Capstone Project - The Battle of the Neighbourhoods (Week 1)

Introduction: Business Problem

An investor is looking to open a new **Indian Style** restaurant in **San Francisco**, but he is not sure about the best location for his new venue. So, you got a call from him asking your input to help him choose the location. San Francisco is a very busy city, best known for tourist attractions and business innovation. Strolling around the city blocks, it is pretty easy to notice that the city already has a lot of restaurants in town. How should we proceed and decide the location?

In The Art of War, Sun Tzu has said: "Know the enemy and know yourself; in a hundred battles you will never be in peril." Following this line of thinking, the basic strategy here is then to know what are the most critical factors that contribute to the restaurant's profitability. According to a report https://www.foodnewsfeed.com/fsr/vendor-bylines/8-factors-choosing-new-restaurant-location by Tom Larkin published in the FSR magazine, these components stand out as the most important ones: visibility, parking, space size, crime rates, surrounding businesses and competitor analysis, accessibility, affordability, and safety. Using public datasets, we could actually address some of these considerations pretty straightforwardly.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighbourhood (any type of restaurant)
- number of and distance to Indian Style restaurants in the neighbourhood, if any
- distance of neighbourhood from city center

We decided to use regularly spaced grid of locations, cantered around city center, to define our neighbourhoods.

Following data sources will be needed to extract/generate the required information:

- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**
- coordinates of San Francisco center will be obtained using **Google Maps API geocoding**

The city of San Francisco has maintained a large data repository hosted on the website DataSF (https://datasf.org/). From there, we can access the city's crime rate, housing price, and public parking information. Sounds good! So, let's proceed and see what insight can be extracted from the data. We will be working with two datasets:

Police Department Incident Reports: 2018 to Present

https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

and

Assessor_Historical_Secured_Property_Tax_Rolls

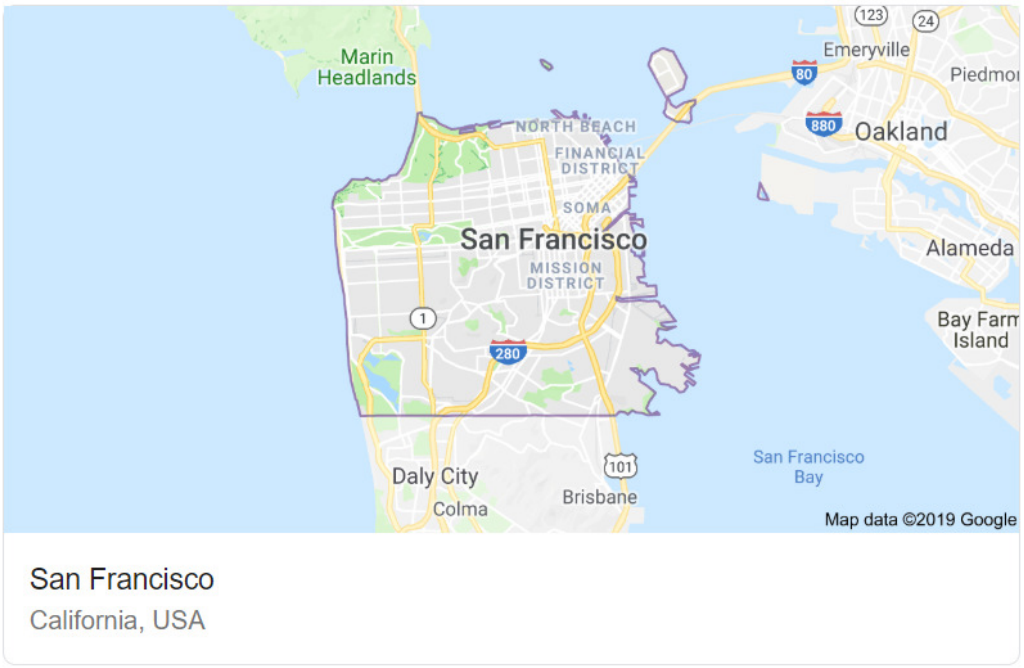
https://data.sfgov.org/Housing-and-Buildings/Assessor-Historical-Secured-Property-Tax-Rolls/wv5m-vpq2

The **Neighbourhood details of San Francisco** is available in this website

https://data.sfgov.org/Geographic-Locations-and-Boundaries/SF-Find-Neighborhoods/pty2-tcw4

Since these data files are in huge size, we cannot refer them as URL. We need to download them to our local system and should refer the file location in our program. Some of the files are in SHAPE files format and some other files are in CSV format. We are using both types of files in our project.

The Map of San Francisco:



San Francisco
California, USA