Capstone Project

Opening a new Restaurant or Bar in Los Angeles

Introduction:

Los Angeles or L.A. is the largest city in California and the second largest city in United States. Los Angeles is known for its Mediterranean climate, ethnic and cultural diversity, Hollywood industry, and its sprawling metropolitan area. Personally, I will be moving there this fall to pursue my higher education. As a result, I wanted to know about this city and where are some of the famous restaurants and bars located. I was curious to know where can one open a new restaurant or a bar. Some neighbourhoods in L.A have amazing nightlife and other activities, some have famous and delicious restaurants whereas some areas have a number of food trucks which serve delicious food as well. As a result, being a foodie myself, I decided to check out various neighbourhoods where one can open a restaurant or a bar. With the help of Foursquare app, we can get a lot of information, which will make our job easier.

Business Problem:

The main objective of this capstone project is to find out which areas are better suited and will be more profitable to open a new restaurant or a bar. This project can be useful for business owners and entrepreneurs as well who are looking to invest in a restaurant in Los Angeles. The main objective of this project is to carefully analyse appropriate data and find recommendations for the stakeholders.

Data:

The data required for this project has been collected from multiple sources. A summary of the data required for this project is given below.

Neighbourhood Data:

The data of the neighbourhoods of Los Angeles was scrapped from https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles . The data is read into a pandas data frame using the Beautifulsoup library. The main reason for doing so is that the Wikipedia page provides a comprehensive and detailed table of the data which can easily be scraped using the Beautifulsoup library of python.
For geolocation data, we will use the Geocoding API.

Venue Data:

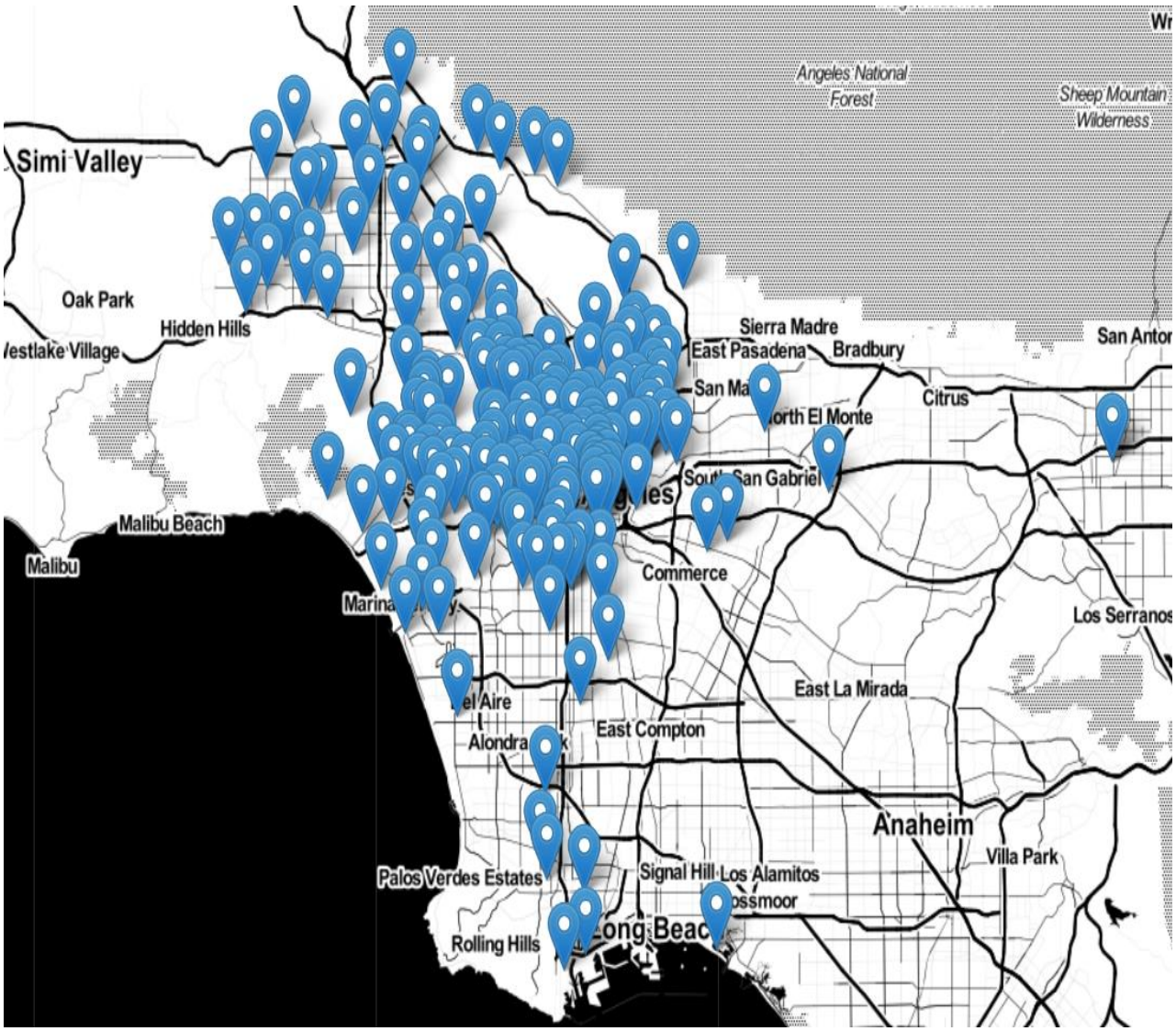
The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighbourhoods in Los Angeles and is used to study the popular venues of different neighbourhoods.

