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

Business problem

where in Toronto should we open the yoga studio?
Area with no yoga studio is risky.
Area with the least competitiveness.

Data

Wikipedia URL:

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

Foursquare API

Methodology

Load data from Wikipedia and CSV file

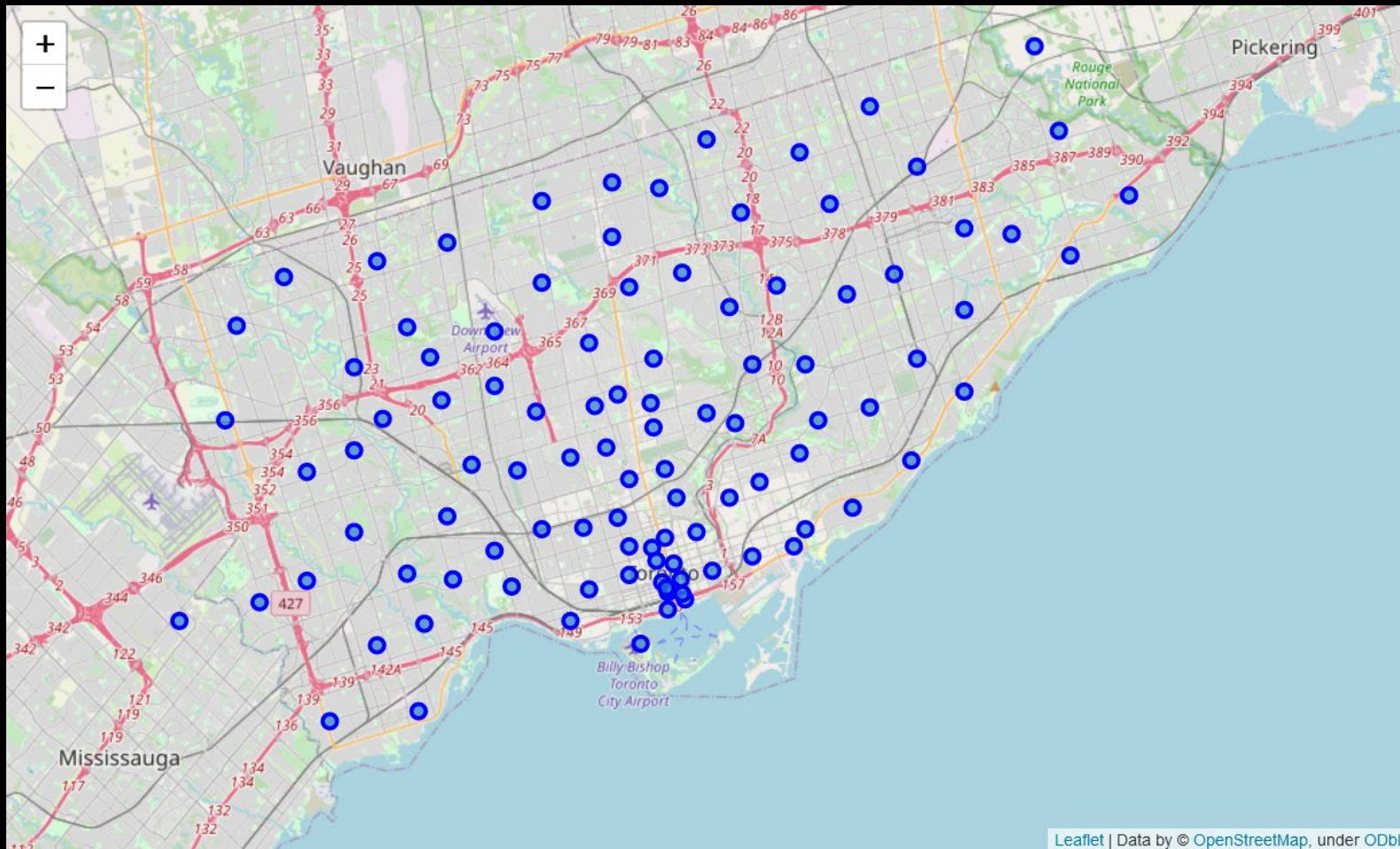|	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4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

