IBM Data Science

Battle of Neighborhoods

1. Introduction

1.1 Background

During the IBM data science training, we explored locations and venues. We used foursquare APIs to get venue details by passing latitude and longitude. Using Manhattan Data and Toronto Data data, we would like to solve the following problems.

1.2 Problems

Matching neighborhoods

How can we compare Manhattan and Toronto neighborhoods to find out if they are matching or not?

Find location to open a restaurant

How do we decide where to open a new Thai restaurant?

2. Data

2.1 Data Sources

We will be using Manhattan Data and Toronto Data neighborhoods data to start with and later explore it further using foursquare venue APIs. Data look like below.

Manhattan Neighborhoods with top 10 most common venues

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	9th Most Common Venue	10th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	0	Discount Store	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Tennis Stadium	Department Store	Pizza Place	Diner	Pharmacy
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Noodle House	Shanghai Restaurant	Salon / Barbershop	Dessert Shop	Optical Shop	Spa
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Grocery Store	Bank	Sandwich Place	Coffee Shop	Park	Spanish Restaurant	Deli / Bodega	New American Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	4	Mexican Restaurant	Café	Lounge	Restaurant	Park	Chinese Restaurant	Frozen Yogurt Shop	Bakery	Caribbean Restaurant	Pizza Place
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Cocktail Bar	Indian Restaurant	Liquor Store	Sushi Restaurant	Park	Deli / Bodega

Toronto Neighborhoods with top 10 most common venues

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	9th Most Common Venue	10th Most Common Venue
0	North York	Parkwoods	43.753259	-79.329656	0.0	Food & Drink Shop	Park	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Women's Store
1	North York	Victoria Village	43.725882	-79.315572	1.0	Portuguese Restaurant	Pizza Place	French Restaurant	Coffee Shop	Hockey Arena	Intersection	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store
2	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.0	Coffee Shop	Park	Pub	Bakery	Breakfast Spot	Café	Theater	Hotel	Chocolate Shop	Spa
3	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Women's Store	Coffee Shop	Boutique	Furniture / Home Store	Gift Shop	Event Space	Accessories Store	Vietnamese Restaurant	Convenience Store
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	1.0	Coffee Shop	Yoga Studio	Bank	Beer Bar	Smoothie Shop	Sandwich Place	Restaurant	Café	Portuguese Restaurant	Chinese Restaurant

2.2 Data cleaning and improvement

Data downloaded from multiple sources like Wikipedia and Foursquare APIs are combined together. Incomplete or not available data was removed. Data is later processed to calculate top 10 most common venues. To plot the map, only Thai restaurant venues are kept to find out places where Thai restaurants are not available.

3. Exploratory data analysis

3.1 Matching neighborhoods

Neighborhoods are matched by comparing top 10 common venues of Manhattan and Toronto neighborhoods. Neighborhoods are considered similar only if 2 out of top 3 venues match. After comparing this data, output look like below:

	Manhattan Neighborhood	Toronto Neighborhood	Common Venue Categories
0	Marble Hill	Bedford Park, Lawrence Manor East	Sandwich Place; Coffee Shop
1	Washington Heights	Christie	Café; Grocery Store
2	Hamilton Heights	Garden District, Ryerson	Coffee Shop; Café
3	Manhattanville	Stn A PO Boxes	Coffee Shop; Seafood Restaurant
4	East Harlem	High Park, The Junction South	Mexican Restaurant; Thai Restaurant
5	Upper East Side	Central Bay Street	Italian Restaurant; Coffee Shop
6	Yorkville	Don Mills	Coffee Shop; Gym
7	Lenox Hill	Central Bay Street	Italian Restaurant; Coffee Shop
8	Upper West Side	Central Bay Street	Italian Restaurant; Café
9	Midtown	Toronto Dominion Centre, Design Exchange	Hotel; Coffee Shop
10	Murray Hill	Toronto Dominion Centre, Design Exchange	Coffee Shop; Hotel
11	Chelsea	Studio District	Coffee Shop; Bakery
12	East Village	High Park, The Junction South	Bar; Mexican Restaurant
13	Lower East Side	Bayview Village	Chinese Restaurant; Japanese Restaurant
14	Little Italy	University of Toronto, Harbord	Café; Bakery
15	Soho	Lawrence Manor, Lawrence Heights	Clothing Store; Coffee Shop
16	Manhattan Valley	Queen's Park, Ontario Provincial Government	Coffee Shop; Yoga Studio
17	Morningside Heights	Regent Park, Harbourfront	Park; Coffee Shop
18	Gramercy	Central Bay Street	Italian Restaurant; Coffee Shop
19	Battery Park City	Canada Post Gateway Processing Centre	Hotel; Gym

