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

Data is extracted majorly from the following Wikipedia page: (https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area), which contains information of 101 Cities and Towns in San Francisco Bay Area. We have used **Web Scarping** techniques to fetch this data from **Wikipedia** with the help of Python requests and Python web scrapping library *BeautifulSoup4*. Once this data is successfully fetched, latitude and longitude details of all the fetched cities and towns will be noted with the help of **Python Geocoder package**.

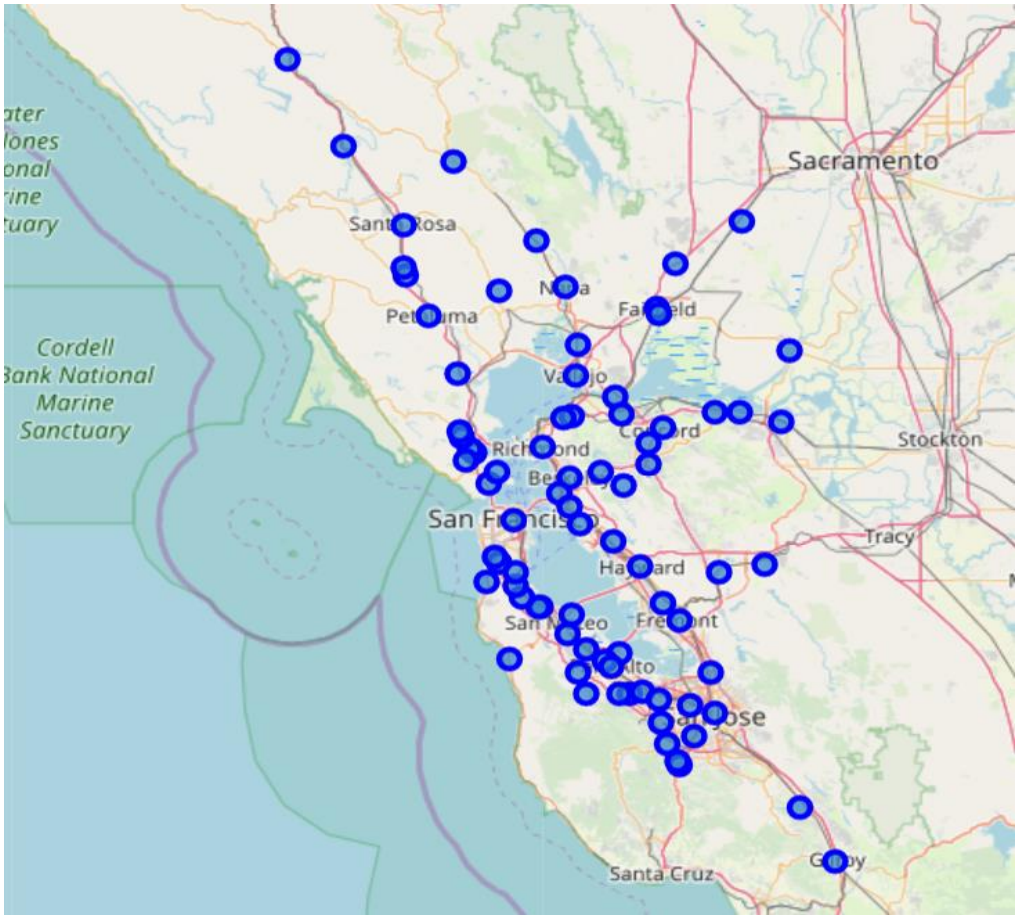
Now, to explore the neighborhood in the cities, we will be using **Foursquare API** to get the venue data for the respective neighborhood. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. We will get information on many categories of venues along with the one we are interested in i.e. restaurants. Other data science skills apart from web scrapping (Wikipedia) and working with API (Foursquare) will be Data Cleaning, Data Wrangling (*pandas* package), **Clustering (Machine Learning – *KMeans* algorithm)** and **Map Visualization (*Folium*)**.

Data Analysis:

Data was collected from the Wikipedia page 'List of cities and towns in San Francisco Bay Area' by Web Scrapping with the help of Python package 'BeautifulSoup4'. This was the list of 101 cities and towns stored in pandas DataFrame.

Now, after collecting the cities & towns in San Francisco Bay Area, we collected location details i.e. ***latitude and longitude information*** for each city and town. This was done with the help of another python package called ***Geocoders***. The same information was then added as new columns to DataFrame.

We located the cities and towns on maps with the help of Python library '***Folium***' and it looked like this:



Then, we explored the cities and towns with the help of Foursquare API. Foursquare is a technology company that built a massive dataset of location details and as of today it's API is most widely used by the developers (around 125,000) to get the numerous details regarding any particular location such as

searching hotels, exploring areas, finding travel destinations, etc. So, in this project we will be using this for exploring the venues for the listed cities and towns listed in DataFrame created.

We will gather Venue Name, Venue Location (i.e. latitude and longitude) and Venue Category from this. Category will let us know what kind of places are nearby. This information will be again added in DataFrame where we had information of Cities and Towns location to see how many categories are present in each area and how many unique categories are present.

So, we got around **324 unique venue categories** in mentioned cities and town. Out of these venues we have sorted down the top 10 venues visited by the people and the some of them are mentioned in below set:

	City / Town	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alameda	Chinese Restaurant	Bubble Tea Shop	Mexican Restaurant	Bar	Sandwich Place
1	Albany	Café	Coffee Shop	Park	Pub	Hotel
2	American Canyon	Chinese Restaurant	Indian Restaurant	Spa	Shopping Mall	Mexican Restaurant
3	Antioch	Train Station	Construction & Landscaping	Bakery	Park	Café
4	Atherton	Supermarket	Pub	Roller Rink	Soccer Field	Sandwich Place

From above table, we can say that there are many types of cafes, restaurants, bar, pub and pizza places as most common venues for the people in Bay Area. This makes a tough competition for any entrepreneur to open any kind of restaurant.

The best location for opening a restaurant will be the location having a smaller number of restaurant and particularly, Indian cafes and restaurant. It will also be beneficial to have restaurants near Multiplexes, Shopping Malls, Supermarkets, Business Centers, Museums, etc. where we can expect a lot of crowd visiting from that city or town and Neighborhood.

From these venues, we also tried to collect the number of total restaurants and Indian restaurants in these areas and we got to know that there are only 39 Indian Restaurants out of 864 restaurants in overall San Francisco Bay Area.

To find the most promising area, we will segment the data in various clusters of top 10 common venues and to do the same, we will use Machine Learning technique 'KMeans'. This will be discussed in Methodology section below.

Methodology

In this section, we will discuss about the Machine Learning technique i.e. Clustering by KMeans to find similar neighborhoods in the cities and towns. KMeans Clustering is one of the most simple and popular unsupervised Machine Learning Technique. Here, k number of centroids are identified, and every data point is assigned to nearest cluster, while keeping the centroid as small as possible.