Methodology

Load data from Wikipedia and CSV file

	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

Methodology

Geographical plot of Neighborhood in Toronto

Methodology

Retrieve Venue details via Foursquare API

	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

Methodology

Transform Venue Category into one-hot encoding

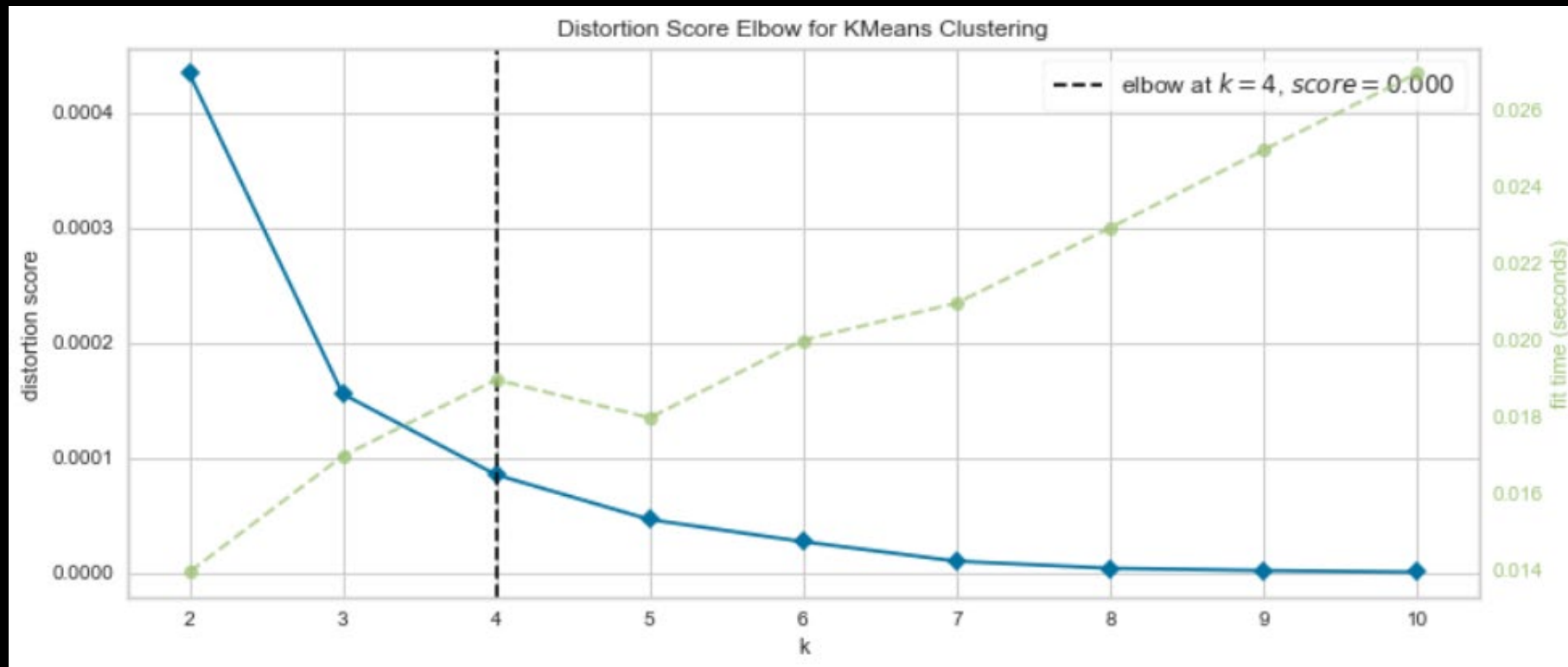|	Neighbourhood	Venue Category_Accessories Store	Venue Category_Adult Boutique	Venue Category_Airport	Venue Category_Airport Food Court	Venue Category_Airport Gate	Venue Category_Airport Lounge	Venue Category_Airport Service	Venue Category_Airport Terminal
0	Parkwoods	0	0	0	0	0	0	0	
1	Parkwoods	0	0	0	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

Methodology

Group data by Neighbor,
find the mean,
and select yoga category
This is input to KMeans.

