

## Capstone Project - The Battle of Neighborhoods

### Business Problem



#### **Introduction:**

In this is a capstone project for Applied Data Science Capstone for IBM Data Science Professional Certificate. I will try to replicate a hypothetical situation where an entrepreneur is interested in opening a refined Spanish restaurant in the city of Toronto. Spanish cuisine, more specifically, mediterranean gastronomy is a very popular and world wide famous for its balanced diet and excellent tastes. Entrepreneur are constantly trying to find new markets to establish and thanks to data analytics we can give insights of possible opportunities.

#### **Business goal:**

Find the best possible location for a refined Spanish restaurant in the city of Toronto. Thanks to data science methodologies and tools such as machine learning algorithms (clustering) I hope to give answer to the question: "Where do I have to locate my refined Spanish restaurant?".

#### **Target audience:**

Any entrepreneur/investor who is looking foward to opening a new business in the city of Toronto.

#### **Data:**

### ***Selection criteria***

For the purposes of this project, the definition of a good location is one that has the absence or fewest identical or similar commercial competitor presence and potentially high income (is normally associated to high housing prices, where upper class people are established):

1. Find location with similar restaurant type competitors (italian, greek)
2. Compare average housing prices

### ***Work flow***

To solve this problem we will need to:

1. List neighborhoods in Toronto city.
2. Location of the different neighborhoods (latitude, longitude).
3. Explore venue data about spanish restaurants, helping us locate a better location.
4. Explore average house pricing of the different neighborhoods.
5. Locate more optimum location for restaurant opening.

### ***Libraries that we will be using:***

*Pandas*: For creating and manipulating dataframes

*Folium*: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

*Scikit Learn*: For importing k-means clustering

*JSON*: Library to handle JSON files

*Geopy*: To retrieve Location Data

*Requests*: Library to handle http requests

*Matplotlib*: Python Plotting Module

For the classification algorithm that we will use will be k-means clustering algorithm. This will help us segment and group similar locations and neighborhoods.

### ***Sources of data:***

The data will be obtained using wikipedia by data scraping techniques thanks to the help of BeautifulSoup packages. Longitude and latitude for the different neighborhoods will be given by the use of Geocoder package. After we have the coordinates we will be able to search for the different venue data thanks to the use of the Foursquare API. K-means algorithm will help to cluster the different neighborhoods.

### **Methodology:**

The methodology used for this business case is the following:

1.- First we create the data frame for all Toronto boroughs and their latitude and longitude.

	<b>Postal Code</b>	<b>Borough</b>	<b>Neighborhood</b>
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...
11	M1R	Scarborough	Wexford, Maryvale
12	M1S	Scarborough	Agincourt
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan
14	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...
15	M1W	Scarborough	Steeles West, L'Amoreaux West
16	M1X	Scarborough	Upper Rouge
17	M2H	North York	Hillcrest Village
18	M2J	North York	Fairview, Henry Farm, Oriole

	<b>Postal Code</b>	<b>Latitude</b>	<b>Longitude</b>
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

2.- We only keep the data for the value Scarborough in borough.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3.- Preview Scarborough map.



4.- Using the Foursquare API we explore the venues in a radius of 700 meters and limiting to 100. Finding the near venues and its categories, extracting top ten categories

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	Disney Store	[{"id": '4bf58dd8d48988d1f3941735', 'name': 'T...	43.775537	-79.256833
1	St. Andrews Fish & Chips	[{"id": '4edd64a0c7ddd24ca188df1a', 'name': 'F...	43.771865	-79.252645
2	SEPHORA	[{"id": '4bf58dd8d48988d10c951735', 'name': 'C...	43.775017	-79.258109
3	DAVIDsTEA	[{"id": '4bf58dd8d48988d1dc931735', 'name': 'T...	43.776320	-79.258688
4	American Eagle Outfitters	[{"id": '4bf58dd8d48988d103951735', 'name': 'C...	43.776012	-79.258334

	name	categories	lat	lng
0	Disney Store	Toy / Game Store	43.775537	-79.256833
1	St. Andrews Fish & Chips	Fish & Chips Shop	43.771865	-79.252645
2	SEPHORA	Cosmetics Shop	43.775017	-79.258109
3	DAVIDsTEA	Tea Room	43.776320	-79.258688
4	American Eagle Outfitters	Clothing Store	43.776012	-79.258334

```

Clothing Store      7
Restaurant          4
Coffee Shop         4
Cosmetics Shop      2
Pharmacy            2
Sandwich Place      2
Gas Station         2
Tea Room            2
Italian Restaurant  1
Jewelry Store       1
Name: categories, dtype: int64

```

And finding that we have 80 unique categories.

