The Battle of Neighborhoods (Week 2)

Deploying data science methodology and machine learning to find the optimal location for a new yoga studio in Toronto.

Introduction

A passionate yoga instructor has decided to open her yoga studio in Toronto, Canada. Practicing yoga has physical benefits, such as increased flexibility, increased muscle strength, weight reduction, etc., and mental benefits, such as reduced stress. However, she has absolutely no idea of where precisely in Toronto should she start her studio.

Business problem

The problem is trying to answer where exactly in Toronto she should open the yoga studio. As of her knowledge, people don't need too many yoga studios nearby since they choose only one. Additionally, doing so would induce unnecessary competition. Like other business lines, she needs the least amount of competitiveness with a fair number of customers. To achieve the goal, it might not be a good idea to start a new yoga studio in the neighborhood with no yoga studio at all because it is possibly no demand in that location in the first place. The logical process may be clustering yoga studios and find the cluster with the least degree of competitiveness.

Data

- Retrieve data of neighborhood and borough of Toronto via Wikipedia URL:https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare API to retrieve the neighborhood's latitude/longitude along with venue category.

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue
6. Venue Latitude
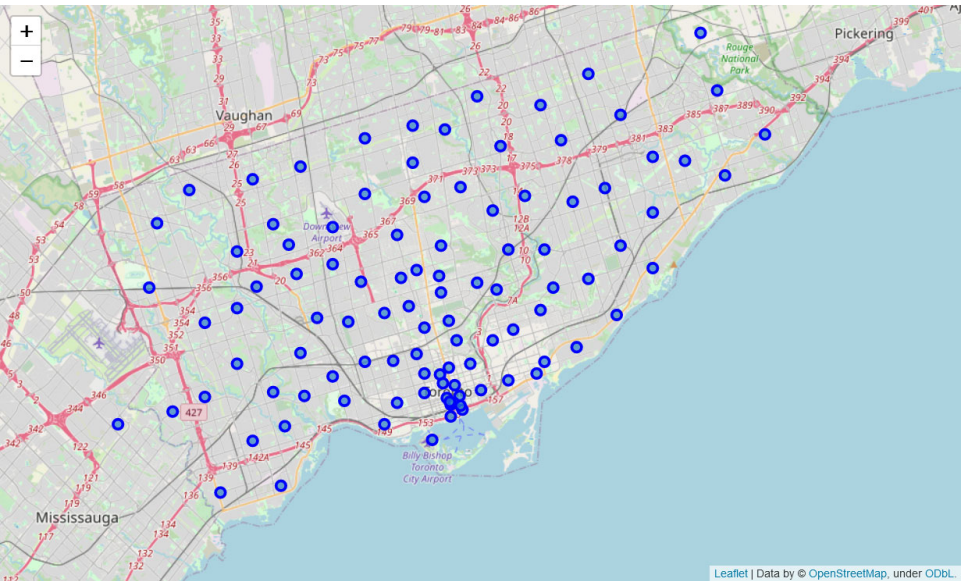7. Venue Longitude
8. Venue Category

Methodology

First, load data from Wikipedia into a data frame and load the postal CSV file into another data frame. Merge the two data frames, got data frame shows below.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Retrieve Toronto's coordinates using geopy library, then use folium library along with latitude/longitude in the data frame to generate a geographical plot of neighborhoods in Toronto.

**Exhibit 1** Geographical plot of Neighbourhood in Toronto.

Foursquare API is used to retrieve more details such as venue, venue's latitude/longitude, and venue category.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

Transform Venue Category into one-hot encoding.

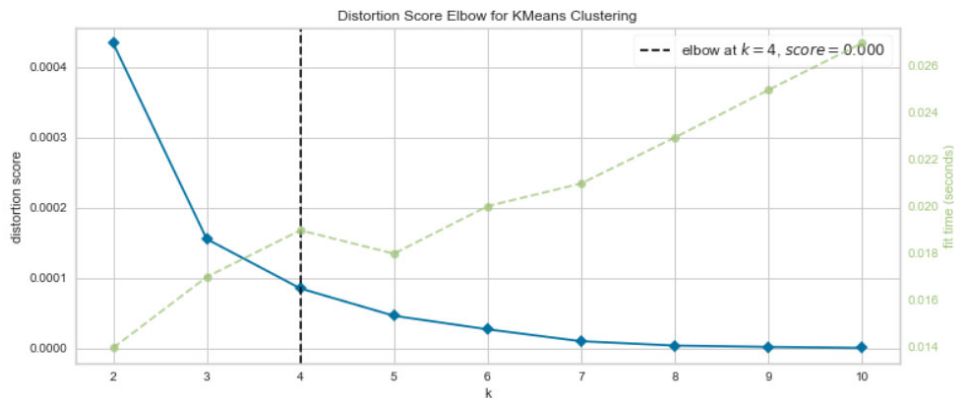|	Neighbourhood	Venue Category_Accessories Store	Venue Category_Adult Boutique	Venue Category_Airport	Venue Category_Airport Food Court	Venue Category_Airport Gate	Venue Category_Airport Lounge	Venue Category_Airport Service	Venue Category_Airport Termina
0	Parkwoods	0	0	0	0	0	0	0	
1	Parkwoods	0	0	0	0	0	0	0	
2	Victoria Village	0	0	0	0	0	0	0	
3	Victoria Village	0	0	0	0	0	0	0	
4	Victoria Village	0	0	0	0	0	0	0	

