A project to identify suitable location(s) for constructing warehouse -Recommender System

Problem Description

- This project tries to help a contractor based in Toronto who regularly supply raw materials (food ingredients and other related items) to 'hotels & restaurants'.

- Contractor’s main business objective is 'good quality raw materials and timely delivery'. To achieve this he needs to have a good warehouse on a right location which is close both to 'farmers & producers' and to his customers. Hence the project aims to suggest a suitable location for the warehouse using suitable machine learning algorithm.

- We tries to develop a recommender system with

1. Sorted list of neighborhood in order to minimize the cost of transportation
2. A neighborhood the contractor can have his warehouse.
3. Some other useful findings on data analysis

Required Data

- The project needs geo-locational information about the specific borough/town and the neighborhoods of the contractor's choice. For example, "Scarborough". We have the borough details with postal codes and the details of the neighborhoods of the corresponding borough/town.
- The project also requires venues in the different neighborhoods of that specific borough. Location data provider - 'Foursquare' can help in this regard.
- Our main data sources are,

  Borough/Town and Neighbourhood information as a csv file.

  Foursqure (Foursquare have different api's for developer community to provide with required data)

  Anything extra needed can be sourced as and when needed.

Methodology

Data Collection.

- Scarborough and its neighborhoods is our focus area. We are having a CSV file with all neighborhood details of Scarborough. We have collected it from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
- 'Foursquare' is the data source for getting location data and using 'Foursquare' APIs we are retrieving location data.
- As we need the details of all venues belonging to 'Scarborough' we use an appropriate API to collect the same from 'Foursquare'. We set the radius limit as 1000 meters. Having fixed that 'Foursquare' will supply us with the venues that are within 1000 meter distance from the center point of the neighborhood.

Scarborough and its neighbourhood dataframe.

Postal code	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353
M1C	Scarborough	Rouge Hill / Port Union / Highland Creek	43.784535	-79.160497
M1E	Scarborough	Guildwood / Morningside / West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476
M1J	Scarborough	Scarborough Village	43.744734	-79.239476
M1K	Scarborough	Kennedy Park / Ionview / East Birchmount Park	43.727929	-79.262029
M1L	Scarborough	Golden Mile / Clairlea / Oakridge	43.711112	-79.284577
M1M	Scarborough	Cliffside / Cliffcrest / Scarborough Village West	43.716316	-79.239476

Scarborough-Neighbourhood venue-dataframe.

Unnamed: 0	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M1B	Malvern / Rouge	43.806686	-79.194353	Harvey's	This spot is popular	Restaurant	807
1	M1B	Malvern / Rouge	43.806686	-79.194353	Wendy's	This spot is popular	Fast Food Restaurant	600
2	M1B	Malvern / Rouge	43.806686	-79.194353	Wendy’s	This spot is popular	Fast Food Restaurant	387
3	M1B	Malvern / Rouge	43.806686	-79.194353	RBC Royal Bank	This spot is popular	Bank	906
4	M1B	Malvern / Rouge	43.806686	-79.194353	Caribbean Wave	This spot is popular	Caribbean Restaurant	912
5	M1B	Malvern / Rouge	43.806686	-79.194353	Staples Morningside	This spot is popular	Paper / Office Supplies Store	735
6	M1B	Malvern / Rouge	43.806686	-79.194353	Tim Hortons	This spot is popular	Coffee Shop	605

Neighborhood Summary Information

Postal Code	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
Neighborhood						
Agincourt	44	44	44	44	44	44
Birch Cliff / Cliffside West	14	14	14	14	14	14
Cedarbrae	29	29	29	29	29	29
Clarks Corners / Tam O'Shanter / Sullivan	35	35	35	35	35	35
Cliffside / Cliffcrest / Scarborough Village West	12	12	12	12	12	12

Contd...

- Further, we have to find desirable features of each venue. We decide category of the venue as the main feature. Then the venue's category is One-hot encoded. After One-hot encoding we are integrating all restaurant columns to one column ‘Total Restaurants’ and all food joint columns to ‘Total Joints’ column, leave the rest columns as it is. Further, we assume that all restaurants use mainly the same raw materials. Hence the dataset is ready for further analysis using a suitable machine learning algorithms.

K means Algorithm

- We decide K-Means Clustering as our Machine Learning algorithm for further analysis. We decide to create 5 clusters of neighborhoods using K-means clustering method. After clustering the neighborhoods we will update our dataset and create a column representing the group for each neighborhood.

- We apply the algorithm and the sample code is as follows:

```
# import k-means
from sklearn.cluster import KMeans
# run k-means clustering
kmeans = KMeans(n_clusters = 5, random_state = 0).fit(scarborough_onehot)
```

Displaying centers of each cluster

Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints	Total Sum
G3	2.000000	1.000000	0.0	0.0	0.00	0.000000	1.00	1.000000	1.000000	19.000000	0.000000	25.000000
G4	1.000000	0.000000	0.0	0.0	0.00	0.666667	1.00	1.666667	1.000000	12.666667	1.333333	19.333333
G2	2.000000	0.500000	0.0	0.5	0.00	2.000000	0.00	1.500000	0.000000	8.000000	2.500000	17.000000
G5	0.750000	0.250000	0.0	0.0	0.25	1.250000	0.25	1.500000	1.000000	6.250000	0.500000	12.000000
G1	0.333333	0.166667	0.5	0.0	0.00	0.333333	0.00	0.833333	0.333333	2.166667	0.333333	5.000000

From the above it is clear that the best group is G5 followed by G1, G4

Inserting "kmeans.labels_“ and the result is as follows

Neighborhood	Group
Agincourt	3
Birch Cliff / Cliffside West	1
Cedarbrae	2
Clarks Corners / Tam O'Shanter / Sullivan	4
Cliffside / Cliffcrest / Scarborough Village West	1
Dorset Park / Wexford Heights / Scarborough To...	4
Golden Mile / Clairlea / Oakridge	1
Guildwood / Morningside / West Hill	5
Kennedy Park / Ionview / East Birchmount Park	5
Malvern / Rouge	5
Milliken / Agincourt North / Steeles East / L'...	4
Rouge Hill / Port Union / Highland Creek	1
Scarborough Village	1
Steeles West / L'Amoreaux West	5
Wexford / Maryvale	2
Woburn	1

Results & Findings

- The group with its center has the highest ‘Total Sum’ will be our best recommendation.
- Based on this analysis, the best recommended neighborhood is as follows,

Neighborhood	Agincourt
Postal Code	M1S
Neighborhood Latitude	43.7942003
Neighborhood Longitude	-79.26202940000002