Methodology:

We get the list of neighbourhoods in LA by scraping the data from https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles which is a Wikipedia page about neighbourhoods in LA. Using Geocoding API, we get the latitudes and longitudes of the location. We parse the data from the Wikipedia page using Beautifulsoup library of Python. There are total 200 neighbourhoods in Los Angeles.

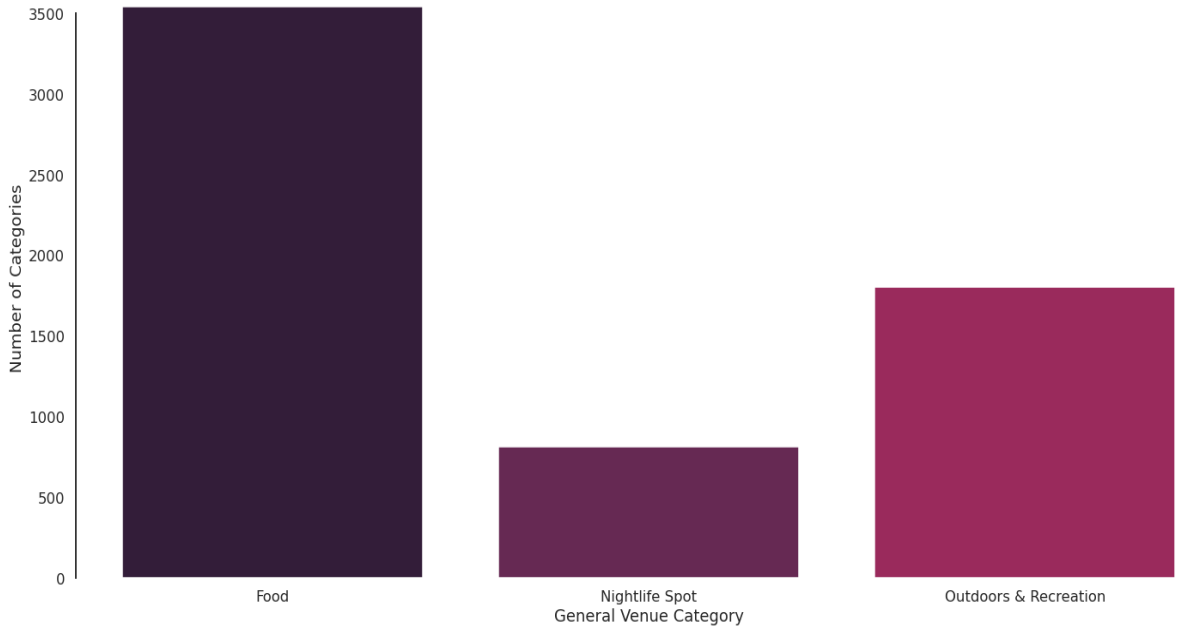
	Neighborhood	Latitude	Longitude
0	Angelino Heights	34.070290	-118.254800
1	Angeles Mesa	32.764074	-116.986171
2	Angelus Vista	34.087575	-118.267156
3	Arleta	34.249050	-118.433490
4	Arlington Heights	34.039890	-118.325160
5	Arts District	34.041952	-118.236385
6	Atwater Village	34.119700	-118.258870
7	Baldwin Hills	34.021570	-118.367650
8	Baldwin Hills/Crenshaw	34.010428	-118.336776
9	Baldwin Village	34.070445	-118.200710
10	Baldwin Vista	34.070445	-118.200710
11	Beachwood Canyon	34.109150	-118.320450

