

A night-time photograph of the Toronto skyline, featuring the CN Tower and various skyscrapers illuminated against a dark blue sky. The city lights reflect on the water in the foreground. The image is overlaid with several dark blue diagonal lines that create a geometric pattern across the frame.

# Predicting the most profitable store type in Toronto

---

Capstone  
Project

Deepti Madanagopalan

# Why is this necessary

Finding the least common store type determines the demand

Who would be interested?

- A new business trying to find the most profitable type of store in Toronto
- An existing business trying to find the best location for their store
- Residents of the neighborhood





# Types of Data Used

---

## Four Different Sources

- Data of different neighborhoods and postal codes in Toronto from Wikipedia
- Latitude and Longitude for the different postal codes as a csv file
- Boundaries of each neighborhood as a geojson file
- Store name, type and location from Foursquare

# Data Cleaning- Toronto Location Data's

The postal code data had some missing and repeated values and needed correction. Also, the location csv data was combined with the postal code data

## Original Data

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
7	M8A	Not assigned	Not assigned
8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
9	M1B	Scarborough	Malvern, Rouge

## Cleaned Data

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686299999996	-79.19435340000001
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.7845351	-79.16049709999999
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.7635726	-79.1887115
3	M1G	Scarborough	Woburn	43.7709921	-79.21691740000001
4	M1H	Scarborough	Cedarbrae	43.773136	-79.23947609999999



# Data Cleaning- Foursquare Store Data

The data from foursquare consists of all venues nearby. We only need nearby store data. This store data was cleaned further to get numeric values for analysis.

