CAPSTONE PROJECT-The Battle of Neighborhoods-Part 2

Applied Data Science By IBM-Coursera

Table of contents

Introduction : Business Problem

In this project, the goal is to find the right neighborhood corresponding to future or actual citizens of Toronto who wants to rent or buy a home regarding special criterias(like the number of restaurants around their house, are cinemas or theatres near by, or in opposite, citizens that don't want to have restaurants or distraction centers near by).

Since citizens have their particular preferences regarding the choice of their houses, we will then use data Science to find and match the right Toronto neighborhoods correspondng to the home nearby distraction insitutions criterias.

Datas

To achieve this goal, we will use:

- Datas from wikipedia at this link: "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
- Datas from Foursquare API
- Data from Geocoder or Geospatial_coordinates_csv

Methodology

1-Import Libraries

All the libraries needed for the importation, process,analysis of the datas will have to be imported first.From pandas to numpy, matplotlib and seaborn and folium for data visualization, json for json processing, requests for get requests into foursquare, and finally sklearn for machine learning, where we will import k-means cluster algorithm

2-Scrap Datas from Wikipedia and put in Dataframe

Datas will first be extracted from Wikipedia list of postal codes page.We will use a requests on Wikipedia page and then extract the results and transform it to csv format

3-Clean and wrangle datas

The datas will be cleaned, and “Not Assigned” datas in “Borough” column will be dropped

4-Import csv coordinates and merge with dataframe

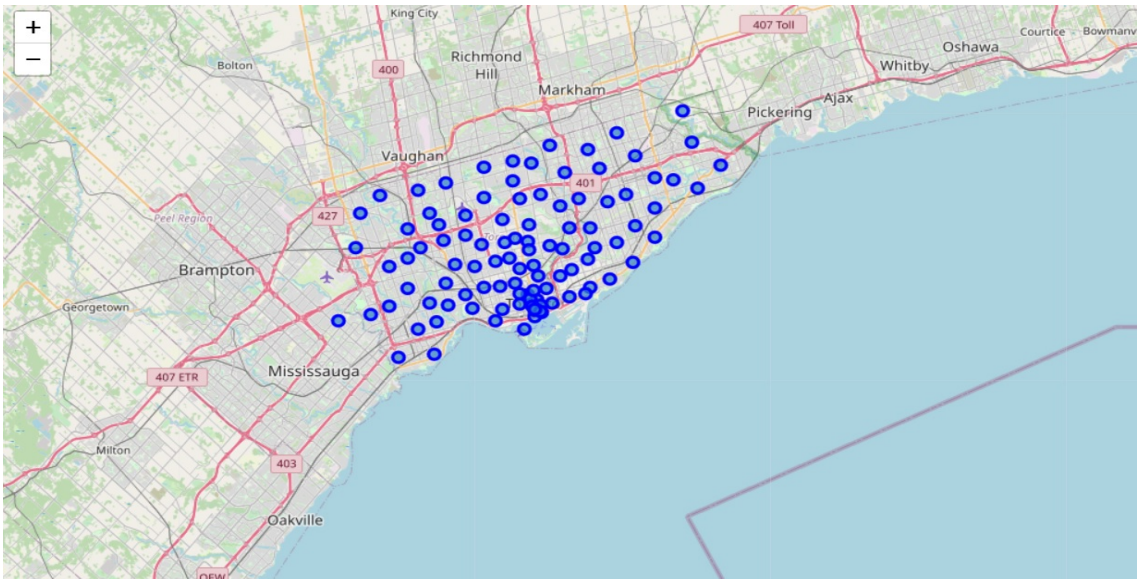
A csv geospatial coordinate will be imported as a dataframe, and then will be merged with the existing datas.We will then have a file with Borough, postal codes, and neighborhoods of Toronto with Longitude and Latitude

[11]:

	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
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724

5 - General Neighborhood map

After defining longitude and latitude, a Toronto neighborhood map will be displayed with folium



Input Foursquare Credentials requests and get json file

We will then need to input our Foursquare Credentials to be able make a requests and get a json file containing the places, restaurants,hotels,Ice creams, etc…

6 - Extract categories, clean json file, get nearby values

We will then extract categories from json file, get nearby values, clean them , define sme functions to help us process those datas

7- Get and process Toronto venues

We will use a get process to get the Toronto venues, clean them and process them.Some grouping and hot coding processes will be needed.Some sorting will also be needed.

