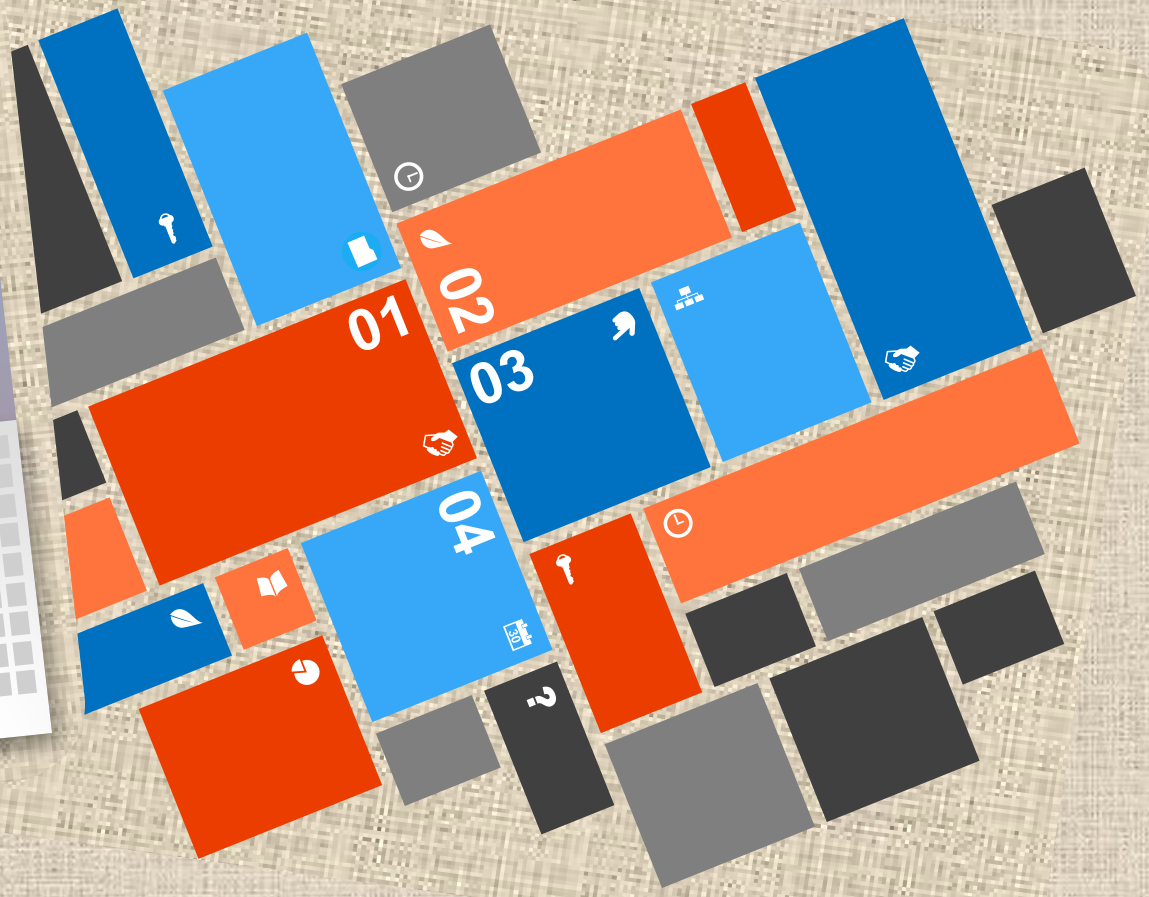


# Recommender System- Best Neighborhoods

Applied Data Science Capstone
IBM Data Science Professional Certificate

# INDEX:

# INTRODUCTION

## Business Problem

An individual migrating to Canada, specifically to Toronto, needs information on the potential neighborhoods in and around Toronto where he can scout for a decent apartment.

*Is it possible for an Individual to get a detailed information about possible neighborhoods in and around Toronto and also the amenities that are available in those neighborhoods even before starting his Journey?*

A prior hand information about various neighborhoods, and facilities available in and around those neighborhoods will be of tremendous value and can help the person to take more informed decision by zeroing in on the neighborhoods that meet his specific requirements. Since the Individual can get all this information even before being physically present there, he can save time and resources that would otherwise have been spent scanning neighborhoods and the facilities they offer.