Using Folium, we can plot these neighbourhoods on a map.

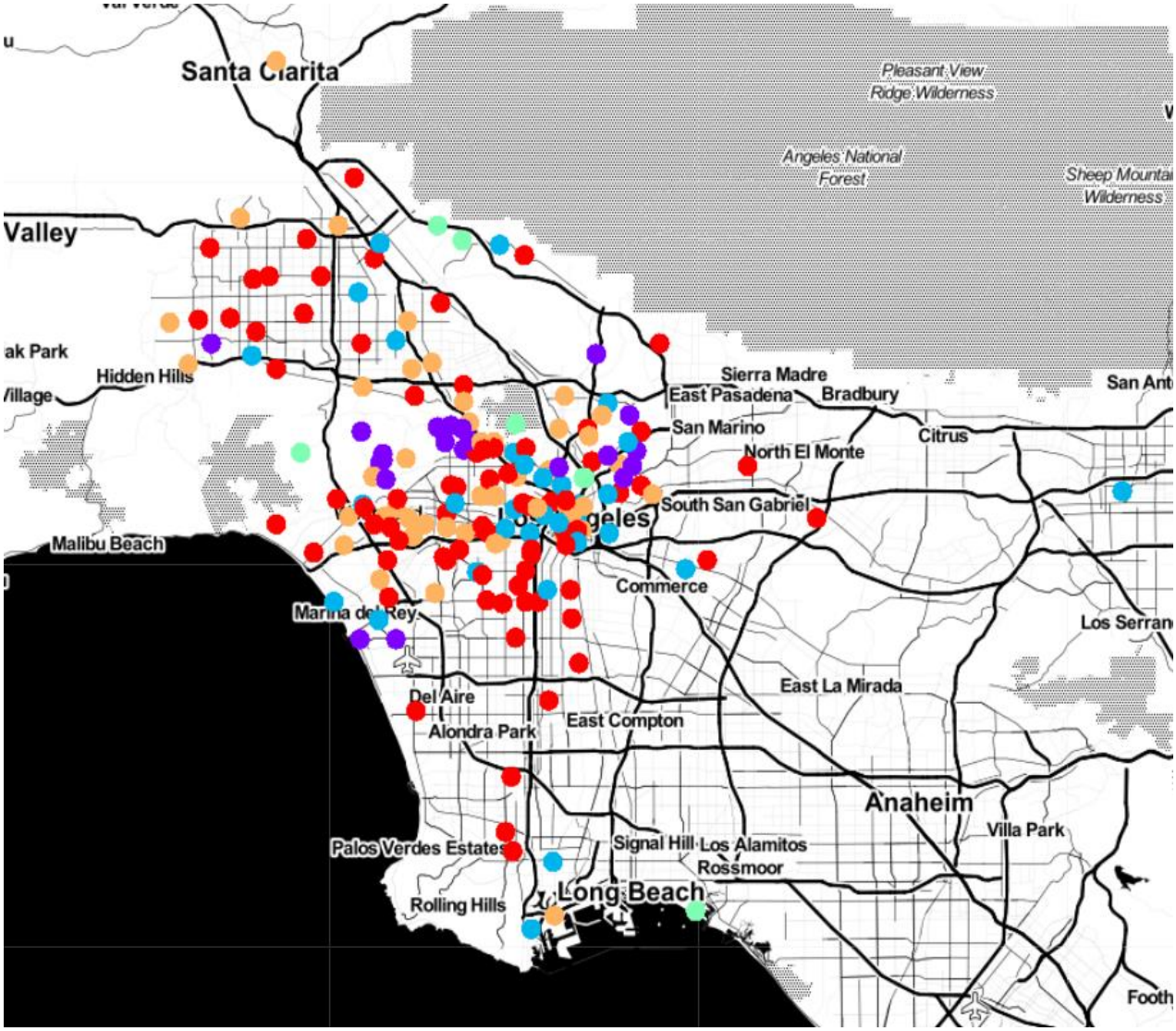


We then use the Foursquare API to get the data of venues. These venues are divided into many categories such as Arts and Entertainments, Travel and Transport, Professional and other places etc. Since we are looking to open a restaurant or a bar, we are only going to select three categories – Food, Outdoor and Recreation, and Nightlife Spot.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	General Venue Category
4	Angelino Heights	34.07029	-118.2548	Rico's Tacos	34.070252	-118.250261	Food Truck	Food
5	Angelino Heights	34.07029	-118.2548	Ridge Wy	34.071826	-118.253959	Trail	Outdoors & Recreation
7	Angelino Heights	34.07029	-118.2548	Bar Henry	34.069062	-118.250465	Cocktail Bar	Nightlife Spot
8	Angelino Heights	34.07029	-118.2548	Tacos Los Primos	34.066974	-118.252916	Food Truck	Food
9	Angelino Heights	34.07029	-118.2548	만수	34.057517	-118.304496	Korean Restaurant	Food
11	Angelino Heights	34.07029	-118.2548	Thunderbolt	34.066298	-118.253917	Restaurant	Food
14	Angelino Heights	34.07029	-118.2548	W. Edgeware Rd.	34.071707	-118.253693	Trail	Outdoors & Recreation
15	Angelino Heights	34.07029	-118.2548	El Toritos	34.068463	-118.257507	Cocktail Bar	Nightlife Spot
