Find a suitable location to open a restaurant in Toronto

Abhishek Deb

April 2020

Introduction

- *As the population is constantly increasing, the requirement of Different services are also increasing. Hence, Opening a Restaurant or a chain of Restaurant is a very common business idea For any entrepreneur nowadays. The Success and profit of the restaurant are dependent on several factors like locality, type, services etc. One of the very first problems is to find a suitable locality for the new Restaurant as location plays a big role in the success of Restaurant.*
- **The goal of this project is to provide the answer to the business question "In Toronto if an entrepreneur wants to open a new restaurant, which neighbourhood/neighbourhoods will be suitable for this?**

Data acquisition

- In this project, we have used the below data to find the similar Neighbourhood.
  - i. Different neighbourhood information of Toronto City from The official website for the City of Toronto (www.toronto.ca).
  - ii. Latitude and Longitude of all neighbourhoods using Geocoder.
  - iii. Different venue details based on all neighbourhoods using Foursquare API

Data Cleaning and Modification

- A. Unwanted observations and missing data removal.
- B. correcting Datatypes for calculation

```
df.dtypes
```

```
Neighbourhood_Name              object
Population_2016                 object
Population_Change_2011_2016     object
Population_density              object
Children                        object
Youth                           object
Working_Age                     object
Pre_retirement                  object
Seniors                         object
Older_Seniors                   object
Family_2_persons                object
Family_3_persons                object
Family_4_persons                object
```

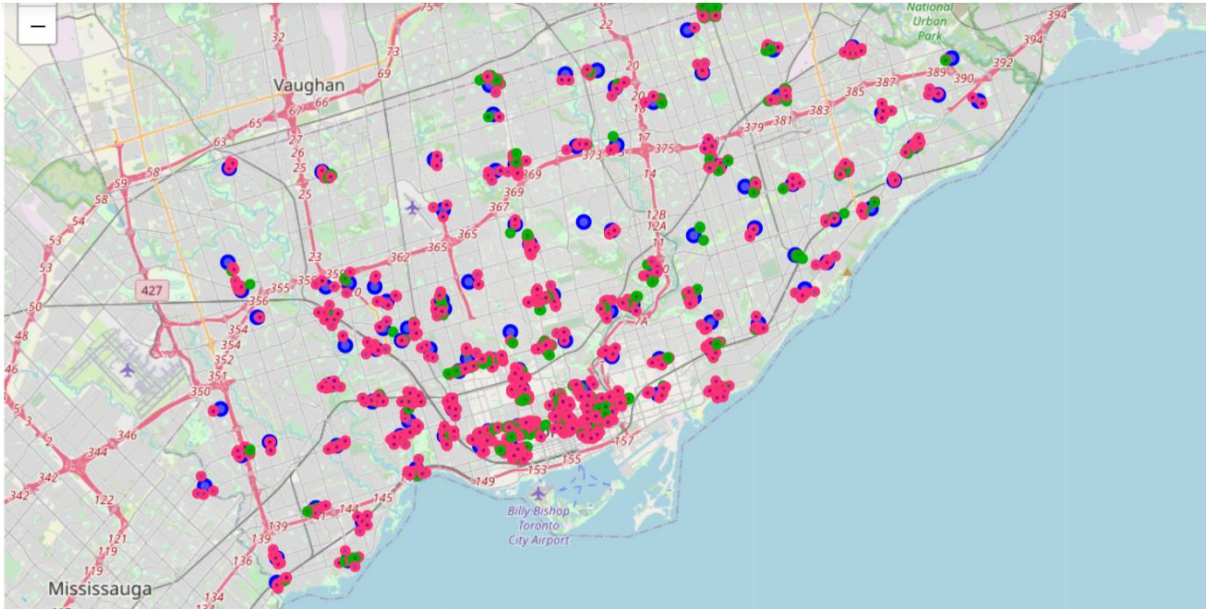
```
df.dtypes
```

```
Neighbourhood_Name              object
Population_2016                  int64
Population_Change_2011_2016    float64
Population_density               int64
Children                         int64
Youth                            int64
Working_Age                      int64
Pre_retirement                   int64
Seniors                          int64
Older_Seniors                    int64
Family_2_persons                 int64
Family_3_persons                 int64
Family_4_persons                 int64
```

- C. Merging related columns

Exploratory Data Analysis

- We have plotted data in maps to show the distribution of restaurant and other venues



Feature Selection

- It has observed in machine learning that if we use important features instead of all features, the machine learning algorithm's output is better.
  - i. Filtering population related data
  - ii. Filtering Venue related data

