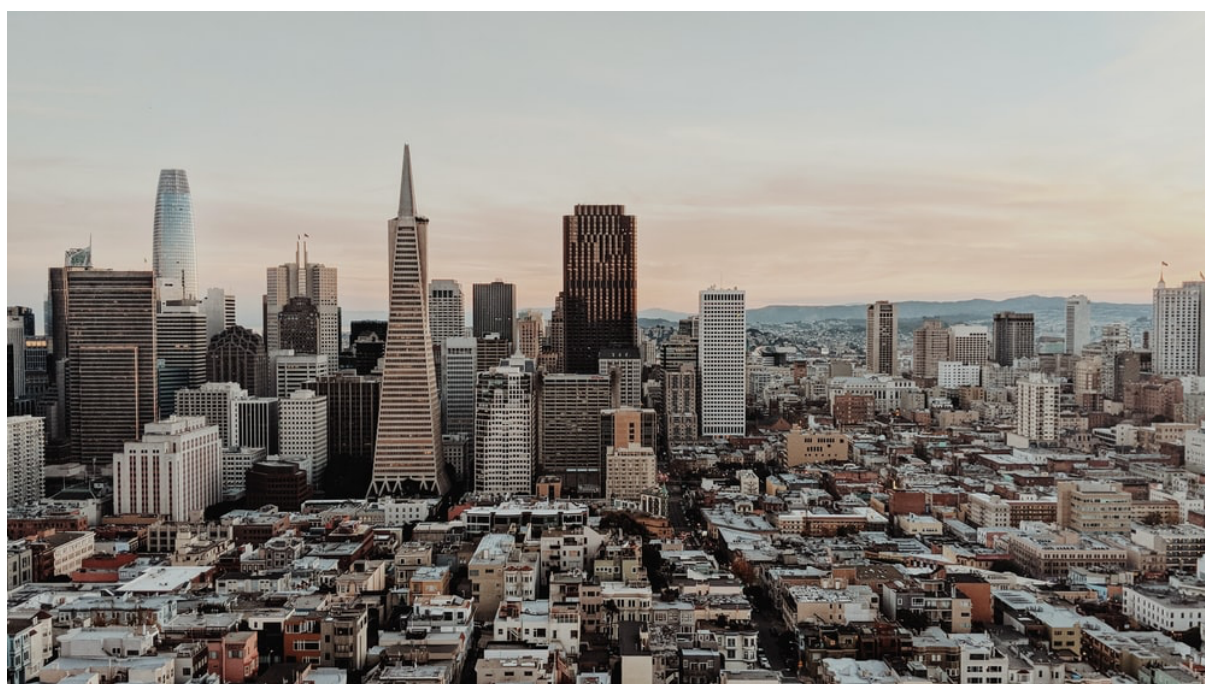


IBM/Coursera Data Science Capstone Project

Opening an Indian Restaurant in San Francisco

by Yash Kamat

September 2020



Introduction

The city of San Francisco is one of the biggest and most diverse cities in the United States. It is a melting pot of people from different backgrounds, nationalities and cultures. The result of this cosmopolitan nature of the city is the wide variety of cuisines available in the restaurants of SF. From popular American and European cuisines to the exotic foods of Asia and Africa, San Francisco boasts a great variety of options to dine.

Of these, Indian food is one of the most popular choices, largely due to the growing South Asian population in the city. In fact, the San Francisco-Oakland-Hayward area contained the fourth-largest population of Asian Indians in United States, according to the 2010 Census data (ProximityOne). The growing Indian population in the San Francisco metropolitan area means that the demand for Indian food is increasing, which is a great opportunity for entrepreneurs and restaurateurs to open an Indian restaurant in the city.

Business Problem

The main question of my Capstone Project is quite simple - **if someone were to open an Indian restaurant in the city of San Francisco, where would should they open it?** In order for the restaurant to be successful, there have to be enough customers, while also being in a place without too many competitors in the neighborhood.

The primary audience of this study is meant to be entrepreneurs and restaurateurs in San Francisco. If they are new to the market, they would be interested in this study to scout potential locations to start their venture in. Or, if they already have existing establishments, to seek new locations to expand into.

Data

Data Requirements

To explore the business problem, we will need the following information:

- List of neighborhoods in San Francisco. This defines the scope of the project which is confined to the city of San Francisco, located in Silicon Valley in California on the west coast of the USA.
- Latitude and longitude of the neighborhoods of San Francisco. This is to geolocate points of interest (especially Indian restaurant) and also to plot the map in the Jupyter notebook.
- Geolocation and venue data, specifically for Indian restaurants in San Francisco. This data will then be used for clustering and its inferences.

Data Sources

As a starting point we will use the Wikipedia page which contains the list of neighbourhoods in San Francisco (https://en.wikipedia.org/wiki/Category:Neighborhoods_in_San_Francisco). This page contains a list of neighbourhoods in San Francisco, with over 90 distinct neighbourhoods listed. Using a Python-based web scraper, we will extract the list of neighbourhoods from the Wikipedia page. Based on this list, we will then get the individual geographical coordinates of the neighbourhoods using a Python package which will give us the latitude and longitude for each neighbourhood.

After that, we will use the Foursquare API to get the geolocation data for those neighbourhoods. Foursquare has a database of 100+ million places and is used by developers all around the world. While the Foursquare API provides multiple categories of venue data, we are mostly interested in the Indian restaurant category for the purpose of this business problem.