5 rows × 276 columns
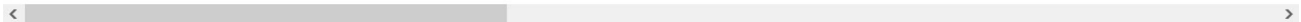
Group data by Neighbour, find the mean, and select yoga category. This data would be input to KMeans clustering.

	Venue Category_Yoga Studio
0	0.015873
1	0.025000
2	0.023810
3	0.058824
4	0.032258
5	0.020833
6	0.027027
7	0.010101
8	0.027027
9	0.023256
10	0.050000
11	0.030303

Implement the elbow method to find the optimal value of k. The result is shown in Exhibit 2, k=4.

Exhibit 2 Elbow plot

Then, fit KMeans clustering with k=4. Some of the results are shown below.

	Neighbourhood	Venue Category_Yoga Studio	cluster
0	Central Bay Street	0.015873	3
1	Church and Wellesley	0.025000	2
2	Little Portugal, Trinity	0.023810	2
3	North Toronto West, Lawrence Park	0.058824	1
4	Queen's Park, Ontario Provincial Government	0.032258	0

Results

Print out lists of the neighborhood in each cluster, as shown in Exhibit 3.

**Exhibit 3** Lists of neighbour in each cluster.

```
Neighbourhood 0
        Queen's Park, Ontario Provincial Government
        Runnymede, Swansea
        Studio District
        University of Toronto, Harbord
=================================================
Neighbourhood 1
        North Toronto West, Lawrence Park
        Thorncliffe Park
=================================================
Neighbourhood 2
        Church and Wellesley
        Little Portugal, Trinity
        Regent Park, Harbourfront
        The Danforth West, Riverdale
=================================================
Neighbourhood 3
        Central Bay Street
        Stn A PO Boxes
=================================================
```

Generate a geographical plot that shows the locations of each cluster, as shown in Exhibit 4.

Exhibit 4 Geographical plot of clusters.
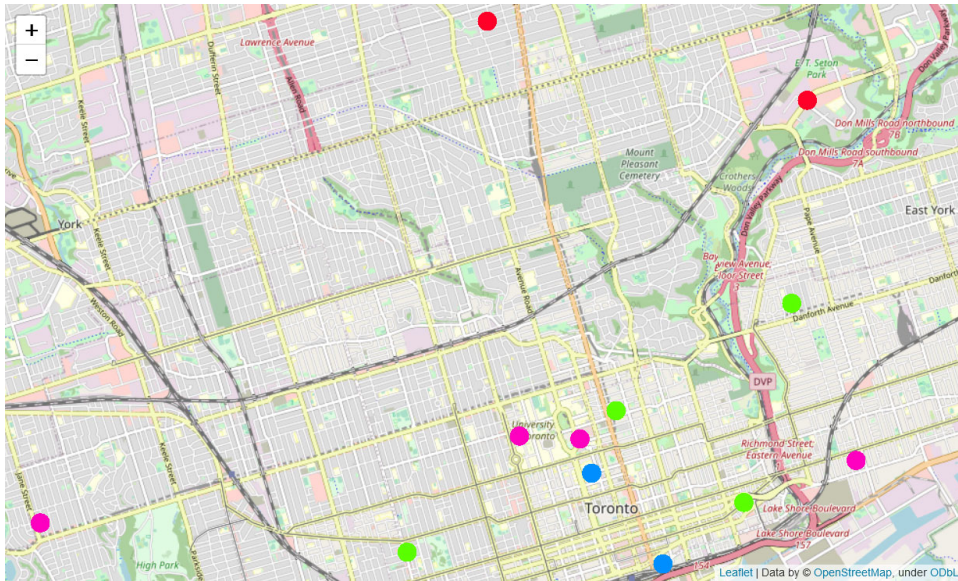
Conclusion

As mentioned earlier, it would be too risky to open a yoga studio in the neighborhood with no yoga studio because there might be no demand in the first place; that is why there are none of the yoga studios in that particular area. Everybody has to eat and more likely to try new food, but not everybody is willing to practice yoga or even think of trying it. The logic of opening a restaurant is not the same as a yoga studio. Our passionate yoga should consider opening a new yoga studio in the same cluster area with the least competitiveness—either cluster 1 in the same location as Lawrence Park and Thorncliffe Park or cluster 3 in the area around Central Bay Street and Stn A PO Boxes.

References

- List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare API: https://developer.foursquare.com/docs/