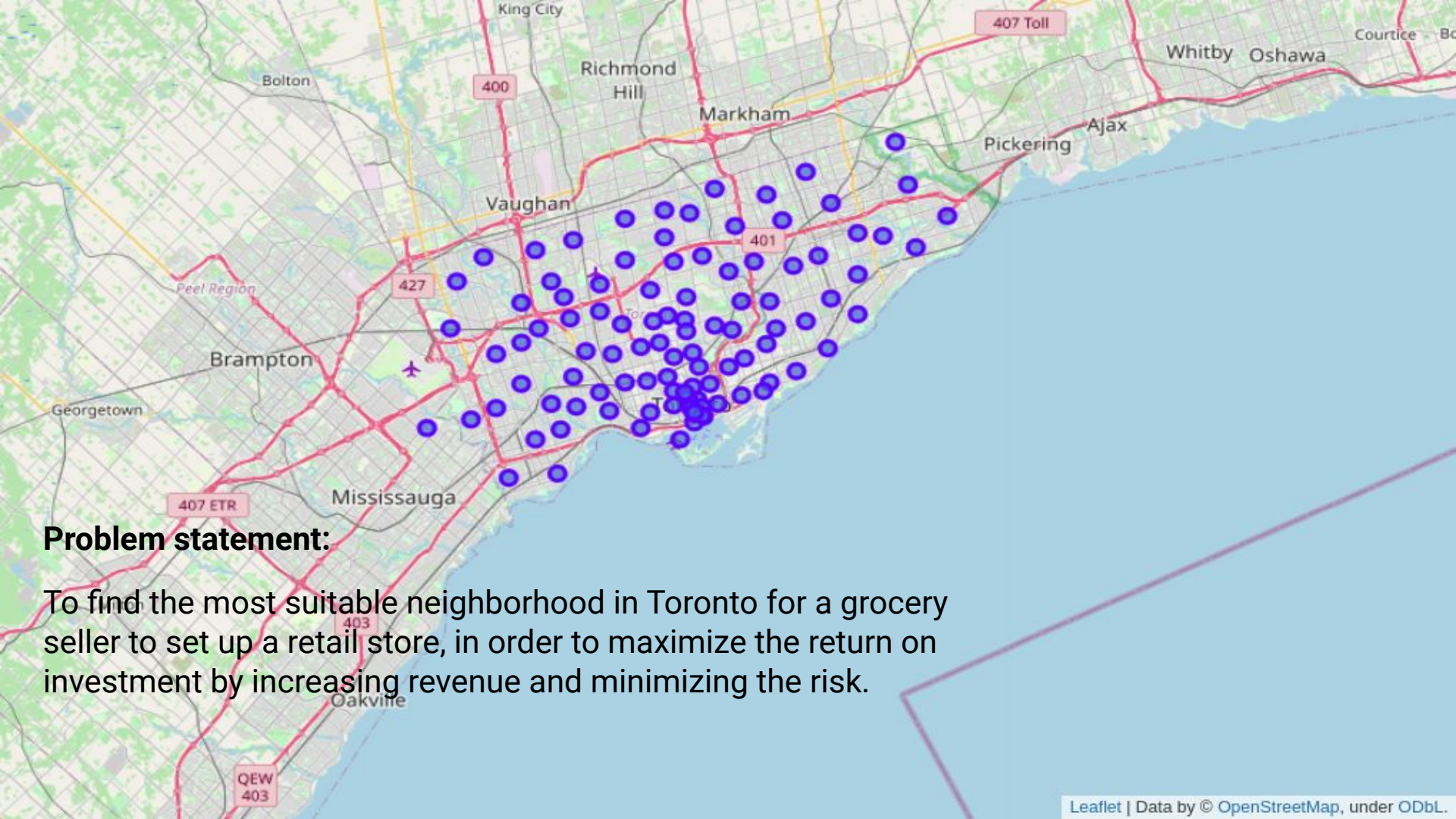# IBM Data Science Capstone

Battle of the Neighborhoods

Problem statement:

To find the most suitable neighborhood in Toronto for a grocery seller to set up a retail store, in order to maximize the return on investment by increasing revenue and minimizing the risk.