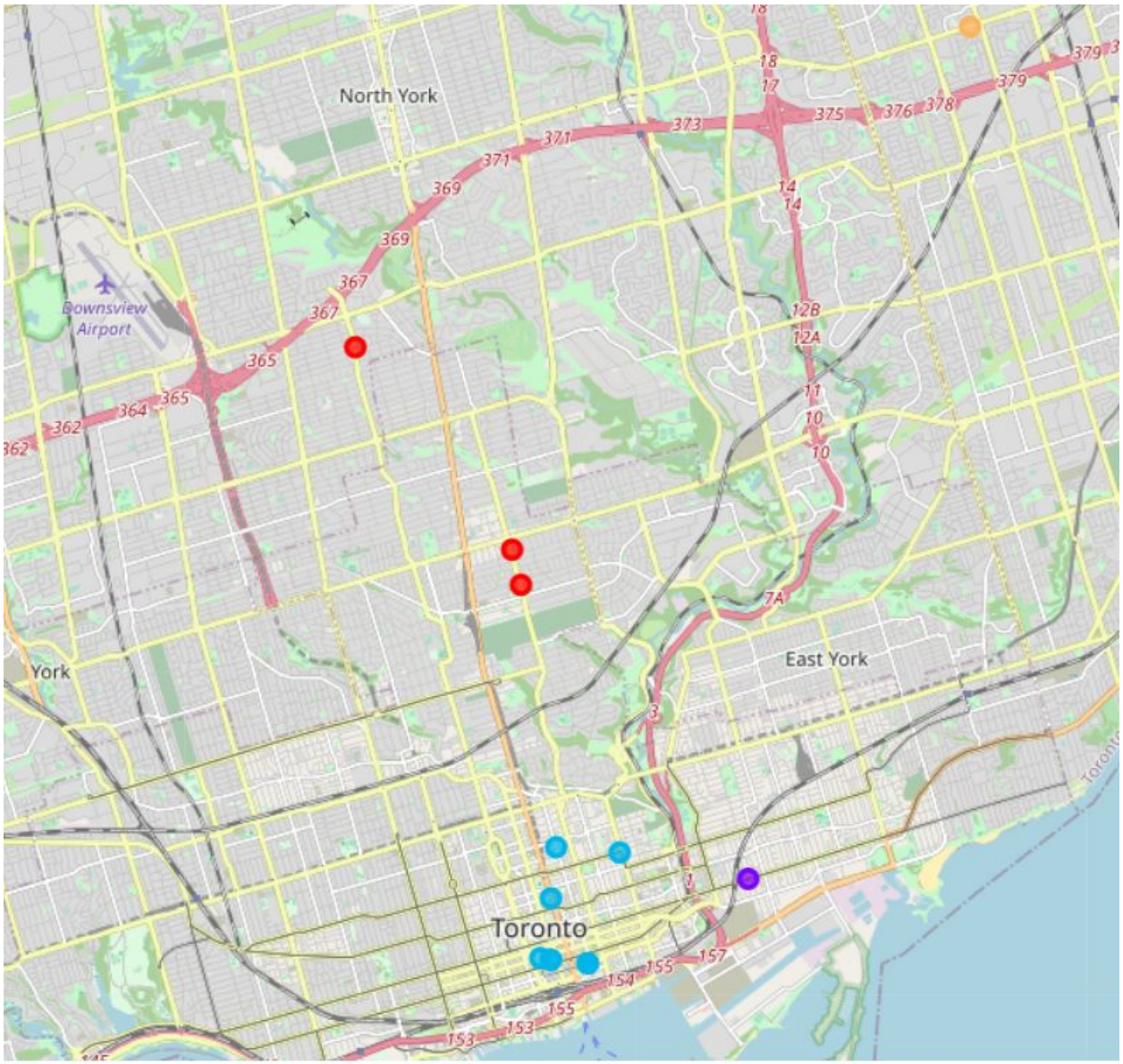
3.2 Finding location for new Thai restaurant

