Best location for a coffee shop

IBM Data Science Capstone Project

I.   Introduction

The integration of machine learning and geography can provide novel approches in addressing a variety of problems occuring in natural and human environments.this article presents an expriment that identifies the best neighborhoods in the city of toronto , which appears as the largest city of canada with an estimation of 6 millions population out there.

Using machine learnings and data science techniques we will try to determine what would the best place (neighboorhood) to start a business (opening a coffee shop).

1. Target Audience

This analysis might be helpful for Entrepreneurs or Business owners
who want to open a new coffee shop or grow their current business.

2. Libraries Used to Develope the Project:

- Matplotlib: Python Plotting Module.
- Pandas: For manipulating dataframes.
- XML: To separate data from presentation and XML stored data.
- Beautiful Soup and Requests: To scrap and library to handle http reque sts.
- Folium: Python visualization library
- Scikit Learn: For importing k-means clustering.

II.   Data aquisation

to solve this problem and get the analysis start we firt need data to act on , the data needed for this project is a combination of CSV files that have been prepared for the purposes of the analysis which will provide the list of neighbourhoods in Toronto. we also need the "Foursquare" for locational information and the Geographical location using (Geocoder package).

As mentioned above we firt get the toronto Neighborhoods via Wikipedia using This link.

the csv file contains almost all the information about the neighborhoods of Toronto . which includes the postal code, borough and the name of the neighborhoods.

III. Methodology section
Data cleaning and preparation

In this section we discuss and describe all exploratory data analysis that we did, and the inferential statistical testing that we performed. so after all the aquisation is done, we must clean our data in order to perform diffrent analysis. for that we have done the follwing

- The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
- Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned.
- More than one neighborhood can exist in one postal code area.
- If a cell has a borough but a Not assigned, neighborhood then the neighborhood will be the same as the borough.

After dowing this the data now look clean and ready to be analysed as shown in the picture bellow

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

figure.**Error! No text of specified style in document.**.1 the data frame table.

Now we add the geographical coordinates using this link . the result is shown bellow

	Postal Code	Latitude	Longitude
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

Figure 2 : The geospatial location.

After getting the data and preparing it, now it's time to start analyzing it using the diffrent technics so we can get insghits from this data.

for that purpous we use folium to print and show the diffrent locations and their proprieties as shown in the picture bellow.
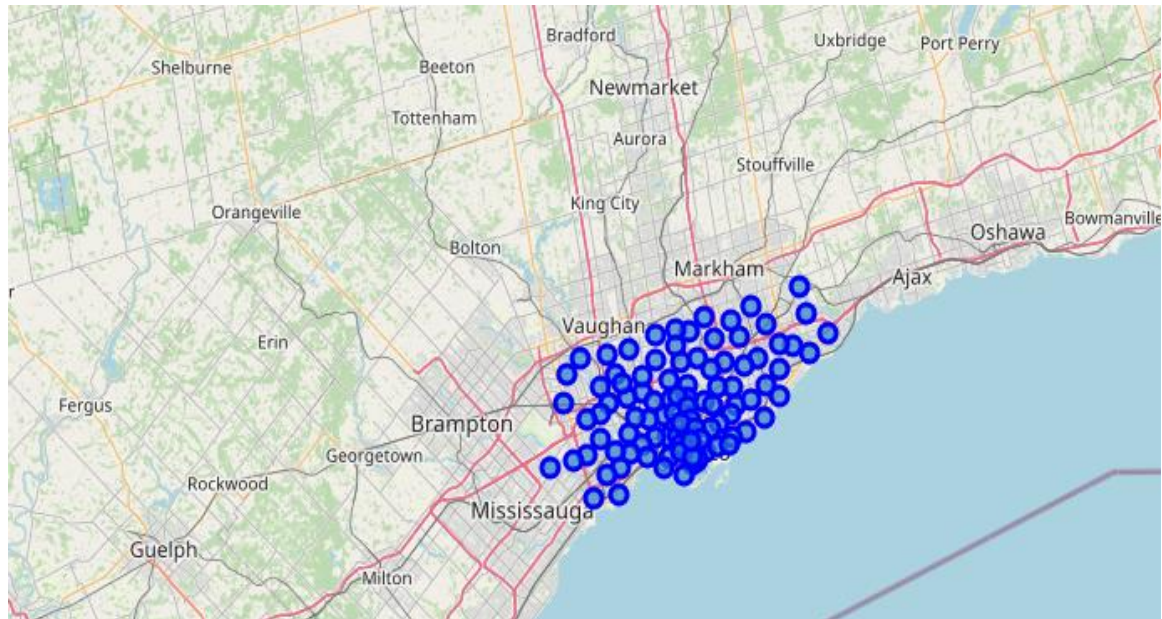

Figure 3 : Neighborhoos of toronto.

