COFFEE SHOP RECOMMENDER SYSTEM
- IBM Applied Data Science Capstone Project

Contents :

1. Introduction
2. Data
3. Methodology
4. Result
5. Recommendation
6. Conclusion

INTRODUCTION :

PROBLEM:

" To develop a recommender system that will help the restaurant manager to find the best suitable place to open an Coffee Shop. "

PROBLEM BACKGROUND:

Toronto is the provincial capital of Ontario and the most populous city in Canada, with estimated population of 2,956,024 and an estimated population of 6,341,935 in the Toronto region. Toronto is an International centre of business, finance, arts, and culture. Its large population of immigrants from around the globe has also made Toronto one of the most multicultural and cosmopolitan cities in the world.

The city has many restaurants, coffee shops, cafe, hotels. The variety of food items are provided by these shops. One of the popular item is an Coffee. Thus, there are many Coffee Shops in the various areas of the city. Therefore, if someone decides to open an Coffee Shop in the city, he would select the best suitable place for the shop. To get the information about this suitable place, the recommender system can be used. This recommender system will get the data, analyze and visualize it; and then provide the best place to open a shop or restaurant.

PROBLEM DESCRIPTION: If the manager decides to open an Coffee Shop in Toronto, then following questions must be addressed –

1. Which place is the most suitable and popular for the Coffee Shop ?

2. What type of Coffee should be provided ? What type is preferred by people in that area ?

3. What type of people live in that area (students, company employees, etc) ?

4. How many similar shops are present in that area ?

5. What other specialities should be provided to attract customers ?

6. What should be the cost of the Coffee provided ?

To get answers of above questions, a recommender system can be implemented. This system can help the managers to decide the best place to open a Coffee Shop based on the rankings of places.

The recommender system should be able to answer the all questions and provide best solution. It should provide – popular items in that area, target audience, their average income, cost estimation, etc.

TARGET AUDIENCE:

Target audience for this system are the managers or people who want to open an Coffee Shop in the specific city or area. These people expect the place which is most popular and well known in the city. They also need the information about popular items, categories, other specialities, etc. Thus using

this recommendation system, the managers can decide the most suitable place for the Coffee Shop.

DATA :

To build a recommender system to find the best suitable place for an Coffee Shop, huge amount of data is required. The datasets are used to analyze the data, visualize the data, and the finally to get suitable solution to the problem.

To open a shop, following things are required –

1. Geographical coordinates of the area
2. The population of the neighbourhood
3. The type of people in the neighbourhood
4. Average income of the people nearby that area
5. The preference of people towards the type of food
6. Other service details such as juice, transport, taxi, etc.

The more popular the area, more chances to successfully provide service to people. The type of people (students, or employee) plays the important role to decide what type food should be provided.

The average income helps to decide the cost of the food items. The preference of the people should be taken in consideration. Other special services such as juice help to increase the business opportunities.

Thus, to get the solutions, a large amount of data is required.

DATA COLLECTION:

The names, postal codes and addresses of the places around the city can be obtained from various websites such as Wikipedia. The geographical coordinates data can be obtained from the various open source websites such as Wikipedia, Google Maps, Government websites, census report websites, etc.

The population of the area can be easily obtained by searching on the websites.

The average income data can be obtained from Wikipedia, or other sources.

Foursquare API can be used to get the information about the nearest locations. This data can be used to visualize, clustering the results.

METHODOLOGY :

The first task was to get the data of neighborhoods in Toronto. The list of
neighborhoods of Toronto was extracted from the Wikipedia page. The list
consists of the postal code, borough and the area name.

[3]:	Postal code	Borough	Neighborhood
0	M1A	NaN	NaN
1	M2A	NaN	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront

Then I used the pandas HTML table scraping method for web scraping.

Then to get the longitude and latitude coordinates of the areas, the CSV
file was used.

	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

After matching the areas and the coordinates, I visualized the map of Toronto using the Folium library package.

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood / Morningside / West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Then to get the list of top nearest venues, I used the Foursquare API, using my ID, secret key and version. Then I obtained the names, addresses, longitude, latitude of the nearest venues using it.

Then I grouped the venues by their categories. Then I selected the category as " Coffee Shop ". Then I used the K-Means clustering method to form the clusters of the data.

Then by analysing the results, the final results were obtained.

RESULT :

The clusters obtained using the K-Means clustering method :

Cluster 0

	Neighborhood	Coffee Shop	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
28	Runnymede / Swansea	0.055556	0	43.651571	-79.484450	Wibke's Espresso Bar	43.649132	-79.484802	Coffee Shop
28	Runnymede / Swansea	0.055556	0	43.651571	-79.484450	Tim Hortons	43.648526	-79.485066	Coffee Shop
25	Richmond / Adelaide / King	0.095745	0	43.650571	-79.384568	Starbucks	43.646891	-79.381871	Coffee Shop
25	Richmond / Adelaide / King	0.095745	0	43.650571	-79.384568	Dineen @CommerceCourt	43.648251	-79.380127	Coffee Shop
25	Richmond / Adelaide / King	0.095745	0	43.650571	-79.384568	Starbucks	43.649028	-79.381593	Coffee Shop
...	...	...	...	...	...	...	...	...	...
13	Garden District, Ryerson	0.090000	0	43.657162	-79.378937	Hailed Coffee	43.658833	-79.383684	Coffee Shop
13	Garden District, Ryerson	0.090000	0	43.657162	-79.378937	Balzac's Coffee	43.657854	-79.379200	Coffee Shop
13	Garden District, Ryerson	0.090000	0	43.657162	-79.378937	Nordstrom Ebar	43.654649	-79.380574	Coffee Shop
	Brockton / Parkdale Village /								

Cluster 1

	Neighborhood	Coffee Shop	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
19	Little Portugal / Trinity	0.02381	1	43.647927	-79.41975	Jimmy's Coffee	43.644521	-79.418908	Coffee Shop

RESULT :

By analyzing nearby venues, we can conclude that the cluster 1 does not have many Coffee shops in that areas. Thus, it would be suitable to select these locations for opening Coffee shops.

RECOMMENDATION :

Therefore, locations like Central Bay Street, Riverdale, The Beaches West, Commerce court will be good to open a new Coffee Shop.

CONCLUSION :

The recommender system correctly recommends the most suitable place to open an Coffee Shop. Thus, it can provide good results to the users of the system. The system can also be used as recommendation system for opening the restaurants, coffee shops, street food shop, etc. Using this method the recommendation system for malls, theatres, shops can also be designed.