CAPSTONE PROJECT

A Presentation by MADHUMATHY K
December 15,2020

BUSINESS PROBLEM

There is a groceries contractor in one of the boroughs of Toronto (Scarborough). This contractor provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, Brewery and Café with fresh and high-quality groceries. The contractor wants to build a warehouse for the groceries it buys from villagers and farmers inside the borough, so that they will support more customers and also bring better "Quality of Service" to the old customers.

For example, if the warehouse is close to those old and famous restaurants, then the vegetables and other groceries would be delivered to the restaurant in the right time and there would be no delay so the restaurant cooks can start their job from the morning and the Quality of Service will be high and this contractor will gain more reputation and income.

DATA REQUIRED

GEO DATA

1- We will need geo-locational information about that specific borough and the neighborhoods in that borough. We specifically and technically mean the latitude and longitude numbers of that borough. We assume that it is "Scarborough" in Toronto. This is easily provided for us by the contractor, because the contractor has already made up his mind about the borough. The Postal Codes that fall into that borough (Scarborough) would also be sufficient fo us. I fact we will first find neighborhoods inside Scarborough by their corresponding Postal Codes

[Postal Code] [Neighborhood(s)] [Neighborhood Latitude] [Neighborhood Longitude] [Venue] [Venue Summary] [Venue Category] [Distance (meter)]

FOURSQUARE DATA

2- We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. By locational information for each venue we mean basic and advanced information about that venue. For example there is a venue in one of the neighborhoods. As basic information, we can obtain its precise latitude and longitude and also its distance from the center of the neighborhood. But we are looking for advanced information such as the category of that venue and whether this venue is a popular one in its category or maybe the average price of the services of this venue.

https://foursquare.com

DERIVE THE NEIGHBOURS OF TORONTO

By scraping the data from the wikipedia find the data set about the toronto

url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:'

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern / Rouge
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek
2	M1E	Scarborough	Guildwood / Morningside / West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park / Ionview / East Birchmount Park
7	M1L	Scarborough	Golden Mile / Clairlea / Oakridge
8	M1M	Scarborough	Cliffside / Cliffcrest / Scarborough Village West
9	M1N	Scarborough	Birch Cliff / Cliffside West
10	M1P	Scarborough	Dorset Park / Wexford Heights / Scarborough To...

Collect Geospatial data

Collect the longitude and latitude of the data from the spatial data

url="http://cocl.us/Geospatial_data"

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Using Foursquare to collect data

Use your client_ id and client_data and version to obtain the data

```
url = 'https://api.foursquare.com/v2/venues/explore?
client_id={}&client_secret={}&ll={},
{}&v={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude_scar,
longitude_scar, VERSION, radius, LIMIT)
```

Using K-means for clustering and segmenting the data

Here we cluster neighborhoods via K-means clustering method. We think that 5 clusters is enough and can cover the complexity of our problem. After clustering we will update our dataset and create a column representing the group for each neighborhood.

from sklearn.cluster import KMeans

Decision Making and Reporting Results

.

Now, we focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Total Restaurants + Total Joints + Other Venues.} This algorithm although is pretty straightforward yet is strongly powerful.

Now, we focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Total Restaurants + Total Joints + Other Venues.} This algorithm although is pretty straightforward yet is strongly powerful.

Results

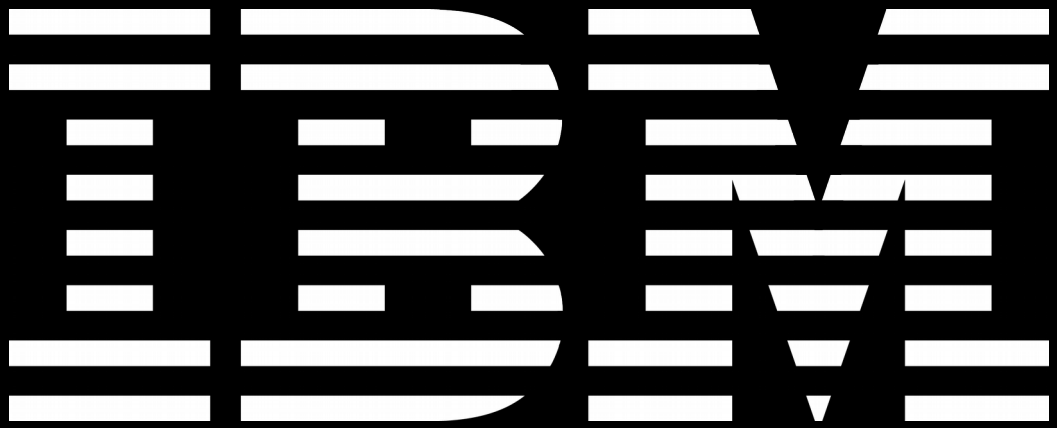
Based on this analysis, the best recommended neighborhood will be:

{'Neighborhood': 'Agincourt',

'Postal Code': 'M1S',

'Neighborhood Latitude': 43.7942003,

'Neighborhood Longitude': -79.26202940000002


IBM