Data Description

Features:

1. Neighborhood
2. Population
3. Land Area
4. Density
5. Population %
6. Income

7. Commuting
8. 2nd Language
9. 2nd Language %
10. Latitude
11. Longitude

Dataset

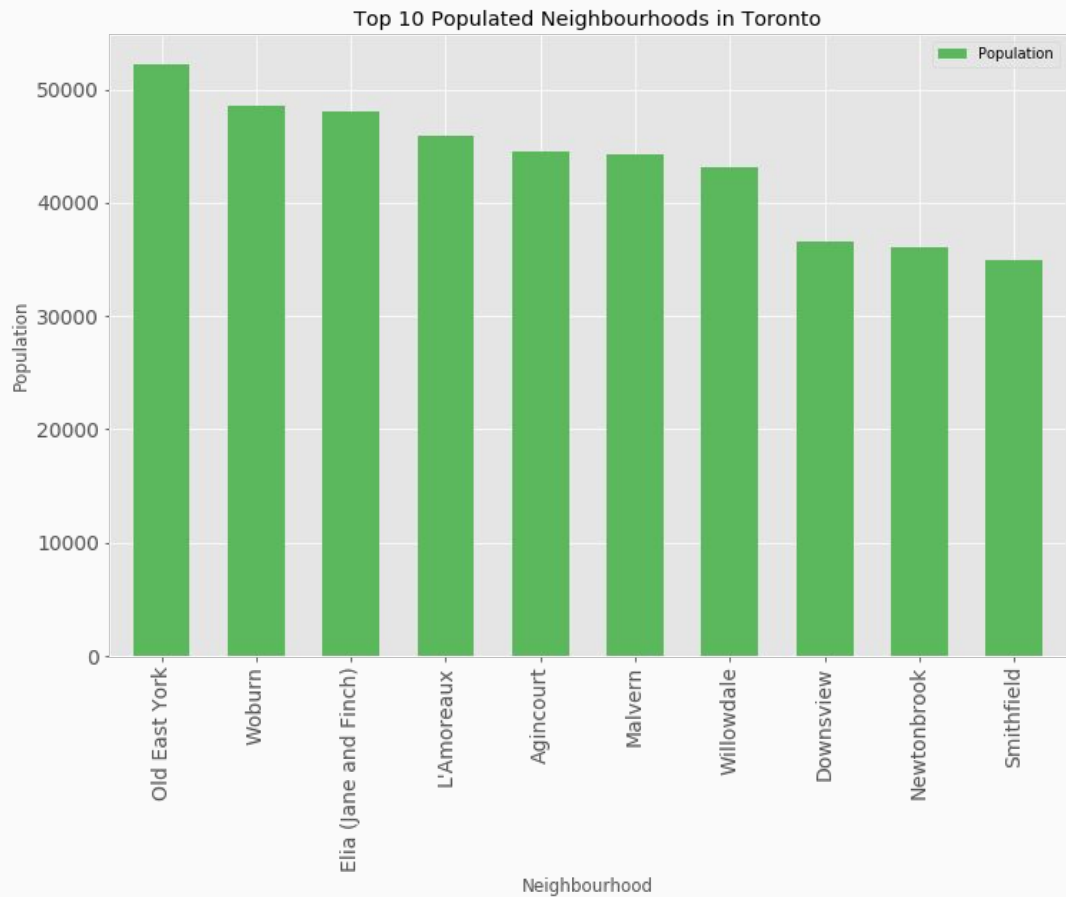|	Neighbourhood	Population	Land Area	Density	Population %	Income	Commuting	2nd Language	2nd Language %	Latitude	Longitude
1	Agincourt	44577	12.45	3580	4.6	25,750	11.1	Cantonese (19.3%)	19.3% Cantonese	43.788	-79.2839
2	Alderwood	11656	4.94	2360	-4.0	35,239	8.8	Polish (6.2%)	06.2% Polish	43.6035	-79.5464
3	Alexandra Park	4355	0.32	13,609	0.0	19,687	13.8	Cantonese (17.9%)	17.9% Cantonese	43.6498	-79.4015
4	Allenby	2513	0.58	4333	-1.0	245,592	5.2	Russian (1.4%)	01.4% Russian	43.7077	-79.4127
5	Amesbury	17318	3.51	4,934	1.1	27,546	16.4	Spanish (6.1%)	06.1% Spanish	43.7011	-79.481
6	Armour Heights	4384	2.29	1914	2.0	116,651	10.8	Russian (9.4%)	09.4% Russian	43.7454	-79.4226
7	Banbury	6641	2.72	2442	5.0	92,319	6.1	Unspecified Chinese (5.1%)	05.1% Unspecified Chinese	43.7491	-79.3664
8	Bathurst Manor	14945	4.69	3187	12.3	34,169	13.4	Russian (9.5%)	09.5% Russian	43.7627	-79.4563
9	Bay Street Corridor	4787	0.11	43,518	3.0	40,598	17.1	Mandarin (9.6%)	09.6% Mandarin	43.6567	-79.3835
10	Bayview Village	12280	4.14	2,966	41.6	46,752	14.4	Cantonese (8.4%)	08.4% Cantonese	43.7782	-79.3828

