745.0	1.392914	-0.444042
8	3	North York	Humber Summit	M9L	43.756303	-79.565963	12416.0	895.0	7.208441	-0.444042

- HIGH percentage of Latin American
- LOW percentage of Mexican Restaurants

In this Cluster, the percentage of Latin American people is high and the percentage of number of Mexican restaurants in these neighborhoods is very low, choosing this cluster would be one of the best options in case of "Humber Summit" neighborhood.

4.2.5 Cluster Label 4: (Orange Circle)

Cluster Label	Borough	Neighborhood	Postal Code	Latitude	Longitude	Total Population	Latin American	latin amer %	% Mexican Restaurants	
2	4	East Toronto	The Beaches	M4E	43.676357	-79.293031	21567.0	215.0	0.996893	0.748294

- LOW percentage of Latin American
- LOW percentage of Mexican Restaurants

In this Cluster, the percentage of Latin American people is low and the percentage of number of Mexican restaurants in these neighborhoods is low, choosing this cluster wouldn't be a right decision.

5. Conclusion

After all of the analysis above, we will come up with “Cluster Label” 1 is the best option we have.

The high population of Latin American people who most probably prefer to enjoy their cuisines such as Mexican food, and the very low number of Mexican restaurants in this neighborhood, will satisfy our goal from this project and the possibility of achieving the expected profits is very high.

So, I would suggest to our clients to consider “Weston” neighborhood as the optimal location to establish a new Mexican restaurants, but if our clients are seeking to expand a franchised restaurants, I would recommend “Cluster Label 3” cause it has two neighborhoods with a good percentage of Latin American like “Humber Summit” and “Woburn” neighborhoods.

6. Future Directions

This project can be improved to define the exact address in a certain area at a certain neighborhood. We have to collect more information from the map of the areas and specify the addresses and do our analysis to find the optimal location with the exact address.