5.- Prepare data frame for clustering using one hot encoding.

	Neighborhood	Asian Restaurant	Athletics & Sports	Auto Garage	Badminton Court	Bakery	Bank	Bar	Beer Store	Board Shop	Breakfast Spot	Brewery
0	Agincourt	0.000000	0.000000	0.0	0.1	0.000000	0.000000	0.0	0.0	0.000000	0.1	0.0
1	Birch Cliff, Cliffside West	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0
2	Cedarbrae	0.055556	0.055556	0.0	0.0	0.111111	0.055556	0.0	0.0	0.055556	0.0	0.0
3	Clarks Corners, Tam O'Shanter, Sullivan	0.000000	0.000000	0.0	0.0	0.000000	0.040000	0.0	0.0	0.000000	0.0	0.0
4	Cliffside, Cliffcrest, Scarborough Village West	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0

## 6.- Top ten venues in the Scarborough neighborhoods and top five.

	Neighborhood	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	9th Most Common Venue
0	Agincourt	Clothing Store	Coffee Shop	Badminton Court	Lounge	Latin American Restaurant	Skating Rink	Pool Hall	Breakfast Spot	Motorcycle Shop
1	Birch Cliff, Cliffside West	Park	Thai Restaurant	College Stadium	Diner	Skating Rink	Café	General Entertainment	Gas Station	Fried Chicken Joint
2	Cedarbrae	Indian Restaurant	Bakery	Coffee Shop	Hakka Restaurant	Athletics & Sports	Bank	Board Shop	Caribbean Restaurant	Chinese Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	Coffee Shop	Intersection	Convenience Store	Pharmacy	Pizza Place	Gas Station	Sandwich Place	Bus Stop	Italian Restaurant
4	Cliffside, Cliffcrest, Scarborough Village West	Wings Joint	Burger Joint	Pharmacy	Chinese Restaurant	Hardware Store	Grocery Store	Greek Restaurant	Coffee Shop	College Stadium

---Agincourt---

venue	freq
0 Coffee Shop	0.1
1 Breakfast Spot	0.1
2 Clothing Store	0.1
3 Pool Hall	0.1
4 Motorcycle Shop	0.1

---Cedarbrae---

venue	freq
0 Indian Restaurant	0.11
1 Bakery	0.11
2 Coffee Shop	0.11
3 Asian Restaurant	0.06
4 Hakka Restaurant	0.06

---Cliffside, Cliffcrest, Scarborough Village West---

venue	freq
0 Hardware Store	0.2
1 Burger Joint	0.2
2 Pharmacy	0.2
3 Chinese Restaurant	0.2
4 Wings Joint	0.2

---Birch Cliff, Cliffside West---

venue	freq
0 Park	0.25
1 General Entertainment	0.12
2 Thai Restaurant	0.12
3 Diner	0.12
4 Café	0.12

---Clarks Corners, Tam O'Shanter, Sullivan---

venue	freq
0 Pizza Place	0.08
1 Pharmacy	0.08
2 Convenience Store	0.08
3 Intersection	0.08
4 Coffee Shop	0.08

---Dorset Park, Wexford Heights, Scarborough Town Centre---

venue	freq
0 Electronics Store	0.13
1 Indian Restaurant	0.13
2 Coffee Shop	0.07
3 Brewery	0.07
4 Fast Food Restaurant	0.07

---Golden Mile, Clairlea, Oakridge---

venue	freq
0 Intersection	0.19
1 Bakery	0.12
2 Bus Line	0.12
3 Diner	0.12
4 Coffee Shop	0.06

---Kennedy Park, Ionview, East Birchmount Park---

venue	freq
0 Coffee Shop	0.21
1 Discount Store	0.14
2 Bus Line	0.07
3 Hobby Shop	0.07
4 Light Rail Station	0.07

