Capstone Project -The Battle of Neighborhoods(Report)

Topic:- Recommendation for Opening a Fitness Center in North York,Toronto

Description of Problem:

Which locality is best to open a fitness center in North York, Toronto ?

Background :

There is a successful Entrepreneur in Toronto and now he wants to expand his business in other boroughs in Toronto. He wants to open a new fitness center in North York,Toronto. But not sure in which locality he should open the fitness center and get profitable business. The borough preference is given by the person. We need to provide a best recommendation to him which helps him to grow his business fast and make it stable.

Analytical approach-

According to the need of client first we have to find the City and then the Locality in that city for Business. Fitness center should be at central place near any park,sports club or shopping mall which will be easy for people to do exercise along with their any kind of work. For this particular problem we will set few Benchmarks for the City and the locality in which opening a fitness center would be successful.

1. 'Park',
2. 'Shopping Mall',
3. 'Recreation Center',
4. 'Pool',
5. 'Residential Building (Apartment / Condo)',
6. 'Convenience Store',
7. 'Toy / Game Store',
8. 'Salon / Barbershop',
9. 'Department Store',
10. 'Sporting Goods Shop',
11. 'Skating Rink',
12. 'Trail',
13. 'Supermarket',
14. 'Sports Bar',
15. 'Lounge',
16. 'Yoga Studio',
17. 'Gym',
18. 'Intersection',
19. 'Bowling Alley',
20. 'Tennis Court',
21. 'Golf Course',
22. 'Playground',
23. 'Road',24. 'Gym / Fitness Center',
25. 'Athletics & Sports',
26. 'Video Game Store',
27. 'Ski Chalet',
28. 'Community Center',
29. 'Ski Area',
30. 'Massage Studio',
31. 'Soccer Field',

32. 'Baseball Field',
33. 'Hockey Arena',
34. "Men's Store",
35. 'Boxing Gym',
36. 'Sports Club'

Data Requirements and Description:

1- We will need Geo-location 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 "North York" in Toronto. This is easily provided for us by the client, because the client has already made up his mind about the borough. The Postal Codes that fall into that borough (North York) would also be sufficient for us. In fact we will first find neighborhoods inside North York by their corresponding Postal Codes.

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" location information. By location 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. A typical request from Foursquare will provide us with the following information.

Source of Data:

Wikipedia:
- ' https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Four Square API:
- For Venues in City and in Each locality in selected city.

Data Description (How to use Data to solve the Problem?)

By Using the Wikipedia data we will first analyze the data and then focus on certain benchmarks to discover the best locality in North York,Toronto to open a Fitness Center. Then by leveraging the Foursquare data we will compare and try to find the best locality. Based on Scores which we get by using clustering methods will try to determine the locality in similar way. Finding the right neighborhood is our mission and our recommend system will provide this person with a sorted list of neighborhoods in which the first element of the list will be the best suggested neighborhood.

Methodology and Discussion:

Exploring Neighborhoods of "North York":

We will use Postal Codes of different regions inside North York to find the list of neighborhoods. We will essentially obtain our information from https://en.hwikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and then process the table inside this site. Images from data frames and also from maps will be provided in the presentation. Here we only present our strategy and how we got the mission accomplished.

Connecting to Foursquare and Retrieving Location Data:

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood. (I think distance is measured by latitude and longitude of venues and neighborhoods, and it is not the walking distance for venues.)

Processing the Retrieved Data and Creating a Data Frame for All the Venue:

When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" will be One-hot encoded and different venues will have different feature-columns. We will set few Benchmarks for the City and the locality in which opening a fitness center would be successful This assumption is made for simplicity and due to not having a very detailed data set about different venues.
Now, the data set is fully ready to be used for machine learning (and statistical analysis) purposes.

Applying one of Machine Learning Techniques (K-Means Clustering):

Here we cluster neighborhoods via K-means clustering method. We think that 4 clusters is enough and can cover the complexity of our problem. After clustering we will update our data set and create a column representing the group for each neighborhood.

Decision Making and Reporting Results:

Now,on the basis of centers of clusters and compare them from their "Total Sum". The group which its center has the highest "Total Sum" will be our best recommendation to the person. This algorithm although is pretty straightforward yet is strongly powerful.

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

```
{'Postal Code': 'M3C',
 'Neighborhood': 'Flemingdon Park, Don Mills South',
 'Neighborhood Latitude': 43.72589970000001,
 'Neighborhood Longitude': -79.340923}
```

Conclusion:

After analyzing the cluster we came to the conclusion that “Flemingdon Park, Don Mills South” is a best place in North York borough to set a fitness center to get good growth in the business.