Methodology

Step 1: Web-scraping data for Toronto neighborhoods from Wikipedia.

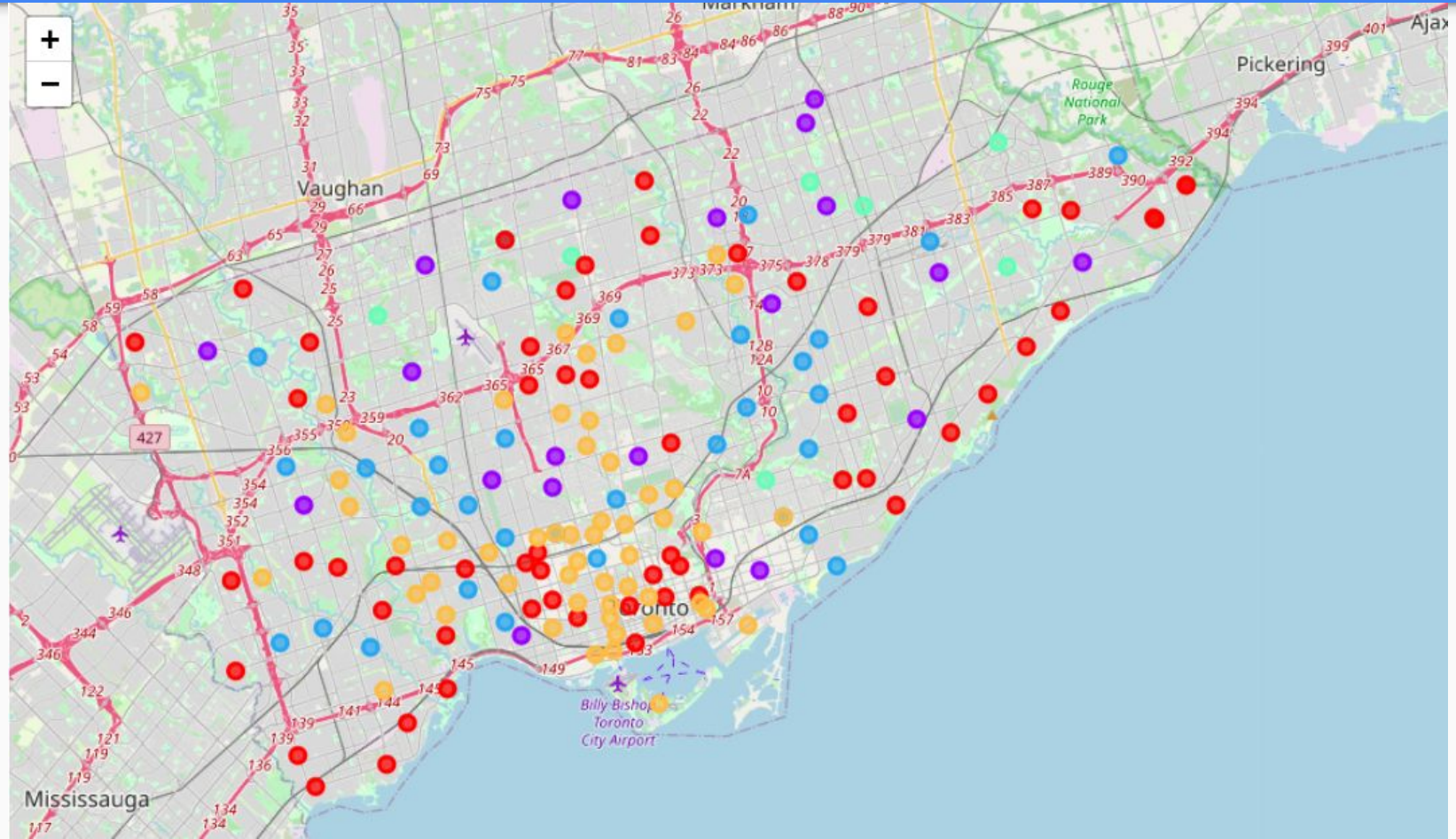Step 2: Using the geocoder API to assign the lat-lng to the neighborhoods.