	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

Methodology

Use elbow method,
Select optimal k=4.

Methodology

Some of outputs

	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

Lists of four clusters


```
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
==============================================================
```
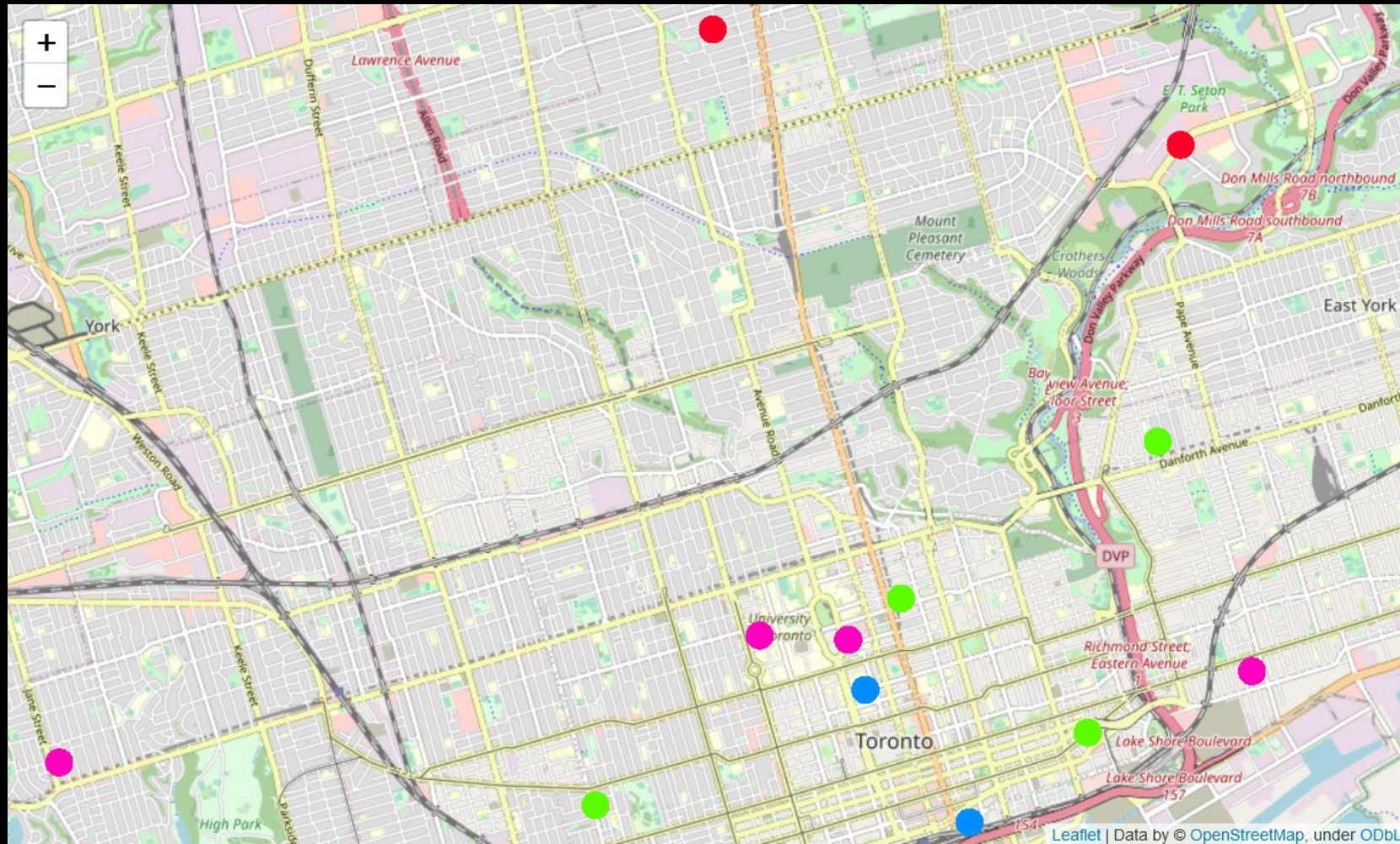
Results

Geographical plot of cluster

Conclusion

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.