---Milliken, Agincourt North, Steeles East, L'Amoreaux East---

venue	freq
0 Pizza Place	0.17
1 Chinese Restaurant	0.17
2 Pharmacy	0.08
3 Noodle House	0.08
4 Caribbean Restaurant	0.08

---Guildwood, Morningside, West Hill---

venue	freq
0 Fast Food Restaurant	0.10
1 Park	0.10
2 Electronics Store	0.05
3 Moving Target	0.05
4 Pizza Place	0.05

---Malvern, Rouge---

venue	freq
0 Coffee Shop	0.33
1 Fast Food Restaurant	0.33
2 Hobby Shop	0.17
3 Spa	0.17
4 Medical Center	0.00

---Rouge Hill, Port Union, Highland Creek---

venue	freq
0 Breakfast Spot	0.50
1 Burger Joint	0.25
2 Bar	0.25
3 Asian Restaurant	0.00
4 Park	0.00

---Scarborough Village---

venue	freq
0 Ice Cream Shop	0.33
1 Convenience Store	0.17
2 Pizza Place	0.17
3 Fast Food Restaurant	0.17
4 Coffee Shop	0.17

---Wexford, Maryvale---

venue	freq
0 Middle Eastern Restaurant	0.11
1 Pizza Place	0.11
2 Asian Restaurant	0.06
3 Burger Joint	0.06
4 Korean Restaurant	0.06

---Steeles West, L'Amoreaux West---

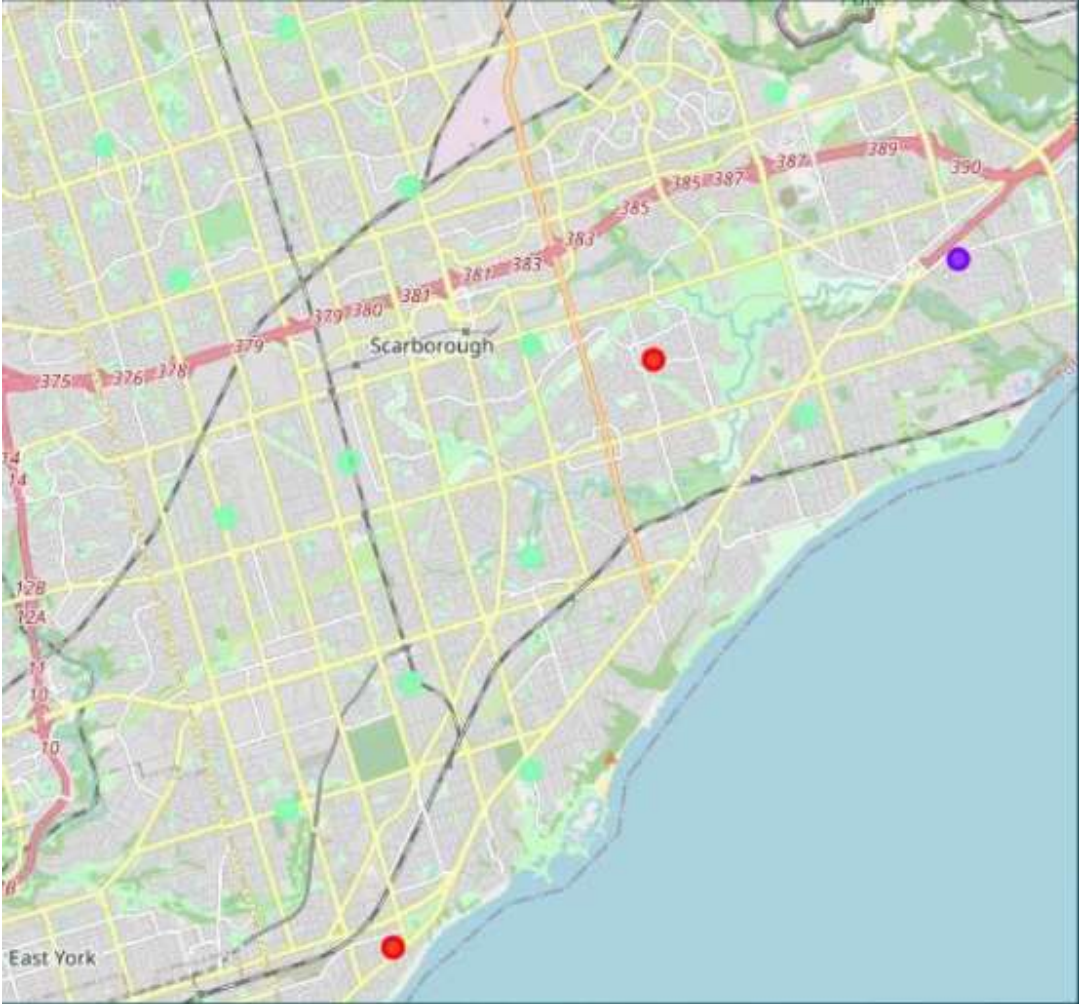
venue	freq
0 Fast Food Restaurant	0.13
1 Chinese Restaurant	0.13
2 Grocery Store	0.07
3 Breakfast Spot	0.07
4 Pizza Place	0.07