Step 3: Clustering neighborhoods using k-means.

Step 4: Concluding the most suitable neighborhood for the new retail grocery store.

Neighborhood Populations

Top 10 Populated Neighbourhoods in Toronto
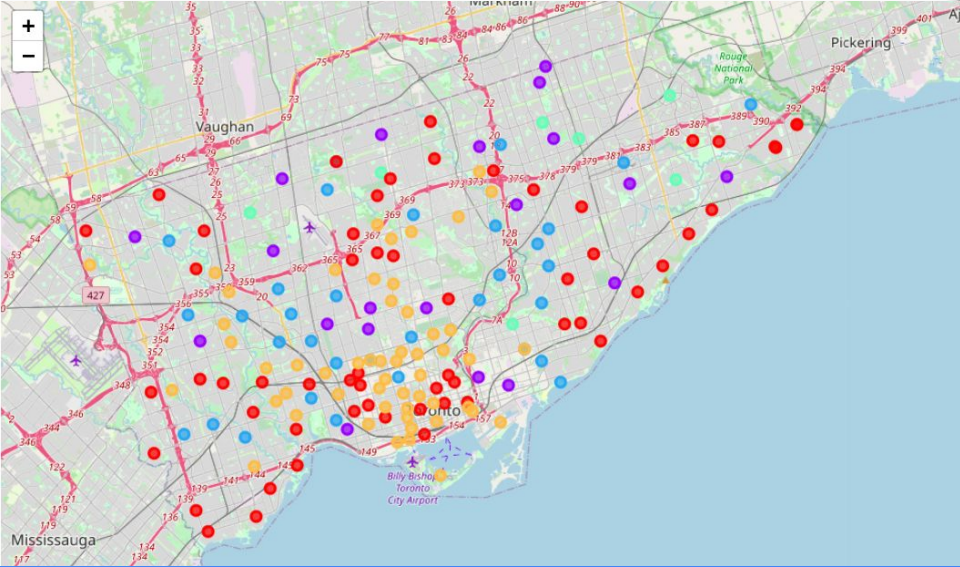
K-Means Clustering (k=5)

Common Venues

	Neighbourhood	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
0	Agincourt	Coffee Shop	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
1	Alderwood	Pizza Place	Gym	Dance Studio	Pub	Coffee Shop	Donut Shop	Bank	Convenience Store
2	Alexandra Park	Bar	Café	Vegetarian / Vegan Restaurant	Restaurant	Coffee Shop	French Restaurant	Dessert Shop	Yoga Studio
3	Allenby	Coffee Shop	Sushi Restaurant	Gym	Italian Restaurant	Café	Fruit & Vegetable Store	Liquor Store	Gastropub
4	Amesbury	Bakery	Fast Food Restaurant	Park	Sandwich Place	Flea Market	Fish Market	Fish & Chips Shop	Filipino Restaurant

Conclusion

Cluster 0 looks to offer a higher number of similar neighborhoods and allow a new retail grocery store to run successfully. The neighborhood **Humbermede** looks to be the most suitable choice as it is the highest populated neighborhood and very few grocery stores are present.

- Suitable cluster: **Cluster 0**
- Suitable Neighborhood: **Humbermede**

There is a also large Punjabi population in that neighborhood and hence, we can also suggest the sale of Punjabi food or related items.

Thank You

Humbermede 🎉

	Neighbourhood	Population	Income	Commuting	2nd Language	2nd Language %	Latitude	Longitude	Population Score	Venue Score	Total Score
60	Humbermede	14778	24,297	11.8	Punjabi (9.7%)	09.7% Punjabi	43.7421	-79.5407	0.611729	0.0	0.305865