Final selected features:

```
['Living_alone', 'Population_density', 'Arcade', 'Art Gallery', 'Arts & Crafts Store', 'BBQ Joint', 'Bagel Shop', 'Bakery', 'Ba
nk', 'Bar', 'Beer Bar', 'Bookstore', 'Boutique', 'Breakfast Spot', 'Bubble Tea Shop', 'Burger Joint', 'Café', 'Chiropractor',
'Cocktail Bar', 'Coffee Shop', 'Deli / Bodega', 'Dessert Shop', 'Diner', 'Event Space', 'Farmers Market', 'Food & Drink Shop',
'Fried Chicken Joint', 'Frozen Yogurt Shop', 'Gastropub', 'Grocery Store', 'Gym', 'Hobby Shop', 'Hospital', 'Ice Cream Shop',
'Indie Movie Theater', 'Jazz Club', 'Jewelry Store', 'Juice Bar', 'Karaoke Bar', 'Lounge', 'Miscellaneous Shop', 'Noodle Hous
e', 'Pet Store', 'Pizza Place', 'Pub', 'Record Shop', 'Restaurant', 'Rock Club', 'Sandwich Place', 'Snack Place', 'Sports Bar',
'Supermarket', 'Taco Place', 'Tea Room', 'Wine Bar', 'Yoga Studio']
```

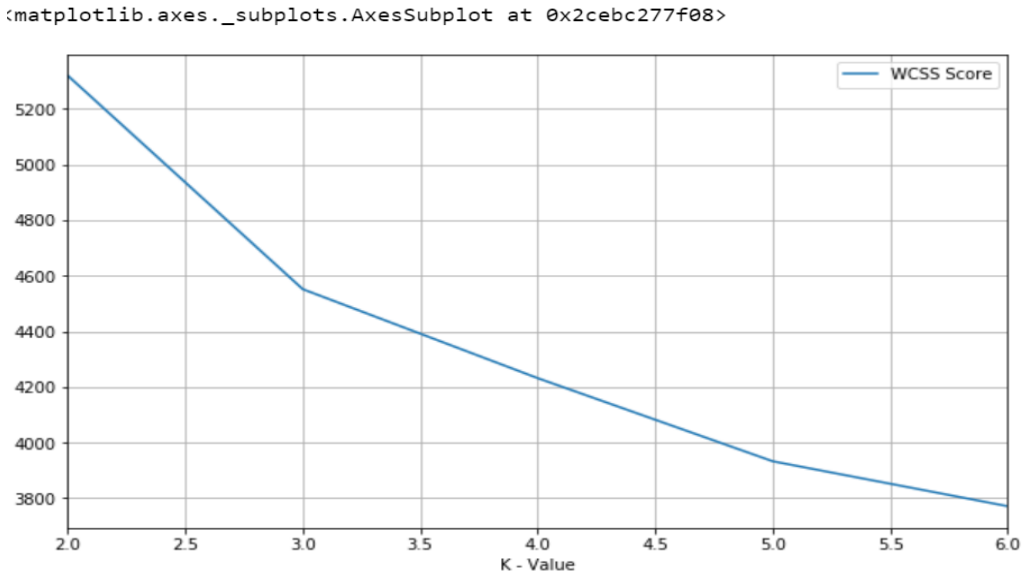
Scaling

- Feature scaling in ML algorithms is one of the most important steps, which use to affect the output of the algorithm. It helps to normalise the data within a particular range. we have used StandardScaler from SKlearn library to perform feature scaling on our data before feeding it to ML algorithm.

	Living_alone	Population_density	Arcade	Art Gallery	Arts & Crafts Store	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	...	Record Shop	Rock Club	Sandwich Place
0	-0.592582	-0.497702	-0.094916	-0.204598	-0.246183	-0.330017	-0.156174	1.701691	2.370289	-0.378658	...	-0.127804	-0.094916	0.848440
1	-0.460640	-0.668695	-0.094916	-0.204598	-0.246183	-0.330017	-0.156174	-0.580817	-0.595883	-0.378658	...	-0.127804	-0.094916	0.848440
2	-0.714751	-0.783136	-0.094916	-0.204598	-0.246183	-0.330017	-0.156174	-0.580817	-0.595883	-0.378658	...	-0.127804	-0.094916	0.848440
3	2.596013	0.827061	-0.094916	-0.204598	-0.246183	-0.330017	-0.156174	0.560437	-0.595883	-0.378658	...	-0.127804	-0.094916	0.848440
4	0.875882	-0.718178	-0.094916	-0.204598	-0.246183	-0.330017	-0.156174	-0.580817	-0.595883	-0.378658	...	-0.127804	-0.094916	-0.684226

Applying ML Algorithm :

- We have used
- K-means clustering
- here.



