Finding a suitable location with respect to relevant business opportunity

1. Introduction and Business Problem:

A client would like to open a new business; however the client is confused about the following:

a) *Which type of business to open at which location?*

In this jupyter notebook, I will analyse the optimal areas in Canada to open a new business, along with their respective category

For the aforementioned cited problem, I will analyse the neighborhoods of Canada, using the Foursquare API. Also, through the help of data wrangling, and analysis, I will be able to make clusters using Kmeans machine learning model, to suggest the right business at the right locations.

2. Sources of Data:

For the problem cited above, the data would be gathered from multiple locations listed below:

a) Wikipedia - Areas of Canada- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

b) Foursquare API to collect venue information for each area.

c) Point "B" will be used to determine the top venues in each area.

After the aforementioned data has been collected, clusters will be formed based on venue categories using K means Machine Learning model. This will allow the model to determine not only the best business idea but its location at the same time.

3. Methodology:

The Wikipedia page was loaded into the jupyter notebook to analyse the data-set present in it. Firstly, some of the data cleaning and wrangling aspect was completed by renaming some of the names, and making the columns consistent for analysis. After that, the location values (both latitude and longitude) with respect to Postal Code, Neighborhood and Borough were appended to the data set of Wikipedia page from the data set of Geospatial Coordinates of locations in Canada, for effective use of grouping functionality.

After importing the relevant libraries, like folium, numpy, pandas, seaborn etc. Foursquare data set was used to append the 10 most common venues in particular locations with the above created data-set.

For the process of getting most common venue with respect to specific neighborhoods, so attributes with two or more distinct categories could be represented effectively.

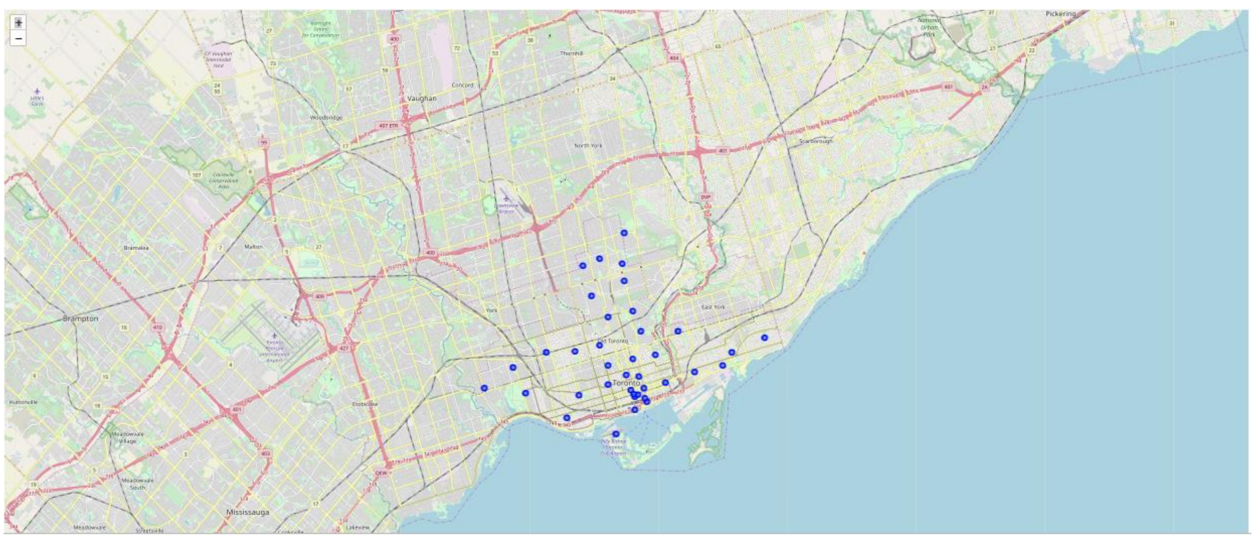
Then the machine learning model of Kmeans was used effectively through cluster analysis of most common venues with respect to relevant neighborhoods.

4. Results

4.1. Appending of Latitude and Longitude with Neighborhoods, Postal Codes etc.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	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

4.2. Toronto and its Neighborhoods mapped through geospatial coordinates



4.3 Most common venues with respect to the neighborhoods extracted

```
The Beaches
The Danforth West, Riverdale
The Beaches West, India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park, Summerhill East
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
Rosedale
Cabbagetown, St. James Town
Church and Wellesley
Harbourfront, Regent Park
Ryerson, Garden District
St. James Town
Berczy Park
Central Bay Street
Adelaide, King, Richmond
Harbourfront East, Toronto Islands, Union Station
Design Exchange, Toronto Dominion Centre
Commerce Court, Victoria Hotel
Roselawn
Forest Hill North, Forest Hill West
The Annex, North Midtown, Yorkville
Harbord, University of Toronto
Chinatown, Grange Park, Kensington Market
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara
Stn A PO Boxes 25 The Esplanade
First Canadian Place, Underground city
Christie
Dovercourt Village, Dufferin
Little Portugal, Trinity
Brockton, Exhibition Place, Parkdale Village
High Park, The Junction South
Parkdale, Roncesvalles
Runnymede, Swansea
Business Reply Mail Processing Centre 969 Eastern
```

