Applied Capstone Project: The Battle of Neighborhood in Kolkata

Searching for place in kolkata city, capital of West Bengal to launch pizza place, indian restaurant and hotel

Swattik Chakrabarty

Introduction:

Kolkata is considered one of the largest cities in India. It is among the top metro cities in India and is the capital of West Bengal. Nowadays, the hotel management industry is becoming one of the leading industries among all. This project deals with the major venue categories in the neighborhoods of Kolkata. This project would specifically help Business personnels plan, to start new restaurants, hotels, pizza places, etc.

Aim:

This project aims to find the best location to open a pizza place,Indian restaurant and hotel in the city of joy, Kolkata, to maximize the profit of the owner. We want to open these in a neighborhood which are attractive in business aspects yet less crowded with already existing venues as mentioned above.

Data Requirements:

- ★ Dataset containing Neighborhoods of Kolkata- This dataset holds the names of Neighborhoods and Landmarks in Kolkata extracted from Nivalink with the help of Beautiful Soup package.

    - Here is a link to the dataset http://www.nivalink.com/localities-and-landmarks/kolkata
- ★ Python library geopy was used to get the coordinates of neighborhoods in the dataset. Data required repeated calls from geopy to correctly give the coordinates of Neighborhoods of Kolkata in the used data frame.
- ★ Foursquare API was used to extract the most common places of a neighborhood in the form of a JSON file and was put in the final dataset to explore the venues.

Methodology:

As our data does not have the coordinates of the location that we extracted, we will use the geocoder package to get the latitudes and longitudes of 65 neighborhoods.With the help of Folium let's visualize the neighborhood data that we collected. This will be done by using the coordinates of the location that we collected and updated in our pandas dataset.Here with the help of the FourSquare API we will try to collect popular venues around our locations. In order to make a good prediction and create a good analysis let's consider only the neighborhoods with more than 40 venues, as the neighborhoods with less than 40 venues are less popular and hence less populated. So as we want the new business owner to make profit in their ventures we should only consider neighborhoods which are popular.We performed one hot encoding on the obtained data set and used it to find the 10 most common venue categories in each neighborhood. Then clustering can be performed on the dataset.We are using the K - Nearest Neighbor clustering technique. To find the optimal number of clusters, silhouette score technique is used. Initially, in the original data frame we only had the neighborhoods with their top 10 venues. After fitting the data on the K-means model we insert a new column in the pandas dataframe, as cluster label, with the help of our model so that each neighborhood can be identified with the name of their cluster and segregation of neighborhoods according to their respective cluster can take place

Results:

- There are 15 Pizza places ,15 Multiplex , 14 Indian Restaurant and 8 Hotels in cluster 1
- As there are no pizza places in the top 10 venues of this cluster hence the number of pizza places must be less than 10 because here the 10th most popular venue is a pub whose value count is 10. There are 18 Multiplex , 30 Indian Restaurant and 30 Hotels in cluster 2
- As there are no Multiplexes and Hotels in top 10 venues of this cluster hence the number of multiplexes and Hotels must be less than 3 because here the 10th most popular venue is lounge whose value count is 3.There are 5 Pizza Places and 5 Indian Restaurant a in cluster 3
- As there are no pizza places and Indian restaurants in top 10 venues of this cluster hence the number of pizza places and Indian restaurant must be less than 2 because here the 10th most popular venue is Bengali restaurant whose value count is 2. There are 3 Multiplexes and 3 Hotels in Cluster 4

Discussion/Observation:

Hotels:

As there are no pizza places and Indian restaurants in the top 10 venues of this cluster hence the number of pizza places and Indian restaurants must be less than 2 because here the 10th most popular venue is Bengali restaurant whose value count is 2. There are 3 Multiplexes and 3 Hotels in Cluster 4

Fast Food Restaurant:

Cluster 3 and Cluster 4 are the least popular with the lowest number of neighborhoods and venues hence opening a fast food restaurant here will not fetch the owner profits Out of the most popular clusters i.e, Cluster 1 and Cluster 2, the neighborhoods in cluster 1 has the greatest number of fast food restaurants(15), hence opening one here is not ideal.

Thus opening a hotel in Cluster 2 is most ideal as it has multiple popular and most neighborhood options and no fast food restaurants. This rarity of fast food restaurants in such a popular cluster can be utilized by business venturers to maximize profits.

Potential Neighborhoods are: Tollygunge Metro Station,Esplanade Metro Station,Fort William

Pizza Place:

Cluster 1 has the most number of Pizza Places, so it doesn't make any sense to open a Pizza Place there.

However Cluster 2 being the most popular and crowded place among all the clusters, and if it does have any pizza place it will be less than 8. Hence the most popular place to open a pizza place is Cluster 2. Alternatively Cluster 3 and Cluster 4 can also be suggested as Cluster 3 has only 5 and Cluster 4 does not have any. But as cluster 4 is the least popular it is probably not a good idea to open it there.

Potential neighborhoods to open: Townhall, Esplanade Metro Station, Fort William

Multiplexes:

The most popular clusters are crowded with multiplexes so opening a multiplex there will face a lot of competition.

Cluster 4 being the least lost popular cluster is still crowded with multiplexes(5). Hence Cluster 3 is the most suitable to open a multiplex as it is relatively popular than Cluster 4 and even if it has multiplexes then they will be less than 3 therefore it will face less competition.

Potential neighborhoods to open: Tollygunge Metro Station, Kalighat

Conclusion:

Reason for this venture was to investigate the neighborhoods of Patna and make a grouping model to recommend spots to begin another business dependent on the class. The neighborhoods information was gotten from nivalink and the Foursquare API was utilized to track down the significant settings in every area. Then, at that point areas were utilized to make a bunching model. The best number of bunches for example 4 was acquired utilizing the outline score. Each bunch was analyzed to track down the most scene classes present, that characterizes the qualities for that specific group. A guide showing the bunches has been given. Both of these can be utilized by partners to choose the area for the necessary kind of business.