Cluster Labels		
0	85	2.35
1	1	21.0
2	1	24.0
3	2	11.0
4	23	12.43

Some details about Cllusters

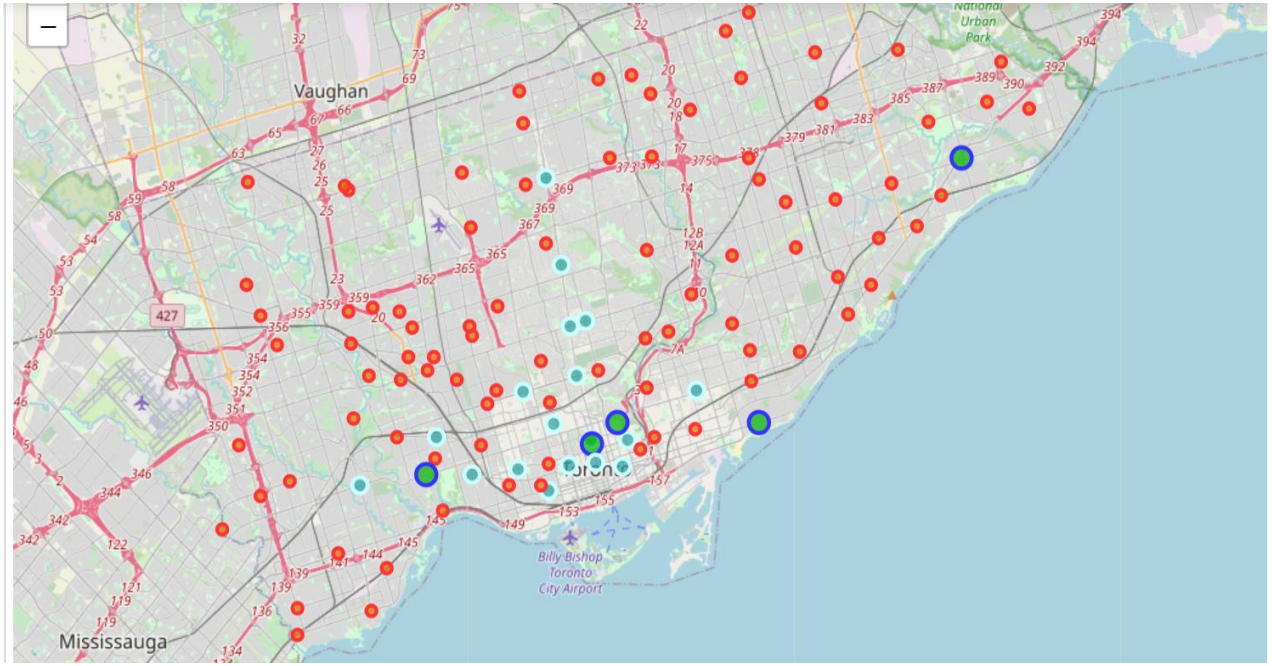
Cluster 4

Number of restaurent in each neighbourhood of cluster 4

	Neighbourhood_Name	Restaurant
107	Wychwood	23
109	Yonge-St.Clair	20
108	Yonge-Eglinton	19
30	Dufferin Grove	19
105	Willowdale East	15
60	Lawrence Park South	14
51	Junction Area	14
54	Kensington-Chinatown	14
70	Moss Park	13
3	Annex	13
18	Cabbagetown South	13
75	Niagara	12
100	University	12
56	Kingsway South	11
6	Bay Street Corridor	11
85	Roncesvalles	10
41	Greenwood	10
72	Mount Pleasant West	10
88	Runnymede-Bloor West Village	8
22	Church-Yonge Corridor	8
97	The Beaches	7
76	North St. James Town	5
102	West Hill	5

Recommendations:

- In cluster number 4 neighbourhoods with less than 10 restaurants are : 'Runnymede-Bloor West Village', 'Church-Yonge Corridor ', 'The Beaches', 'North St. James Town' and 'West Hill‘ but the neighbourhood is similar to other neighbourhood which has a higher number of restaurants.



Conclusion and future directions

- *I have gone through the process of identifying the business problem. I have analyzed, collected and prepared to find a solution performing the machine learning (i.e. k-means clustering). In the end, I have provided recommendations to the entrepreneur based on the result.*

- There are many factors that can be taken into consideration such as connectivity of the neighbourhood, rating of the existing venues, how much the neighbourhood is developing etc. Additionally, if we can categorize existing venues in a better way, it will affect the result as well. Future research can take into consideration these factors.