4.4. Creation of Dummy Variables (Extract):

	Neighborhoods	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Wings Joint
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0

5 rows × 238 columns

4.5. Grouping of Variables, with respect to neighborhoods:

	Neighborhoods	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar
0	Adelaide, King, Richmond	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.04	...	0.0	0.0	0.0	0.010000	0.0	0.0	0.01
1	Berczy Park	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	...	0.0	0.0	0.0	0.017544	0.0	0.0	0.00
2	Brockton, Exhibition Place, Parkdale Village	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.00
3	Business Reply Mail Processing Centre 969 Eastern	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.00
4	CN Tower, Bathurst Quay, Island airport, Harbo...	0.0	0.0	0.066667	0.066667	0.066667	0.133333	0.2	0.133333	0.00	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.00

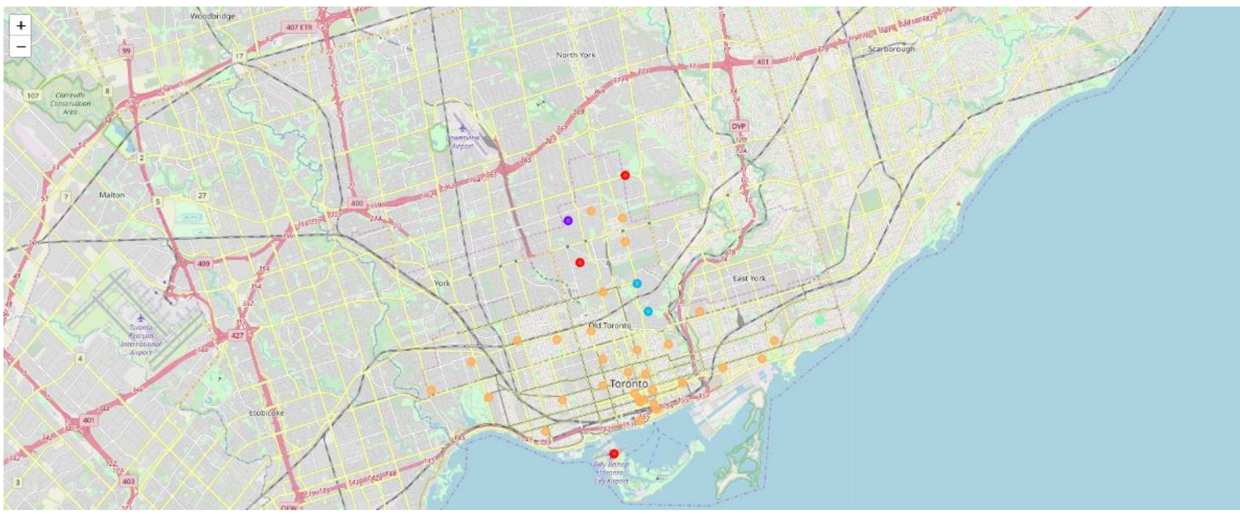
5 rows × 238 columns

4.6. Most frequent venues sorted with respect to their neighborhoods:

	Neighborhoods	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	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Thai Restaurant	American Restaurant	Bar	Hotel	Restaurant	Bakery	Burger Joint
1	Berczy Park	Coffee Shop	Cocktail Bar	Restaurant	Café	Seafood Restaurant	Cheese Shop	Farmers Market	Steakhouse	Beer Bar	Bakery
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Grocery Store	Climbing Gym	Burrito Place	Stadium	Bar	Restaurant	Caribbean Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Auto Workshop	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Comic Shop	Park	Recording Studio
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Terminal	Airport Lounge	Plane	Harbor / Marina	Sculpture Garden	Boutique	Boat or Ferry	Airport Gate	Airport Food Court

ta/.../534d22bc-7dca-408e-b0ad-06999ad6e7f5?pr...

4.7. Use of Folium Library:



4.8. K means Cluster 1-3 results (Part of the results) – Machine Learning Model:

Out[27]:

	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	3	Health Food Store	Pub	Neighborhood	Music Venue	Convenience Store	Cosmetics Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

```
]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[28]:

	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	Trail	Playground	Tennis Court	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Yoga Studio
10	Downtown Toronto	2	Park	Trail	Playground	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Yoga Studio

```
]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 1
                     , toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[29]:

	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	1	Garden	Filipino Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop

5. Discussion:

The most common venue in the first cluster is Garden in the Central Toronto and the least common is a Donut Shop. Whereas when the results of cluster 3 in East Toronto are compared, the most common venue is Health Food Store, and the least common one is Electronics Store. Hence the type and locations of businesses to open can be clearly seen through the machine learning model of Kmeans after proper data wrangling and visualization.

6. Conclusion:

Well the above model can be refined with more machine learning techniques and more advanced analysis with same or different parameters. These techniques can effectively assist any businessmen in any place in the world to figure out the most profitable businesses in his/her state/province, district and even to small denominations of areas.