Toronto - The Battle of the Neighborhoods

by melih canal

24.02.2021

INTRODUCTION:

In this project we tried to find the most intense locations where restaurants are intensively located. Specifically, this report will be targeted to investor who is already run wholesale food company and interested in opening an warehouse in Toronto, Canada.

Since there are lots of restaurants in Toronto, we will divide the amount of restaurants according to postcode and we will give to our investor the most crowded with restaurants.

DATA:

Based on definition of our problem, the main data that we will use is all existing restaurants in the neighborhood. We obtained the neccesary data by using Forsquare.*

We also need the full list of postcodes and their neighborhood in city of Toronto,Canada. Those datas has been imported from Wikipedia.**

* https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
** https://api.foursquare.com

METHODOLOGY:

We used the methodology of K-Means Clustering Algorithm. K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

Steps we took for the analysis:

- Collected required data that contains of every restaurant within our lat and lng.
- Explored the 'restaurant density' across different areas of Toronto - we will use K-mean to identify a few promising areas close to center with low number of restaurants and their type.
- We will take into consideration locations with restaurants in radius of 500 meters

ANALYSIS:

1. **Exploring data:** On this stage, we imported our data from *Wikipedia* as *HTML code* and converted it as *Pandas DataFrame* by using *BeautifulSoup* and *Requests* packages.

	PostalCode	Borough	Neighborhood
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
...	...	...	...
175	M5Z	Not assigned	Not assigned
176	M6Z	Not assigned	Not assigned
177	M7Z	Not assigned	Not assigned
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...
179	M9Z	Not assigned	Not assigned

180 rows × 3 columns

2. **Data cleaning:** On this stage, we cleaned our dataset, prepare for analysing. We red rid of “Not Assigned” datas on our dataset.

	PostalCode	Borough	Neighborhood
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
...	...	...	...
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

103 rows × 3 columns

Then , we combined different data source and sorting neighborhood based on longitude and latitude for working on geographical locations. The list of longitude and latitude were imported from outsources. Next stage was to tidy our dataset and combine all necessary data on our *DataFrame.*

	PostalCode	Borough	Neighborhood	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
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

In order to to integrate the list of restaurants with out dataset, we create Forsquare link.

```
LIMIT = 100
radius = 500
results_parkwoods = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={
CLIENT_ID,
CLIENT_SECRET,
VERSION,
neighborhood_latitude,
neighborhood_longitude,
radius,
LIMIT)
results_parkwoods
```

```
'https://api.foursquare.com/v2/venues/explore?&client_id=EULG4N4YWJMED5K3U3Z5R3XE5PIVIIUEJZGX3SVVXKJ1ZD2J&client_secr
et=1O20ZXC4ZRKHE2ZBZ350ZL5AATRCLN4Q0YHG5QOG03UDTBHD&v=20180605&ll=43.7532586,-79.3296565&radius=500&limit=100'
```

On this point, we noticed the problem that Forsquare did not classify stores with main restaurant title, therefore, we selected the *venue values* that are our target clients and we have obtains the data set below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
5	Victoria Village	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
6	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
7	Victoria Village	43.725882	-79.315572	Cash Money	43.725486	-79.312665	Financial or Legal Service
8	Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery

Now, we roughly have an idea where is the area that restaurants are mostly located. However, we should not analyse part by part but all Toronto, that is why, we are going to use Clustering method.

3. **Explore the Toronto's neighborhoods (Identify the top 10 venues for each neighborhood.):** On this stage , we have given the most 10 popular venues category. Information are may be beneficial for our clients which neighbourhood are the most suitable place for their product chain.

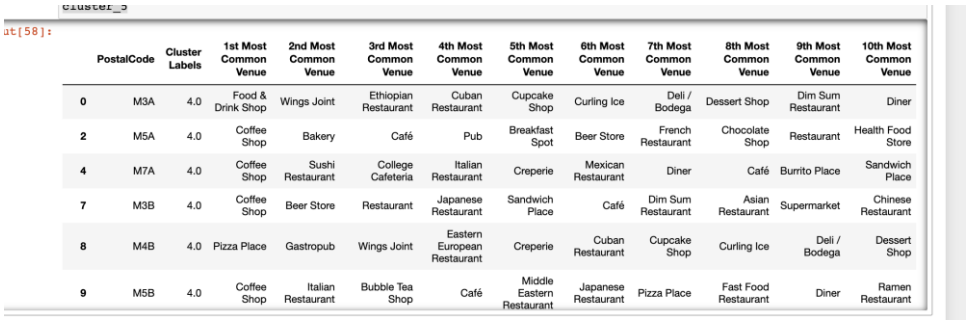
Out[42]:

	PostalCode	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 Co
0	M3A	North York	Parkwoods	43.753259	-79.329656	4.0	Food & Drink Shop	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Cupcake Shop	Curling Ice	Deli / Bodega	D
1	M4A	North York	Victoria Village	43.725882	-79.315572	1.0	Coffee Shop	Pizza Place	Portuguese Restaurant	Wings Joint	Comfort Food Restaurant	Cuban Restaurant	Cupcake Shop	C
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	4.0	Coffee Shop	Bakery	Café	Pub	Breakfast Spot	Beer Store	French Restaurant	Cho
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Vietnamese Restaurant	Coffee Shop	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Cupcake Shop	Curling Ice	B
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	4.0	Coffee Shop	Sushi Restaurant	College Cafeteria	Italian Restaurant	Creperie	Mexican Restaurant	Diner	

4. **Clustering(With an assumption of 5 clusters):** The final step of our project is *"clustering"*. With an assumption of 5 clusters, use K-Cluster algorithm to come up with 5 different clusters in Toronto with similar set of Venues.Explore each cluster and determine the discriminating venue categories that distinguish each cluster.

RESULTS:

Based on our initial assumption of the cluster with maximum number of restaurants will have the best possibility to have a new restaurant due to the need in the area. Based on the resultant clusters it looks like Cluster 5 have higher number of resturants then rest of the clusters.



	PostalCode	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	M3A	4.0	Food & Drink Shop	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Cupcake Shop	Curling Ice	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner
2	M5A	4.0	Coffee Shop	Bakery	Café	Pub	Breakfast Spot	Beer Store	French Restaurant	Chocolate Shop	Restaurant	Health Food Store
4	M7A	4.0	Coffee Shop	Sushi Restaurant	College Cafeteria	Italian Restaurant	Creperie	Mexican Restaurant	Diner	Café	Burrito Place	Sandwich Place
7	M3B	4.0	Coffee Shop	Beer Store	Restaurant	Japanese Restaurant	Sandwich Place	Café	Dim Sum Restaurant	Asian Restaurant	Supermarket	Chinese Restaurant
8	M4B	4.0	Pizza Place	Gastropub	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Cupcake Shop	Curling Ice	Deli / Bodega	Dessert Shop
9	M5B	4.0	Coffee Shop	Italian Restaurant	Bubble Tea Shop	Café	Middle Eastern Restaurant	Japanese Restaurant	Pizza Place	Fast Food Restaurant	Diner	Ramen Restaurant

Borough	Amounts of Rest.
Central Toronto	65
Downtown Toronto	819
East Toronto	76
East York	31
Etobicoke	36
Mississauga	10
North York	137
Scarborough	40
West Toronto	106
York	3

CONCLUSION:

The areas will be given to investor according to "Cluster 5" as the most intense areas which company's target clients are mostly located.. Now, our investor will make his final decision according to availibilty of warehouse facilities, rent fees and other factorts.