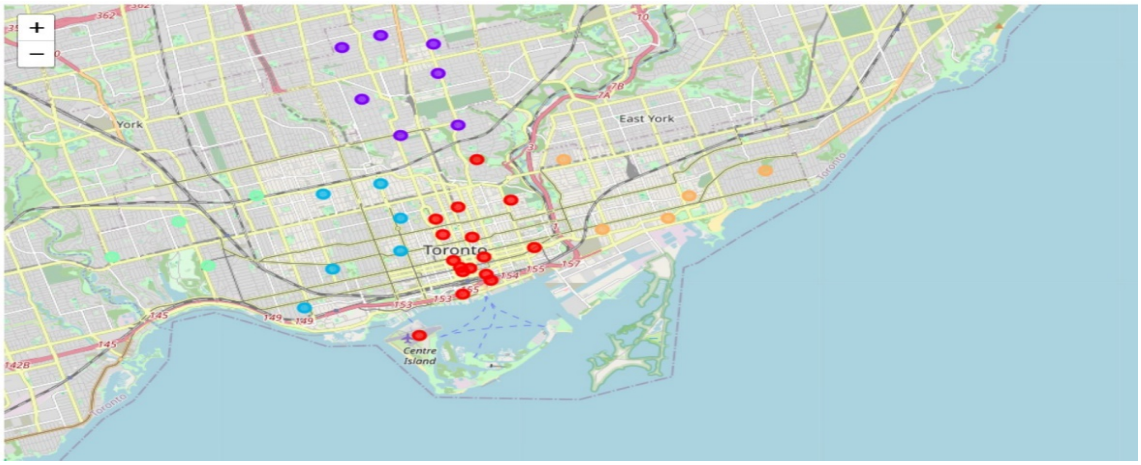
	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Central Toronto	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
1	Downtown Toronto	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
2	East Toronto	Greek Restaurant	Coffee Shop	Brewery	Italian Restaurant	Ice Cream Shop	Park	Pub	American Restaurant	Bakery	Bookstore
3	East York	Coffee Shop	Bank	Sandwich Place	Park	Burger Joint	Sporting Goods Shop	Pizza Place	Pharmacy	Intersection	Grocery Store
4	Etobicoke	Pizza Place	Sandwich Place	Gym	Coffee Shop	Grocery Store	Fast Food Restaurant	Pharmacy	Park	Discount Store	Café
5	Mississauga	Coffee Shop	Hotel	Sandwich Place	Gym	Fried Chicken Joint	Intersection	Mediterranean Restaurant	Middle Eastern Restaurant	American Restaurant	Gas Station
6	North York	Coffee Shop	Clothing Store	Restaurant	Pizza Place	Japanese Restaurant	Bank	Sandwich Place	Park	Shopping Mall	Grocery Store

8- K-Means clustering

For clustering, a k-mean clustering Machine Learning Algorythm will be used. A value of k=5 will be used for commodity, with randomstate=0.We will fit this model to Toronto clustering

9- Map the clusters

The clusters will be mapped with folium, to see a repartition of the different neighborhoods of Toronto.We will then wrangle the datas to examine precisely every cluster



Cluster 0

	Borough	ClusterLabels	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
61	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
62	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
67	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
68	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
73	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
74	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
79	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
83	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy
86	Central Toronto	0	Coffee Shop	Sandwich Place	Pizza Place	Café	Park	Sushi Restaurant	Dessert Shop	Restaurant	Clothing Store	Pharmacy

Cluster 3

	Borough	ClusterLabels	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
2	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
4	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
9	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
15	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
19	East Toronto	3	Greek Restaurant	Coffee Shop	Brewery	Italian Restaurant	Ice Cream Shop	Park	Pub	American Restaurant	Bakery	Bookstore
20	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
24	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
25	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
30	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Park	Bakery	Seafood Restaurant	Gym
31	West Toronto	3	Café	Bar	Coffee Shop	Italian Restaurant	Restaurant	Bakery	Breakfast Spot	Furniture / Home Store	Gift Shop	Grocery Store

Conclusion

- Cluster 0 neighborhood is more likely for citizen who like coffee, sushis, and some fast food
- Cluster3 is more likely for citizen who like diversified culture with restaurants from different countries, mixed population there maybe
- Cluster 0 and cluster 3 appeared to bé in the center town, regarding the density of shops, or near by a commercial area
- For citizen who want peacefull areas, then the remaining clusters might bé the best fit