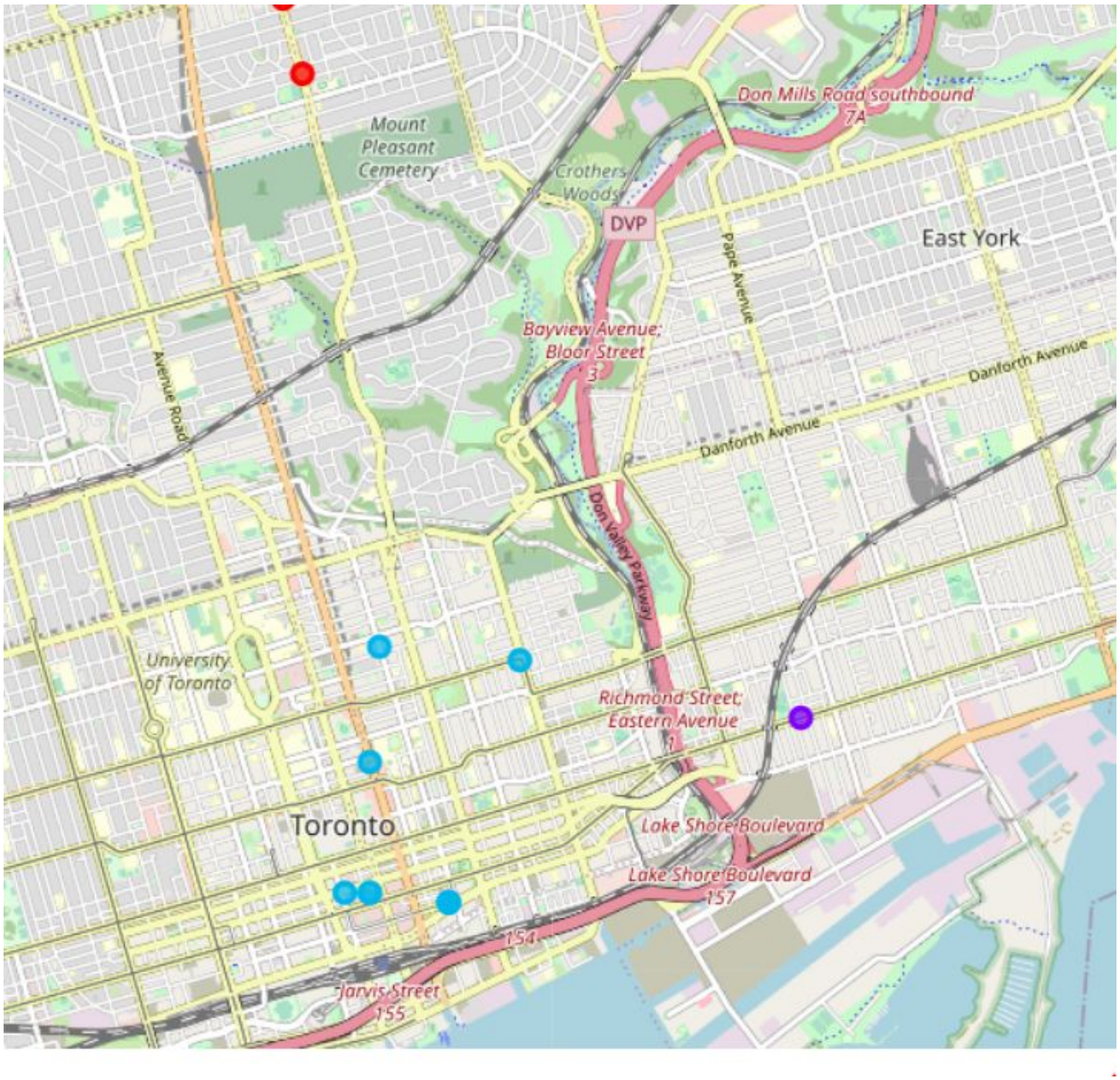
To find out a new location for Thai restaurant, initial data analysis is done to find out places where Thai restaurants are not in the top 10 most common venues.

Below image shows a list of neighborhoods where Thai restaurants are not in top 10 most common venues.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
0	North York	Parkwoods	43.753259	-79.329656	0.0	Food & Drink Shop	Park	Drugstore	Diner	Discount Store	Distribution Center	Dog Run
1	North York	Victoria Village	43.725882	-79.315572	1.0	Portuguese Restaurant	Pizza Place	French Restaurant	Coffee Shop	Hockey Arena	Intersection	Doner Restaurant
2	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.0	Coffee Shop	Park	Pub	Bakery	Breakfast Spot	Café	Theater
3	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Women's Store	Coffee Shop	Boutique	Furniture / Home Store	Gift Shop	Event Space
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	1.0	Coffee Shop	Yoga Studio	Bank	Beer Bar	Smoothie Shop	Sandwich Place	Restaurant
5	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	Scarborough	Malvern, Rouge	43.806686	-79.194353	4.0	Fast Food Restaurant	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
7	North York	Don Mills	43.745906	-79.352188	1.0	Gym	Beer Store	Coffee Shop	Japanese Restaurant	Art Gallery	Sporting Goods Shop	Bike Shop
8	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	1.0	Pizza Place	Pet Store	Athletics & Sports	Gastropub	Intersection	Pharmacy	Bus Line

After finding out neighborhoods, the next step is to plot graph of Thai restaurants as circle maker.