---Woburn---

venue	freq
0 Park	0.4
1 Coffee Shop	0.4
2 Business Service	0.2
3 Asian Restaurant	0.0
4 Medical Center	0.0



7.- Apply K-means algorithm (k = 3, clusters) for Scarborough borough clustering, giving as a result the following map and data frame with the top tenvenues per neighborhood. Alsowe explore the top venues of each cluster.



Postal Code	Borough	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	M1B	Scarborough	Malvern, Rouge	43.806686 -79.194353	2.0	Fast Food Restaurant	Coffee Shop	Hobby Shop	Spa	Wings Joint	College Stadium	Convenience Store	Department Store	Diner	Discount Store
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535 -79.160497	1.0	Breakfast Spot	Bar	Burger Joint	Flower Shop	Convenience Store	Department Store	Diner	Discount Store	Electronics Store	Fast Food Restaurant
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573 -79.188711	2.0	Fast Food Restaurant	Park	Moving Target	Restaurant	Grocery Store	Fried Chicken Joint	Electronics Store	Intersection	Medical Center	Mexican Restaurant
3	M1G	Scarborough	Woburn	43.770992 -79.216917	0.0	Park	Coffee Shop	Business Service	Chinese Restaurant	College Stadium	Convenience Store	Department Store	Diner	Discount Store	Electronics Store
4	M1H	Scarborough	Cedarbrae	43.773136 -79.239476	2.0	Indian Restaurant	Bakery	Coffee Shop	Hakka Restaurant	Athletics & Sports	Bank	Board Shop	Caribbean Restaurant	Chinese Restaurant	Flower Shop
5	M1J	Scarborough	Scarborough Village	43.744734 -79.239476	2.0	Ice Cream Shop	Fast Food Restaurant	Coffee Shop	Pizza Place	Convenience Store	Wings Joint	College Stadium	Department Store	Diner	Discount Store
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929 -79.262029	2.0	Coffee Shop	Discount Store	Intersection	Convenience Store	Department Store	Bus Line	Light Rail Station	Sandwich Place	Grocery Store	Hobby Shop
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112 -79.284577	2.0	Intersection	Bakery	Diner	Bus Line	Coffee Shop	Convenience Store	Park	Ice Cream Shop	Soccer Field	Pharmacy
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316 -79.239476	2.0	Wings Joint	Burger Joint	Pharmacy	Chinese Restaurant	Hardware Store	Grocery Store	Greek Restaurant	Coffee Shop	College Stadium	Convenience Store
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657 -79.264848	0.0	Park	Thai Restaurant	College Stadium	Diner	Skating Rink	Café	General Entertainment	Gas Station	Fried Chicken Joint	Flower Shop
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town	43.757410 -79.273304	2.0	Indian Restaurant	Electronics Store	Brewery	Light Rail Station	Vietnamese Restaurant	Chinese Restaurant	Wine Shop	Bakery	Gym	Pet Store
11	M1R	Scarborough	Wexford, Maryvale	43.750072 -79.295849	2.0	Middle Eastern Restaurant	Pizza Place	Asian Restaurant	Breakfast Spot	Grocery Store	Korean Restaurant	Burger Joint	Pharmacy	Rental Car Location	Restaurant
12	M1S	Scarborough	Agnicourt	43.794200 -79.262029	2.0	Clothing Store	Coffee Shop	Badminton Court	Lounge	Latin American Restaurant	Skating Rink	Pool Hall	Breakfast Spot	Motorcycle Shop	Sandwich Place
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	43.781638 -79.304302	2.0	Coffee Shop	Intersection	Convenience Store	Pharmacy	Pizza Place	Gas Station	Sandwich Place	Bus Stop	Italian Restaurant	Noodle House

