The Battle of Neighborhoods – Capstone Project

(Identifying Suitable Location for your Business Expansion)

1. Introduction:

1.1 Background:

Your friend is running coffee shop business for past few years in different states of U.S. As a company, they felt that itss time to expand their business. They selected for e.g., Los Angeles, CA city as the location of expansion. Their business model only cares about different neighborhoods in this city which are less competitive in nature (assumption). This type of business model worked for them in all their previous scenarios. They thought data can provide solution to their problem. As an experienced data scientist, they approached you for this task.

1.2 Business Problem:

The primary business problem you as a data scientist needs to solve is:

Given a county/city, you have to look for different neighborhoods in this county/city with comparatively small number of coffee shops in them and recommend these neighborhoods to the company team. This helps the company to expand their presence accordingly to their business model.

2. Data:

2.1 Data Source:

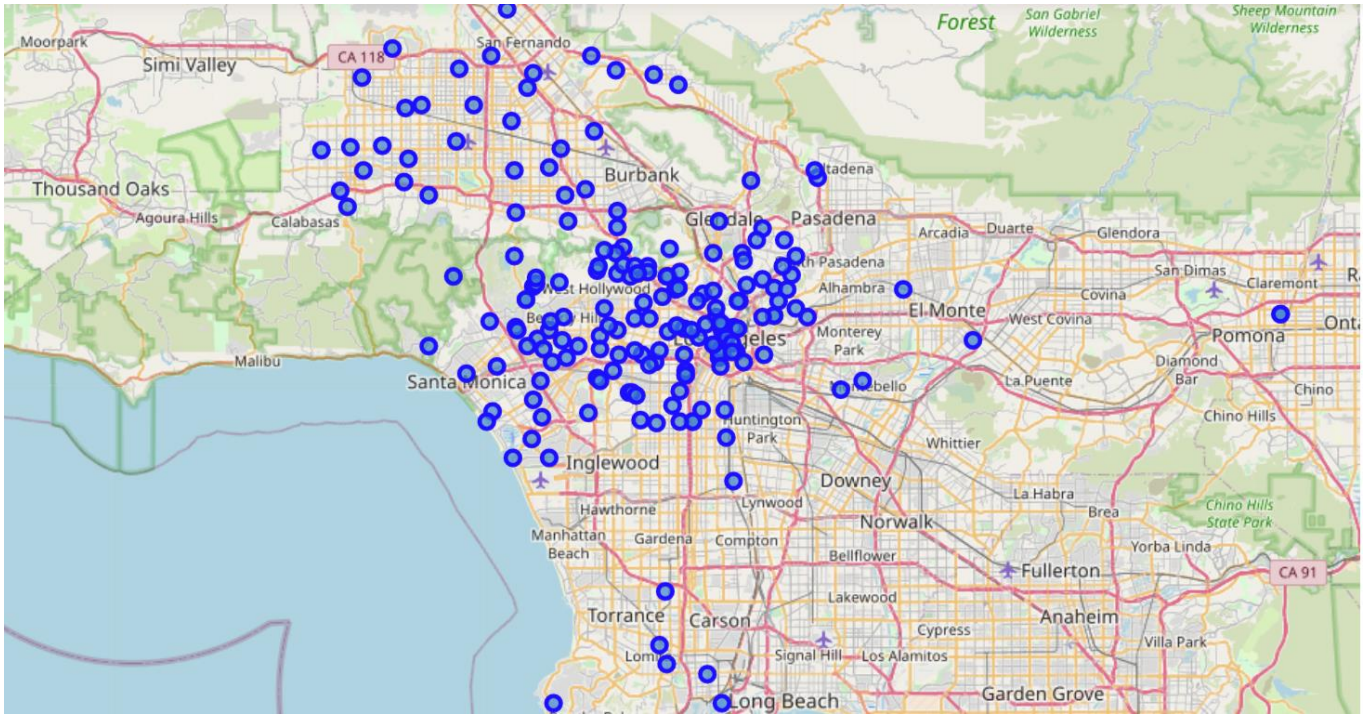
All the neighborhoods that are located in *Los Angeles, CA* are taken from a Wikipedia page *[1]*. This page has total of 200 neighborhoods. We use these neighborhoods as a starting point for our project. As all this data only has names of the neighborhoods, to continue we need to extract more information.

2.2 Data Extraction:

Firstly, all the neighborhood locations in the page can be extracted using Python's BeautifulSoup library. For each neighborhood, we extract its latitude and longitude values using geocoder library. An initial dataframe is created with 'Neighborhood', 'Latitude' and 'Longitude' as its columns. The top 5 rows of dataframe would look like this:

	Neighborhood	Latitude	Longitude
0	Angelino Heights	34.070290	-118.254800
1	Angeles Mesa	2.421100	-76.917380
2	Angelus Vista	34.087575	-118.267156
3	Arleta	34.249050	-118.433490
4	Arlington Heights	34.039890	-118.325160

To visualize geographic details of above neighborhoods on map, we use folium library in Python. I created a map of Los Angeles, CA using its latitude and longitude values. Then, I added markers to this map for each neighborhood location from the dataframe using its latitude and longitude values.



Secondly, we use FourSquare API to extract nearby venues data for each neighborhood in the initial dataframe. To elaborate on FourSquare API, it explores the neighborhood by taking its name, latitude and longitude information along with the user credentials like Client ID and Access Token in extracting the venues that are nearby to the given neighborhood. It is mainly used to access the venues information like venue name, venue location (both latitude and longitude of the venue) and venue category. All this data is combined to form a new dataframe with columns as 'Neighborhood', 'Latitude', 'Longitude', 'Venue', 'Venue Latitude', 'Venue Longitude' and 'Venue Category'. After all this extraction, the tail of the dataframe would look like below:

	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4542	Yucca Corridor	34.10392	-118.33	Hollywood Burger	34.100978	-118.325924	American Restaurant
4543	Yucca Corridor	34.10392	-118.33	Dream Hollywood	34.099879	-118.330173	Hotel
4544	Yucca Corridor	34.10392	-118.33	Trejo's Cantina	34.099513	-118.329077	Mexican Restaurant
4545	Yucca Corridor	34.10392	-118.33	Mamas Shelter Restaurant	34.099590	-118.331391	Lounge
4546	Yucca Corridor	34.10392	-118.33	Wood & Vine	34.101533	-118.326315	American Restaurant

There are total of 4547 venues across all the neighborhoods which are categorized to 363 unique venue categories.

References:

1. https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles