Capstone Project – Battle of Neighborhood

IBM – Applied Data Science Project

Indian Restaurant in San Francisco Bay Area

By: Mayuree Deshmukh
(17th October 2019)

Introduction

San Francisco Bay Area is home to approximately 7.75 million people, coming from various parts of counties, states, cities, etc. North-California's nine-county Bay Area has some of the best Schools, Universities and most innovative technology companies in the world, making it famous for its abundance of talent in education & tech industry, making it very diverse human population. Considering the population, it has almost every kind of Restaurants like Italian, Chinese, Mexican, Indian, Japanese, etc. So, a business of opening any restaurant of specific cuisine is a challenging, yet profitable business. One of the most loved cuisine is Indian, because of its spices and flavors. As with any business decisions, it does require serious considerations and one of the most important decision will the 'Location' of the restaurant.

Business Problem

The objective of this Capstone project is to analyze and select the best locations in the San Francisco bay Area to open a new Indian Restaurant. Using data science methodology and Machine Learning techniques like clustering, this project aims to provide solution to answer the business question: Which is the best location for opening an Indian Restaurant in San Francisco Bay Area?

Target Audience of this Project

This project is useful for the entrepreneurs and investors, who are looking for opening Indian Restaurant in San Francisco Area.

Data

To solve this problem, we will need following data:

- List of Neighborhood in San Francisco Bay Area
- Latitude and Longitude coordinates of those Neighborhood
- Venue data related to the Indian Restaurant. This will be required to find suitable Neighborhood to open Indian Restaurant

Extracting the Data & Method

- Web Scraping: List of Neighborhoods in San Francisco from Wikipedia.
- Getting Latitude and Longitude data of neighborhoods via *Geocoder* python package.
- Using Foursquare API, to get venue data related to the neighborhood.
- Data cleaning and data Wrangling with the help of python package *pandas.*
- Machine Learning K-Means Clustering algorithm, to cluster the neighbors with same venue category.
- Data Visualization with the help of python package *Folium*.