# INTRODUCTION

## *Potential Solution*

The Recommender System is a potential solution *targeted at those potential Immigrants who plan to migrate and settle abroad (Toronto in this case). This System can be adapted to any province or for that matter any country in the world as long as we have Geo-locational data available.* In this case we have incorporated data pertaining to Utilities, Hospitals, Educational Institutions, Restaurants etc. to suggest the best neighborhood, but this system is scalable and can be changed as per specific needs. The system will cluster neighborhoods based on available amenities and suggest the best cluster. This is better, since individual is provided with information about multiple neighborhoods which are similar in terms of available amenities. He can knit pick what's important for him and choose the location that suits him the best.

***The Recommender System*** will be of tremendous help to any potential immigrant planning to migrate to any corner of the world, as he would have a first hand detailed information about the place he is going even before being there.

# DATA

To build the recommender system:

- We need is the data containing the geographical coordinates of the boroughs and the neighborhoods within them.
- We have selected data pertaining to Toronto and it's neighborhood, specifically selected only East Toronto, West Toronto, Central Toronto and Downtown Toronto.
- We identified the neighborhoods within these boroughs using postal codes* and got the geo - coordinates for them. We used a geospatial data which contains the geographical coordinates of these postal codes.


* We got the postal code information about Toronto and it's neighborhoods from a Wikipedia page 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'.

# DATA

**Example of the Neighborhood data got from the Wikepedia page**

| Postcode | Borough | Neighborhood |
|---|---|---|
| M4E | East Toronto | The Beaches |

**Example of the Geo spatial data**

| Postcode | Latitude | Longitude |
|---|---|---|
| M4E | 43.676357 | -79.293031 |

**Example from a Four Square Explore API, converted to a dataFrame**

| Neighbourhood | Neighbourhood Latitude | Neighbhourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| The Beaches | 43.676357 | -79.293031 | Beaches Bake shop | 43.680363 | -79.289692 | Bakery |

# DATA

- We used the Four Square APIs to explore the neighborhoods of Toronto. A typical request from Foursquare will provide us with the following information:

| | Postal Code | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Summary | Venue Category | Distance |
|---|---|---|---|---|---|---|---|---|
| 0 | M1W | Steeles West | 43.799525 | -79.318389 | Mr Congee Chinese Cuisine 龍粥記 | This spot is popular | Chinese Restaurant | 72 |
| 1 | M1W | Steeles West | 43.799525 | -79.318389 | Agincourt Bakery | This spot is popular | Bakery | 759 |
| 2 | M1W | Steeles West | 43.799525 | -79.318389 | Little Sheep Mongolian Hot Pot 小肥羊 | This spot is popular | Hotpot Restaurant | 972 |
| 3 | M1W | Steeles West | 43.799525 | -79.318389 | Phoenix Restaurant 金凰餐廳 | This spot is popular | Chinese Restaurant | 147 |
| 4 | M1W | Steeles West | 43.799525 | -79.318389 | Price Chopper | This spot is popular | Grocery Store | 16 |

# METHODOLOGY

To suggest the best neighbourhood in Toronto, first we need to identify the neighbourhoods in Toronto.

I. We scraped the Wikipedia page and wrangled the data, cleaned it, and then read it into a panda's data frame
II. We then acquired the latitudes and longitudes for these neighbourhoods in Toronto from the https://cocl.us/Geospatial_data.
III. Once we have the list of neighbourhoods and their geographical coordinates, the next step in the analysis was to obtain amenities in each neighbourhoods. For that we used Four Square api's to gather information about the venues in each neighbourhoods.
IV. Once we had all the required data, we did a one-hot encoding of the categories column and summarised it based on the neighbourhoods. Now our data is ready for Segmentation.

# METHODOLOGY

V. We used k-means clustering to segment our neighbourhoods into 5 clusters. We then added the cluster labels back to the data.