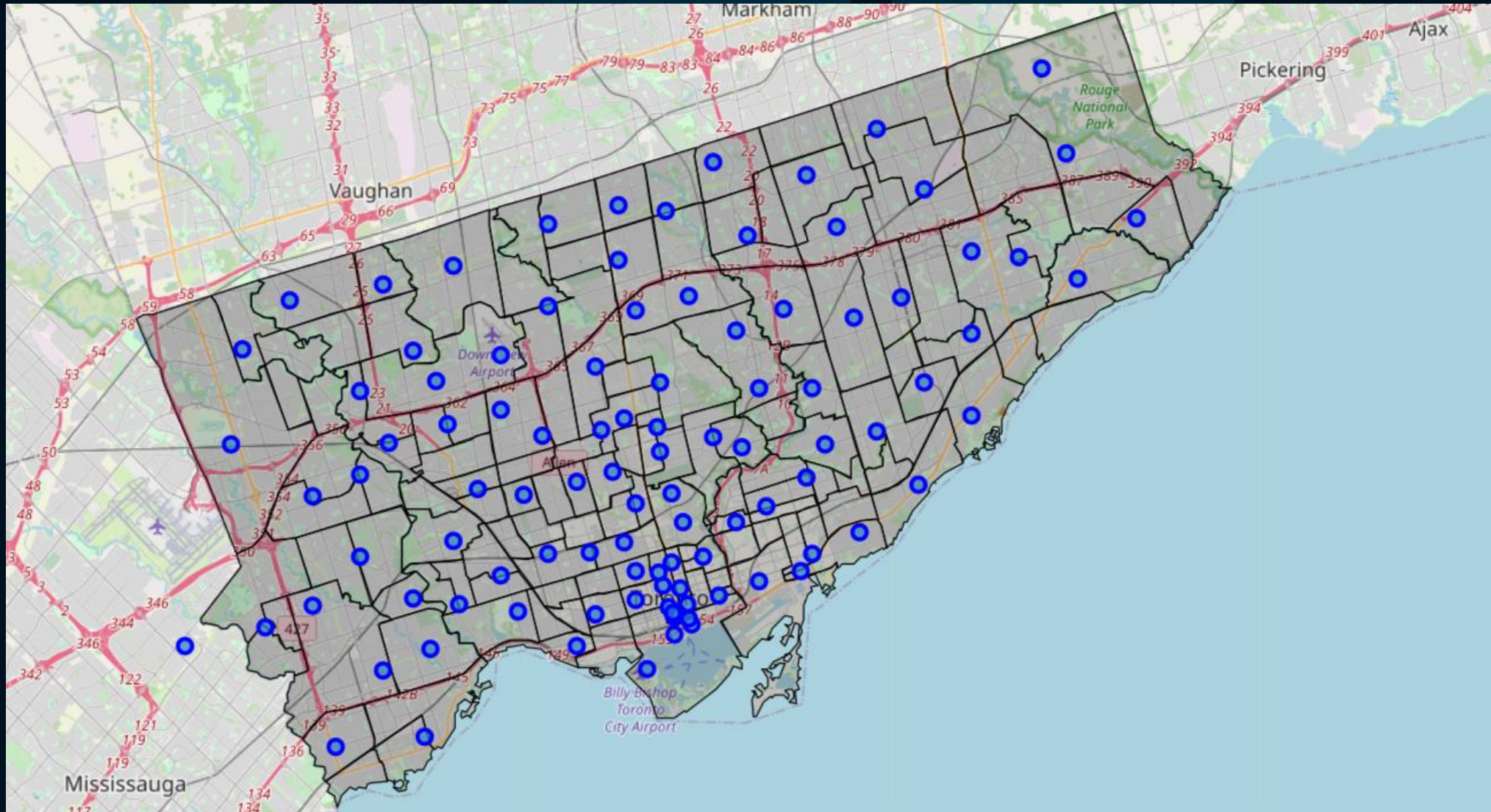
## Original Data

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686299999996	-79.19435340000001	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.7845351	-79.16049709999999	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.7635726	-79.1887115	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.7635726	-79.1887115	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.7635726	-79.1887115	Sail Sushi	43.765951	-79.191275	Restaurant

## Cleaned Data

	Neighbourhood	Accessories Store	Arts & Crafts Store	Baby Store	Beer Store	Camera Store	Candy Store	Clothing Store	Convenience Store	Department Store	...	Music Store	Pet Store	Shoe Store	Stationery Store	Thrift / Vintage Store	Toy / Game Store	Video Game Store	Video Store	Warehouse Store	Women's Store
0	Bedford Park, Lawrence Manor East	0.0	0.0	0.000000	0.0	0.0	0.000000	0.00	0.00	0.00	...	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0
1	Berczy Park	0.0	0.0	0.000000	0.0	0.0	0.000000	0.25	0.00	0.25	...	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0
2	Brockton, Parkdale Village, Exhibition Place	0.0	0.0	0.000000	0.0	0.0	0.000000	0.00	0.25	0.00	...	0.0	0.25	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0
3	Caledonia-Fairbanks	0.0	0.0	0.000000	0.0	0.0	0.000000	0.00	0.00	0.00	...	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.0	1.0
4	Central Bay Street	0.0	0.0	0.000000	0.0	0.0	0.000000	0.00	0.00	0.25	...	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.25	0.0	0.0
5	Christie	0.0	0.0	0.166667	0.0	0.0	0.166667	0.00	0.00	0.00	...	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0
6	Church and Wellesley	0.0	0.0	0.000000	0.0	0.0	0.000000	0.20	0.00	0.00	...	0.0	0.00	0.2	0.0	0.0	0.0	0.0	0.00	0.0	0.0

