Capstone Project -The Battle of Neighborhoods

Geolocation analysis for opening shopping mall in Toronto

Mohamed Mostafa Fathy Ahmed Awad

February 20, 2021

Business Problem

Finding the suitable location is a main step to start any business. In this project, I will try to explore the data of Toronto city using the data of the districts and the Foursquare's 'Places API' to help finding the best location to build a shopping mall.

Dataset preparation

In order to answer the above question, data on Toronto City neighborhoods, boroughs to include boundaries, latitude, longitude, restaurants, and restaurant ratings and tips are required. Below is the list of datasets used in this project:

Datasets - Toronto City postal code data

Toronto City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

	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

Datasets - Toronto City geospatial coordinates

For each postal code get the related longitude and latitude.

Postal Code	Latitude	Longitude
M1B	43.806686	-79.194353
M1C	43.784535	-79.160497
M1E	43.763573	-79.188711
M1G	43.770992	-79.216917
M1H	43.773136	-79.239476

Datasets – Get postal code coordinates

Combine postal code and geospatial coordinates:

|	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.7533	-79.3297
1	M4A	North York	Victoria Village	43.7259	-79.3156
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.6543	-79.3606
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.7185	-79.4648
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.6623	-79.3895

Foursquare API data

Using the Foursquare API, explore the venues related to the districts of Toronto city.

	District	Population	Code	Latitude	Longitude	VenueName	VenueId	VenueLatitude	VenueLongitude	VenueCategory
0	North York	Parkwoods	M3A	43.753259	-79.329656	Allwyn's Bakery	4b8991cbf964a520814232e3	43.759840	-79.324719	Caribbean Restaurant
1	North York	Parkwoods	M3A	43.753259	-79.329656	Donalda Golf & Country Club	4bd4846a6798ef3bd0c5618d	43.752816	-79.342741	Golf Course
2	North York	Parkwoods	M3A	43.753259	-79.329656	Galleria Supermarket	4ccec87654f0b1f7f32824ca	43.753520	-79.349518	Supermarket
3	North York	Parkwoods	M3A	43.753259	-79.329656	Tim Hortons	57e286f2498e43d84d92d34a	43.760668	-79.326368	Café

Data analysis

- Total number of shopping malls for each district:

Shopping mall in Toronto
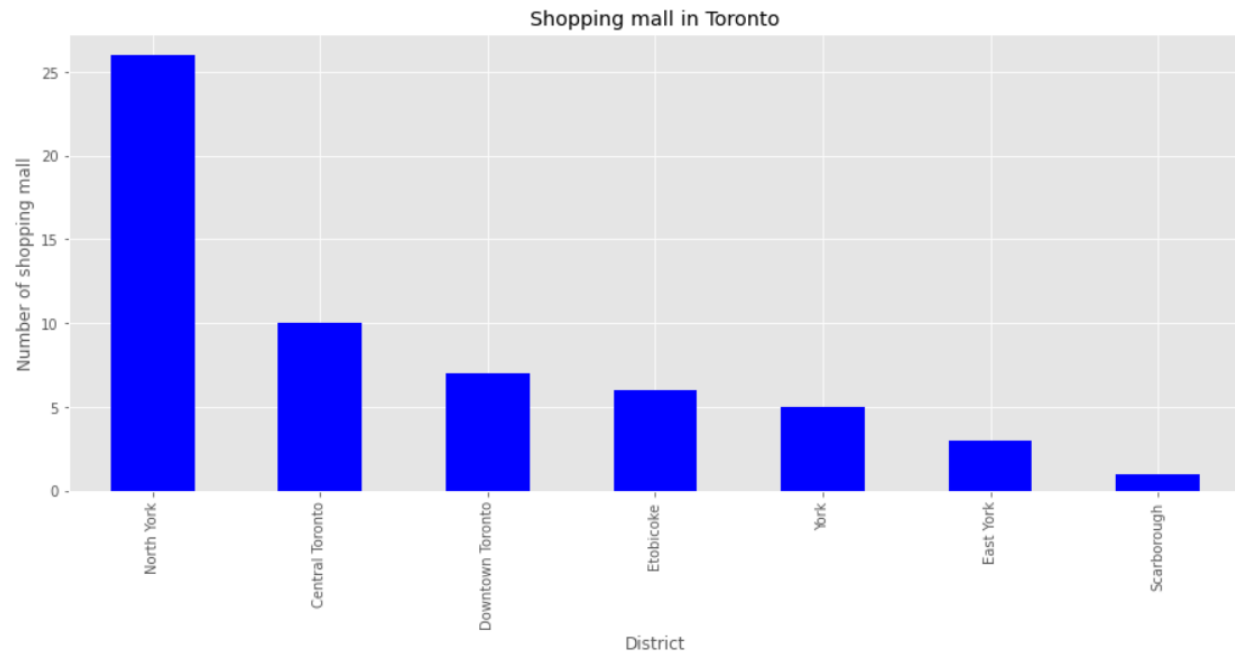
Display districts on map
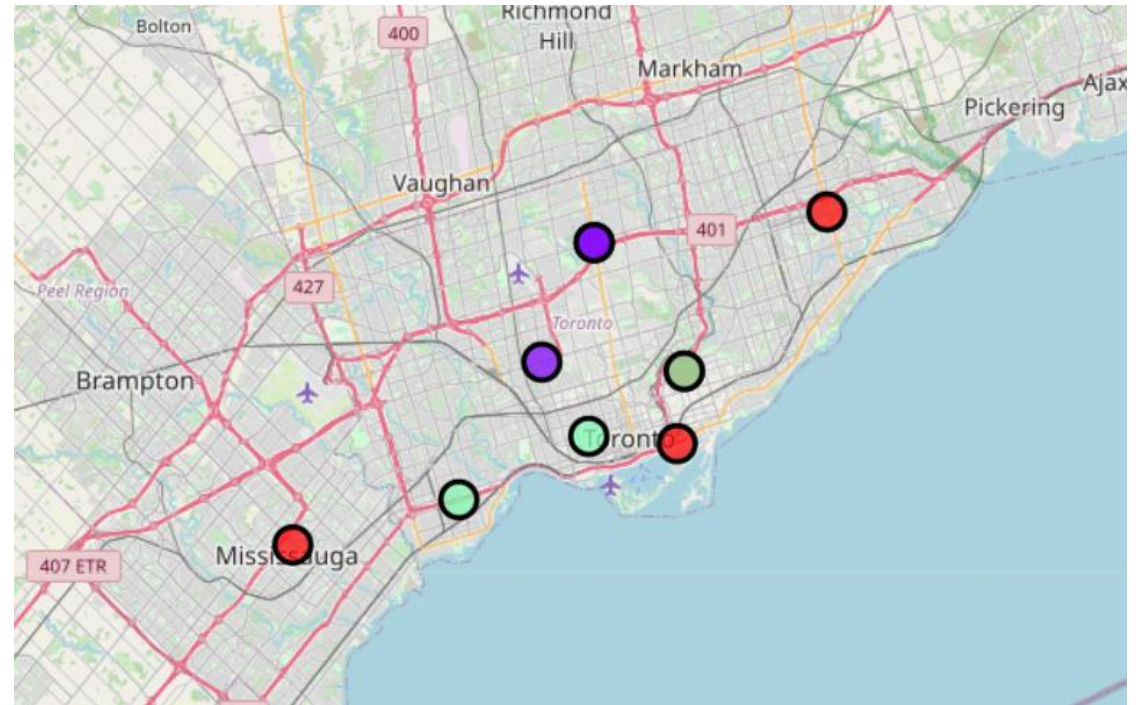
Machine Learning (K-Mean)

