

Coursera IBM Data Science Professional
Applied Data Science Capstone

**Clustering Analysis For Restaurant Location in San
Francisco**

By Jatin Selmokar

January 2020

Introduction/Business Problem

Opening a restaurant in a densely populated city is always challenging and often requires understanding of the current food preferences, location, competition, and the capital investment associated with it.

San Francisco being one of the most populated cities in the US has plethora of restaurants offering different cuisines. Besides offering great food at a lower cost, there are external factors that define restaurant's success. Thus, in order to establish a prosperous business model, it is imperative for a business owner to understand and gauge the restaurant business in and around San Francisco.

Our main objective in this capstone project is to guide the business client in choosing the perfect location to open a restaurant. This project aims to analyze and provide insights on restaurant businesses around SF neighborhoods using data science so that the business owner can make an informed decision.

Data

For the data analysis in this project, we will need the following data

- San Francisco neighborhood locations
- Geo-coordinates for the SF neighborhoods.
- Foursquare venue data

Data Sources And Methods To Extract Them

The San Francisco neighborhood data is extracted from the Wikipedia page (http://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco) using Beautiful Soup package. Once extracted, the geo-coordinates(latitudes & longitudes) are populated using Google Maps API for the neighborhoods. This geo data will be primarily used for visualization and in getting different venues around the neighborhoods using Foursquare API. Our focus will be on the restaurant category data download from foursquare API which will be then fed to clustering model.