PROJECT REPORT

A. Introduction

Problem Statement- In Delhi, if someone is looking to open an Indian restaurant, where would you recommend that they start it?

This project aims to find the best location to open an Indian restaurant in the city of Delhi, India, to maximize the profit of the owner. The target audience for this project is the people who want to start an Indian restaurant.

Our subject wants to own a place at a location that would be best-suited to open an Indian restaurant. Buying property for your Indian restaurant will probably be your most significant start-up expense. Our subject wants to invest in the best location, so our subject desires us to analyze all the findings and then recommend some of them. After finding some suitable locations, our subject e will filter those locations out according to comforts and budget.

So, we have to provide her the best recommendations, the best locations in Delhi, and she'll choose them accordingly.

B. Data Description

We have been dealing with the following data in our project:-

- ✓ A dataset containing Neighborhoods of Delhi, the dataset holds the names Neighborhood in Delhi extracted from Wikipedia. Here is a link to the dataset https://github.com/Sukrati192/Coursera_Capstone/blob/master/delhi.csv

- ✓ I used python library geopy to distill coordinates of neighborhoods in the data frame. Data required some scraping for geopy to correctly give the coordinates of Neighborhoods of Delhi in a data frame.
- ✓ I used Foursquare API to explore the most common places of a neighborhood in the form of a JSON file.

C. Methodology

I extracted the neighborhoods of Delhi from Wikipedia into a CSV file. Then, read the record into a pandas dataframe. I found that there were 117 neighborhoods of Delhi. Using geopy, I extracted the geographical coordinates of the areas. Using folium, I visualized all the communities of Delhi on a map.

Using Foursquare, I extracted the nearby venues of all the neighborhoods of Delhi and counted all the sites in a dataframe. I concluded that there were 164 unique categories for all the nearby centers. Then, I analyzed the neighborhoods by each venue category.

After all the analysis, I took out the extraction of 10 most ordinary venues for each neighborhood in a dataframe out. Then, I used the K-means clustering to cluster similar parts of Delhi.

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 provided. Data points are clustered based on feature similarity.

So, I grouped the neighborhoods in 5 clusters and then visualized them in a map using folium.

D. Results

After doing the fit on the data, the model returned 5 clusters, i.e., Cluster 0 to Cluster 4. Each batch comprises of many neighborhoods similar to each other. That means sections or localities which are in cluster 0 are related to all other parts or sites in cluster 0. same goes for different groups. The algorithm computes the similarity between neighborhood by computing the Euclidian distance, the lesser the Euclidian distance, the more the similarity.

The neighborhoods in each cluster are:

- Cluster 0- 31 Neighborhoods
- Cluster 1- 3 Neighborhood
- Cluster 2- 2 Neighborhoods
- Cluster 3- 71 Neighborhood
- Cluster 4- 3 Neighborhoods

By observation at each cluster's dataframe, it is acutely explicit that Cluster 3 contains fewer restaurants but comprises of areas with grand footfall. So, to open an Indian restaurant, Cluster 3 is the best match.

I eliminated the neighborhoods which have any restaurant as their nearby venue. So, there are a total of 10 areas. From these 10 regions where should our subject open an Indian restaurant.

Now, the subject can choose the best location according to the budget among them.

D. Discussion

Delhi is a big city with diverse neighborhoods and venues. Different clustering algorithms can yield different clusters.

I used the K Means technique since I have suitable experience with it. I tested the algorithm for different values of k ranging from 1 to 10 and used the elbow method to check the accuracy scores with each value of k. I found k=5 to be the optimum value of k. I used the coordinates of all the neighborhoods of Delhi. Some localities didn't return to popular venues. Therefore, I couldn't determine their cluster by K means algorithm, although their number is exceedingly small, and so they do not deviate our accuracy.

I used python library **folium** to create map visualizations of cluster data.

I ended the study by finding the ten best neighborhoods to open an Indian restaurant. In future studies, I will experiment with some other criteria.

F. Conclusion

To start a new business, people usually look for the best locations to maximize profit with the least investment. Studies like this can help people to find the best neighborhoods to start a new business.

Real Estate companies can help their customers by providing a better and more regular location to start a new business by using similar data analysis.