VI. After performing the clustering, we then focus on the cluster centers. The Group whose center has the highest amenities (from Total Facilities) will be the suggested for “the Best neighbourhood”

# RESULTS

| Groups | Arts and Entertainment | Athletics and Sports | Educational | Food and Dine | Gym and Fitness Center | Health and Beauty Service | Hospitals | Others | Recreation | Stores and Utilities | Transportation | Facilities |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C3 | 8.272727 | 0.727273 | 48.636364 | 68.090909 | 2.545455 | 0.909091 | 4.64E+00 | 3.727273 | 2.545455 | 12.272727 | 0.454545 | 152.818182 |
| C1 | 5.090909 | 1 | 21.545455 | 70.818182 | 3 | 1.181818 | 4.44E-16 | 2.454545 | 2.727273 | 13.272727 | 0.181818 | 121.272727 |
| C4 | 4 | 1.666667 | 12.5 | 53.333333 | 2.5 | 1.166667 | 3.33E-01 | 2.166667 | 7.666667 | 9.666667 | 1.333333 | 96.333333 |
| C2 | 1.5 | 1.833333 | 15.5 | 30.333333 | 2.333333 | 0.666667 | 1.67E-01 | 1.666667 | 6 | 6.666667 | 0 | 66.666667 |
| C5 | 0.25 | 0.75 | 9.75 | 7.5 | 0.5 | 0.5 | 5.00E-01 | 1.5 | 2.75 | 2.75 | 0.75 | 27.5 |

We can see that Cluster 3 (C3) has emerged as the best cluster with a set of Neighbourhoods that score high in terms of amenities, followed by C1, C4,C2 & C5.

Clusters are color coded in the attached map.

# RESULTS

| Neighborhood | Postcode | Borough | Latitude | Longitude | Arts and Entertainment | Athletics and Sports | Educational | Food and Dine | Gym and Fitness Center | Health and Beauty Service | Hospitals | Others | Recreation | Stores and Utilities | Transportation | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Church and Wellesley | M4Y | Downtown Toronto | 43.66586 | -79.38316 | 9 | 0 | 50 | 70 | 2 | 2 | 6 | 4 | 3 | 10 | 0 | 3 |
| Ryerson,Garden District | M5B | Downtown Toronto | 43.657162 | -79.378937 | 6 | 0 | 50 | 65 | 2 | 1 | 7 | 3 | 4 | 19 | 0 | 3 |
| St. James Town | M5C | Downtown Toronto | 43.651494 | -79.375418 | 8 | 0 | 50 | 65 | 3 | 0 | 2 | 7 | 3 | 14 | 0 | 3 |
| Central Bay Street | M5G | Downtown Toronto | 43.657952 | -79.387383 | 9 | 0 | 51 | 63 | 2 | 2 | 5 | 4 | 3 | 16 | 0 | 3 |
| Adelaide,King,Richmond | M5H | Downtown Toronto | 43.650571 | -79.384568 | 13 | 0 | 51 | 60 | 4 | 2 | 6 | 4 | 1 | 14 | 1 | 3 |
| Design Exchange,Toronto Dominion Centre | M5K | Downtown Toronto | 43.647177 | -79.381576 | 8 | 2 | 48 | 71 | 3 | 1 | 3 | 2 | 6 | 6 | 1 | 3 |
| Commerce Court,Victoria Hotel | M5L | Downtown Toronto | 43.648198 | -79.379817 | 7 | 2 | 49 | 70 | 3 | 0 | 3 | 5 | 3 | 9 | 1 | 3 |
| Harbord,University of Toronto | M5S | Downtown Toronto | 43.662696 | -79.400049 | 8 | 0 | 49 | 71 | 0 | 0 | 6 | 3 | 2 | 13 | 0 | 3 |
| Chinatown,Grange Park,Kensington Market | M5T | Downtown Toronto | 43.653206 | -79.400049 | 8 | 0 | 44 | 74 | 3 | 0 | 5 | 2 | 0 | 13 | 0 | 3 |
| Stn A PO Boxes 25 The Esplanade | M5W | Downtown Toronto | 43.646435 | -79.374846 | 6 | 2 | 44 | 69 | 3 | 0 | 3 | 4 | 2 | 13 | 1 | 3 |
| First Canadian Place,Underground city | M5X | Downtown Toronto | 43.648429 | -79.38228 | 9 | 2 | 49 | 71 | 3 | 2 | 5 | 3 | 1 | 8 | 1 | 3 |

# CONCLUSION

Recommender System is an effective and reliable tool in this Project of identifying best neighbourhoods in Toronto city.

Eventhough we have taken Toronto as an example in this case study, but the Recommender system can be implemented across geographies as long as we have the required Geo-locational data available. The locational data provided by the locational data providers like Four Square, Google Places etc. can help us fine tune and adapt the system to a granular level of information that we need, to cater to our specific requirements.

# THANK YOU!!