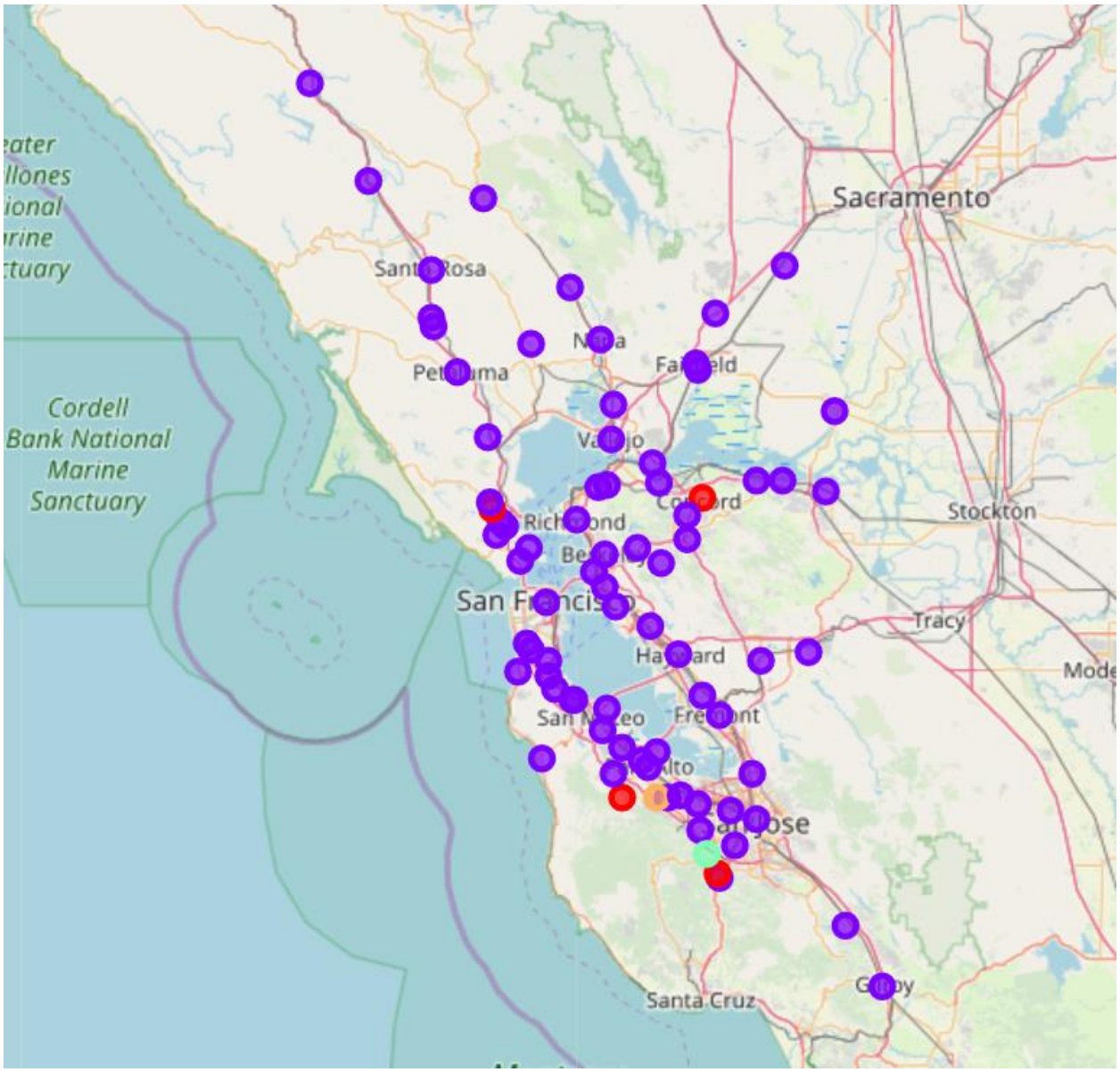
Here, we will create clusters using all the venue data instead of just concentrating on restaurants. This is because, we would like to consider all the other food joints like cafes, bars, pubs and the category with just word 'Restaurant' into consideration. Also, this will include the most common public places in the area which might impact the number of customers in any restaurant.

So, here we will create 5 clusters based on top 10 common venues visited in Bay Area and figure out the area with smaller number of restaurants and try to consider the location does have public places in top 10 common venues.

Let's discuss the result of this clustering in following Result section.

Results & Discussion

We have got 5 clusters from our method and the same are shown in the map below:



From above map, we study the results from the clusters and got the following understanding:

- **Cluster 0:**
  This cluster consist of five area with many diners, restaurants, cafes, etc., so it might be little competitive area for opening an Indian Restaurant.

- **Cluster 1:**
  This cluster consist of 91 area with variety of restaurants, cafes, and many more food joints. So, these areas are not at all recommended for opening a new Indian Restaurant.

- **Cluster 2:**
  This cluster consists of only one area i.e. **San Mateo,** which has 2 restaurants and 1 snack center, which makes it one of the preferred locations for opening an Indian Restaurant.

- **Cluster 3:**
  This cluster consists of only one are i.e. **Saratoga**, which has only 2 restaurants, making it another preferred location for opening a new Indian Restaurant.

- **Cluster 4:**
  This cluster also like Cluster 3 and 4 i.e. with one location **Los Altos Hills** with 1 restaurant, which makes it another preferred location to open Indian Restaurant.

Limitations and Suggestions for Future Research

In this project, we have considered the common venues into consideration while creating clusters. There are many other factors such as kind of population (how many people prefer Indian food), income of residents, rent or land rate of the area, etc. There factors can be considered in future project with the help of other data source and Machine Learning Techniques to recommend perfect location.

Conclusion

Purpose of this project was to identify places in San Francisco Bay Area with low number of restaurants (particularly Indian restaurants) in order to provide the optimal location for a new Indian restaurant.

Most of the restaurants are present with highest number in Cluster 2 and with moderate number in Cluster 1, so theses will be very competitive area for opening new restaurant of any cuisine. While rest of the clusters i.e. 3, 4 and 5 has very a smaller number of restaurants and a good number of public places. This represents a great opportunity and high potential areas to open new Indian Restaurant. This cluster areas are San Mateo, Saratoga and Las Altos Hills.

Final decision on optimal restaurant location will be made by entrepreneurs based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

References

- List of Cities and Towns in San Francisco Bay Area: https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area

- Foursquare Developer Documentation: https://developer.foursquare.com/docs