16	Angelino Heights	34.07029	-118.2548	THE D3EP HAU$	34.069892	-118.257090	Lounge	Nightlife Spot
18	Angelino Heights	34.07029	-118.2548	Tacos La Movidita (aka Bellevue Steakhouse)	34.069535	-118.258367	Steakhouse	Food



As we can see, there are total 3545 venues in Food category, 1408 in Outdoors and Recreation category and 815 in Nightlife spot category. After clustering these venues into 5 clusters using k-means and plotting them on the map, we get the following map:



Results and Discussions:

- All venues are grouped into five clusters
- Cluster 1 has the most number of neighbourhoods (82) and Cluster 4 has the least number of neighbourhoods (6).
- “Food” venue category is the most popular category in Cluster 1,3 and 5 whereas “Outdoor and Recreation” is the most popular category in Cluster 2 and 4.
- Food seems to be the most popular category followed by Outdoor and Recreation.

Conclusions:

- In this project, we analysed the neighbourhoods in Los Angeles. We get the list of neighbourhoods in LA by scraping the data from the Wikipedia page about neighbourhoods in LA. Using Geocoding API, we get the latitudes and longitudes of the location and Folium helps us to plot these neighbourhoods on a map
- We then used the Foursquare API to get the data of venues and divided them into five clusters. Finally, we plotted them on the map.