Using the K-Mean algorithm, we will cluster the district into 3 clusters based on their frequency of occurrence for "Shopping Mall".

Before using K-Mean, we have to transform data using one-hot encoder. The results will allow us to identify which districts have a higher concentration of shopping malls while which districts have a fewer number of shopping malls by using k-means clustering. Based on the occurrence of shopping malls in a different district, it will help us to plan which district we should acquire land for developing new shopping mall.

Results

From the results, we found that the shopping malls were clustered using number of shopping malls in each district, so it can be classified to high, medium and low number of shopping malls in each area.

However, there are other factors which can be used to cluster such as number of tourists, income of people around that area, number of people traveled around these areas. These information can make clustering more accurate and will generate more information for decision making.

Cluster data

```
df_clustered.loc[df_clustered['Cluster'] == 2].head()
```

	District	Shopping Mall	Cluster	Latitude	Longitude
1	Downtown Toronto	0.003684	2	43.654753	-79.414187
3	East York	0.006000	2	43.691200	-79.341667
4	Etobicoke	0.005076	2	43.620495	-79.513199

```
df_clustered.loc[df_clustered['Cluster'] == 0].head()
```

	District	Shopping Mall	Cluster	Latitude	Longitude
2	East Toronto	0.000000	0	43.691200	-79.341667
5	Mississauga	0.000000	0	43.595310	-79.640579
7	Scarborough	0.000592	0	43.777702	-79.233238
8	West Toronto	0.000000	0	43.651070	-79.347015

```
df_clustered.loc[df_clustered['Cluster'] == 1].head()
```

	District	Shopping Mall	Cluster	Latitude	Longitude
0	Central Toronto	0.011111	1	43.761539	-79.411079
6	North York	0.010888	1	43.761539	-79.411079
9	York	0.010000	1	43.695683	-79.450279

Conclusion

The objective of this project is to support decision making of building a new shopping mall in Toronto. Using the following process:

- Identifying the business problem.

- Specifying the required data.

- Extract and prepare the data.

- Performing machine learning by clustering the data into 3 clusters based on their similarities

- Finally providing recommendations.

References

- List of postal codes for Toronto
  - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Foursquare Developers Documentation. Foursquare. Retrieved from
  - https://developer.foursquare.com/docs