Now that the region has been clearly visualized, the Foursquare API was used to explore each neighborhood and return the top 200 venues. And clustering into 5 cluster like we can see in the picture below.
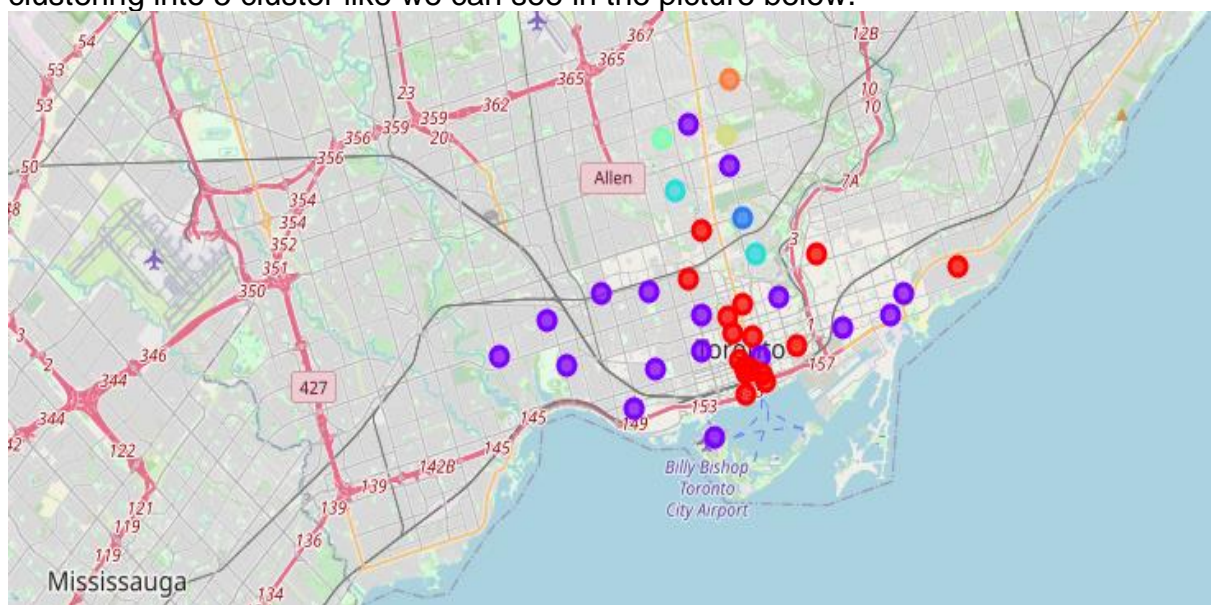

Figure 3 : the five cluster of the toronto city.

After that we can print the five cluster into screen to see what businesses were successful in each cluster.

	Borough	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	East Toronto	0	Asian Restaurant	Coffee Shop	Trail	Pub	Health Food Store	Dog Run	Dessert Shop	Diner	Discount Store	Distribution Center
1	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Yoga Studio	Pub	Indian Restaurant	Spa	Caribbean Restaurant
9	Central Toronto	0	Coffee Shop	American Restaurant	Liquor Store	Restaurant	Bank	Bagel Shop	Pub	Supermarket	Fried Chicken Joint	Sushi Restaurant
12	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Yoga Studio	Mediterranean Restaurant	Pub	Hotel	Smoke Shop
13	Downtown Toronto	0	Coffee Shop	Bakery	Park	Breakfast Spot	Café	Pub	Theater	Cosmetics Shop	Beer Store	Shoe Store

Figure 4: first cluster.

	Borough	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
2	East Toronto	1	Park	Fast Food Restaurant	Gym	Pub	Sandwich Place	Brewery	Italian Restaurant	Restaurant	Ice Cream Shop	Steakhouse
3	East Toronto	1	Coffee Shop	American Restaurant	Brewery	Gastropub	Café	Bakery	Yoga Studio	Ice Cream Shop	Bank	Bar
6	Central Toronto	1	Clothing Store	Coffee Shop	Yoga Studio	Shoe Store	Salon / Barbershop	Diner	Spa	Sporting Goods Shop	Restaurant	Fast Food Restaurant
7	Central Toronto	1	Pizza Place	Dessert Shop	Sandwich Place	Italian Restaurant	Coffee Shop	Café	Sushi Restaurant	Gym	Gourmet Shop	Greek Restaurant
11	Downtown Toronto	1	Pizza Place	Coffee Shop	Café	Italian Restaurant	Restaurant	Pub	Chinese Restaurant	Bakery	Pet Store	Pharmacy
15	Downtown Toronto	1	Coffee Shop	Café	Cocktail Bar	Restaurant	Gastropub	Gym	Clothing Store	Park	Moroccan Restaurant	Lingerie Store

Figure 5: second cluster with revenue.

	Borough	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
8	Central Toronto	2	Gym	Trail	Department Store	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant

Figure 6 : third cluster with revenue.

	Borough	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
10	Downtown Toronto	3	Park	Trail	Playground	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant
23	Central Toronto	3	Park	Trail	Jewelry Store	Bus Line	Sushi Restaurant	Department Store	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop

Figure 7 : the fourth cluster.

	Borough	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
22	Central Toronto	4	Home Service	Garden	Wine Bar	Department Store	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop

Figure 8 : the five cluster.

IV. **Discussion**

From the analysis above It's concluded that the best neighborhood to open a coffee shop is cluster number one since it was clear that the first most common revenue was from coffee shop.

V. Conclusion

Starting a business in a new area is frustrating specially if you are new in a city and don't now the people neither their attitudes, data analysis and data science can help as we have done in this mini project. The aim of this project was to use different machine learning technics and libraries to segment and cluster the neighborhood of Toronto city , in order to find and caption the best neighborhood to open a coffee shop. as we conclude this report, we answered the business question introduced at the beginning