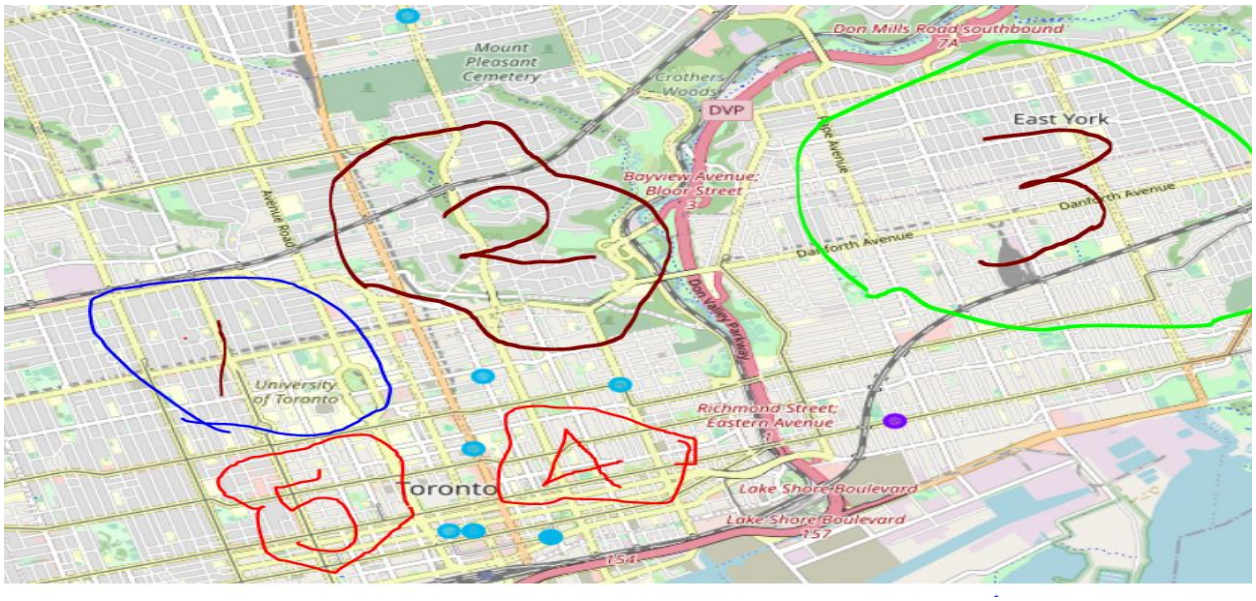

North York
Downsview Airport
York
East York
Toronto
379
378
376
375
373
371
369
367
365
364
362
18
17
14
12B
12A
11
10
7A
3
1
157
155
154
153



As you can see above, Thai restaurants are highlighted in different colored circles as these are divided into clusters using K-means.

4. Results

Plot shows locations with Thai restaurants. Areas marked in numbers 1 to 5 could be potential areas to open a new restaurant.



5. Conclusion

By working on a project, I understood how to compare two dataframes to effectively compare two neighborhoods.I understood the importance of location data extracted using foursquare APIs help us in analysing existing data. Also, plotting a graph helps us to visualize location data and to make important decisions easily.

There are several areas where Thai restaurants are not present so lots of options available to open a new restaurant. Numbers 1, 2, 3 mentioned in the above graph are good options as there are no nearby Thai restaurants. Option 4 and 5 could be good options but higher competition as there are more nearby Thai restaurants.

6. Future directions

There are few points not captured by the program which are:

1. Customer base based on location
2. Cost associated based on location
3. Owner preference

Customer base is a very important factor as if we open a restaurant in an area where people prefer more Thai food, it could turn things very easily.

Cost like Infrastructure cost, rental, transportation, taxes based on location of restaurant could become decision making factors. So it all depends on how much the owner would like to spend and then choose the location wisely.

Owner preference is also an important factor, so after providing the options, the owner will have to choose a location which could be more suitable to him/her. But initial analysis done could help the owner to choose the option.