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