# Neighborhoods of Toronto





# Type of Analysis

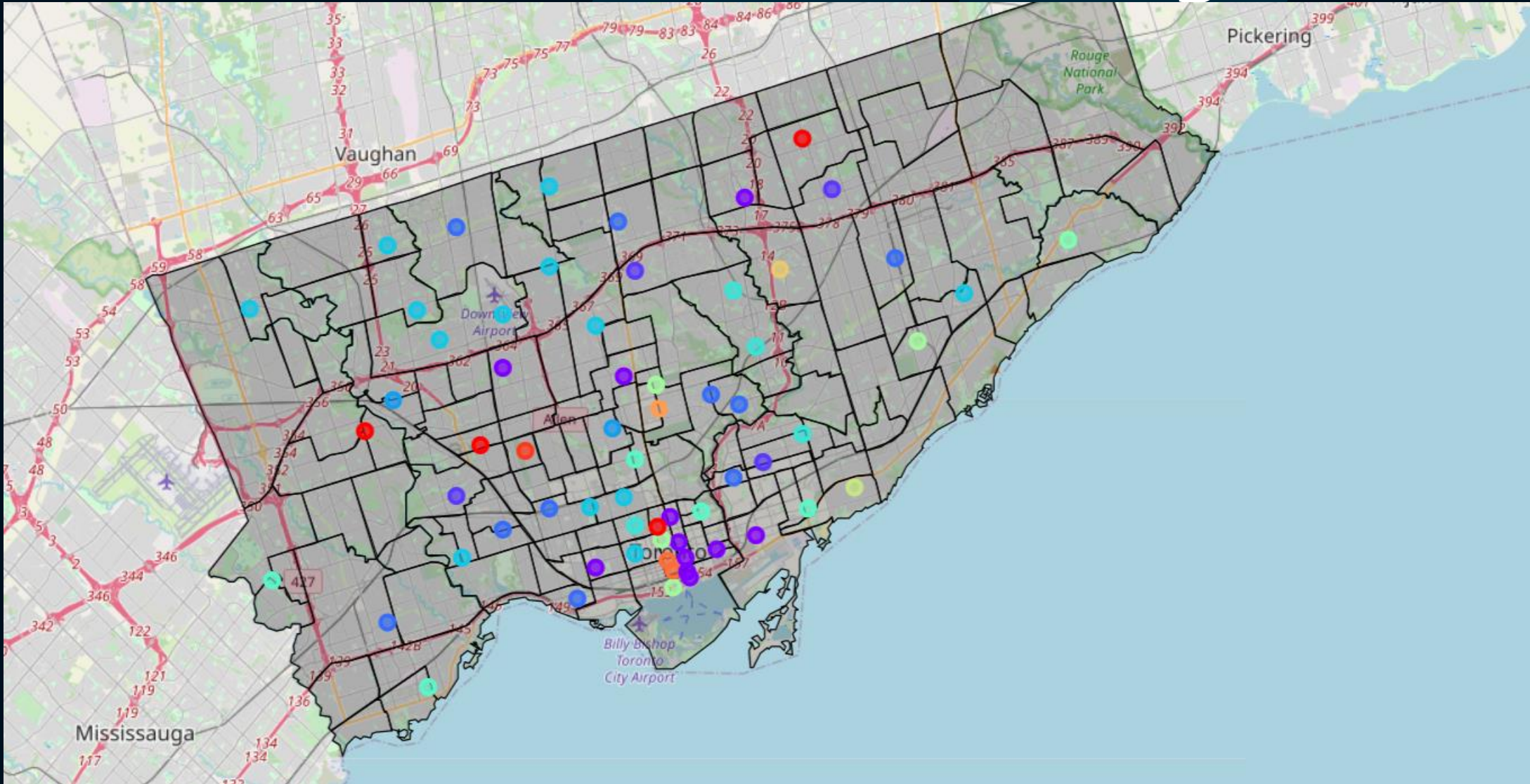
---

## Unsupervised Machine Learning

- The stores need to be categorized based on similarity
- Using K-means clustering, we can get similar groups
- The number of groups was chosen to 15
- After sorting into groups, we can visualize the result on the neighborhood map



# K-means Clustering





# Result

The least common store type in Toronto is

- A new business will be able to use this information to open a store as it has the most potential to be successful
- An existing business can find the neighborhood where their store would be most profitable





# Next Steps

---

- To increase the accuracy of prediction, the dependence of the analysis on the number of groups should be studied.
- To get a full picture of profitability, additional economic factors can be included such as the finances of the stores



**THANK  
YOU!**