### Cluster 0

```
scarborough_merged.loc[scarborough_merged['Cluster Labels'] == 0, scarborough_merged.columns[[1] + list(range(5, scarborough_merged.shape[1]))]]["1st Most Common Venue"].value_counts()[0:10]
```

Park 2  
Name: 1st Most Common Venue, dtype: int64

### Cluster 1

```
scarborough_merged.loc[scarborough_merged['Cluster Labels'] == 1, scarborough_merged.columns[[1] + list(range(5, scarborough_merged.shape[1]))]]["1st Most Common Venue"].value_counts()[0:10]
```

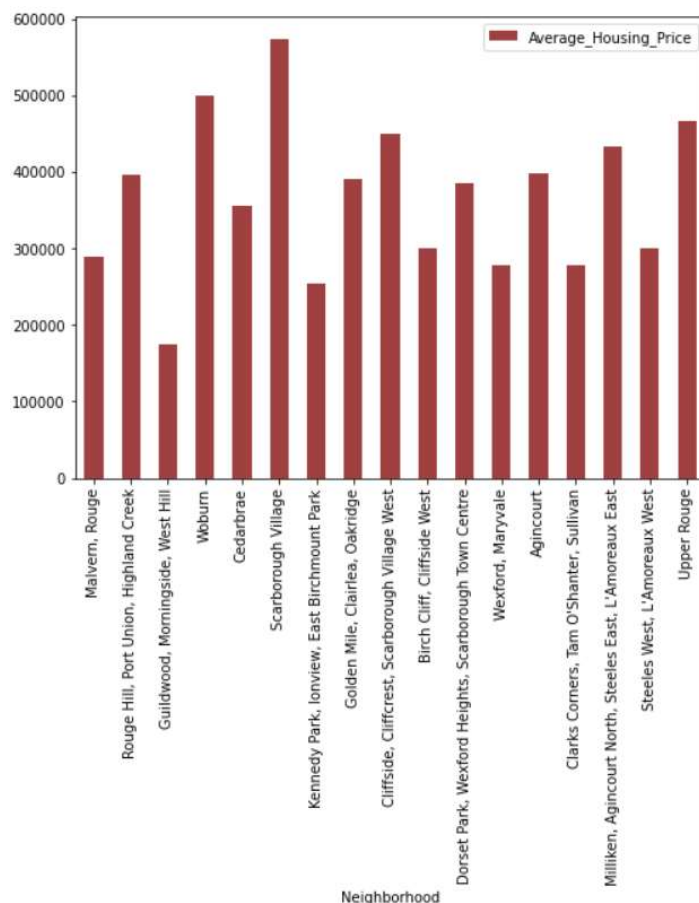
Breakfast Spot 1  
Name: 1st Most Common Venue, dtype: int64

### Cluster 2

```
scarborough_merged.loc[scarborough_merged['Cluster Labels'] == 2, scarborough_merged.columns[[1] + list(range(5, scarborough_merged.shape[1]))]]["1st Most Common Venue"].value_counts()[0:10]
```

Fast Food Restaurant 3  
Coffee Shop 2  
Indian Restaurant 2  
Pizza Place 1  
Middle Eastern Restaurant 1  
Ice Cream Shop 1  
Intersection 1  
Clothing Store 1  
Wings Joint 1  
Name: 1st Most Common Venue, dtype: int64

8.- Graph average house pricing for each of the neighborhoods in the Scarborough borough.





**Conclusion:**

With the data obtained the best location for refined Spanish restaurant in the borough of Scarborough (Toronto, CA), would be in the neighborhood of Woburn. In this neighborhood the average of housing prices is the second highest and there is a lack of restaurants in the area. This is why the best location would be in Woburn neighborhood.