IBM Applied Data Science Capstone

Week 4 Assignment Submission

1. Introduction / Business Problem

A restaurant business company would like open an Indian restaurant in the San Francisco Bay Area, California. There are a lot of Indians working in the Bay area tech companies. Also, considerable number of Indian students are studying in the universities and colleges located in the Bay area. Along with Indians, people from other nations also like Indian food very much. Therefore, San Francisco Bay area has good potential to open an Indian restaurant.

San Francisco Bay area has made up of nine counties, namely, Santa Clara, Alameda, San Mateo, San Francisco, Contra Costa, Marin, Solano, Napa, and Sonoma. There are 186 ZIP codes and 97 boroughs in the Bay area. Since there are a large number of boroughs in the Bay area, it is important to understand which place is the best to start a new branch of an Indian restaurant.

As a data analyst, we should help the restaurant business company that wants to start an Indian restaurant by recommending the best suitable borough and explain why this place is better than others in the Bay area. The success criteria of the project will be a good recommendation of borough for an Indian restaurant business.

2. Data

The San Francisco Bay Area ZIP codes and PO_NAME are downloaded from DataSF open data website. The following csv file downloadable link is used to create a Pandas dataframe in Python. https://data.sfgov.org/api/views/f9wk-m4qb/rows.csv?accessType=DOWNLOAD
From this dataframe, a new dataframe is generated with ZIP codes and PO_NAME, and then PO_NAME is renamed to 'Borough'.

The geographical coordinates are retrieved for each borough in the dataframe using the geocoder and then leveraged the Foursquare AIP service (https://foursquare.com) to collect the location data, explore each borough of the Bay area and obtain information about different venue categories, especially restaurants, cafes, bars, pubs, etc. within 1000m radius from the geographical coordinates. This data is used to understand better that how many eateries available in the Bay area boroughs. Then the dataset is analyzed systematically to identify best suitable boroughs to open an Indian restaurant business.