Exploring Restaurants in Los Angeles

1. Introduction

Restaurant business is one of the most sought-after businesses in Los Angeles (LA). LA, in particular, is an amazing place to dine out owing to its wide variety of international cuisines, the quality of food and the services offered. Given the high demand for quality, a restaurant needs to get several factors right in order to survive here. In this brief report, I tried to investigate some of these factors by exploring the restaurants based in the various neighborhoods of Los Angeles (LA). I have tried to keep the discussion general when it comes to the kinds of restaurant and their features and rather focus on the global features like density of restaurants in a region, their frequency and so on.

The *business* direction that I have in mind is actually two-fold. One is directed towards individuals looking for places to eat or even people looking for places to rent based on restaurant types or frequency. The other direction is for corporations/individuals looking to open a new restaurant.

1.1 Specific Plan

I plan to use 2 ideas for segregating/clustering the neighborhoods based on their restaurant venues. One idea is to find the density of restaurants in each neighborhood. This will help anyone trying to open a new restaurant by either avoiding overcrowded areas or alternately could help finding popular venues to avoid competition.
The next idea to explore is the kind of restaurant. Here, one can again use clustering but now based on the frequency of occurrence of each restaurant in a Neighborhood just like the one done in the New York data

set. Once we find the relevant cluster here, we can then look for its intersection with the clusters found above to fine tune the relevant neighborhood where one wants to open his/her restaurant.

2. Data collection and cleaning

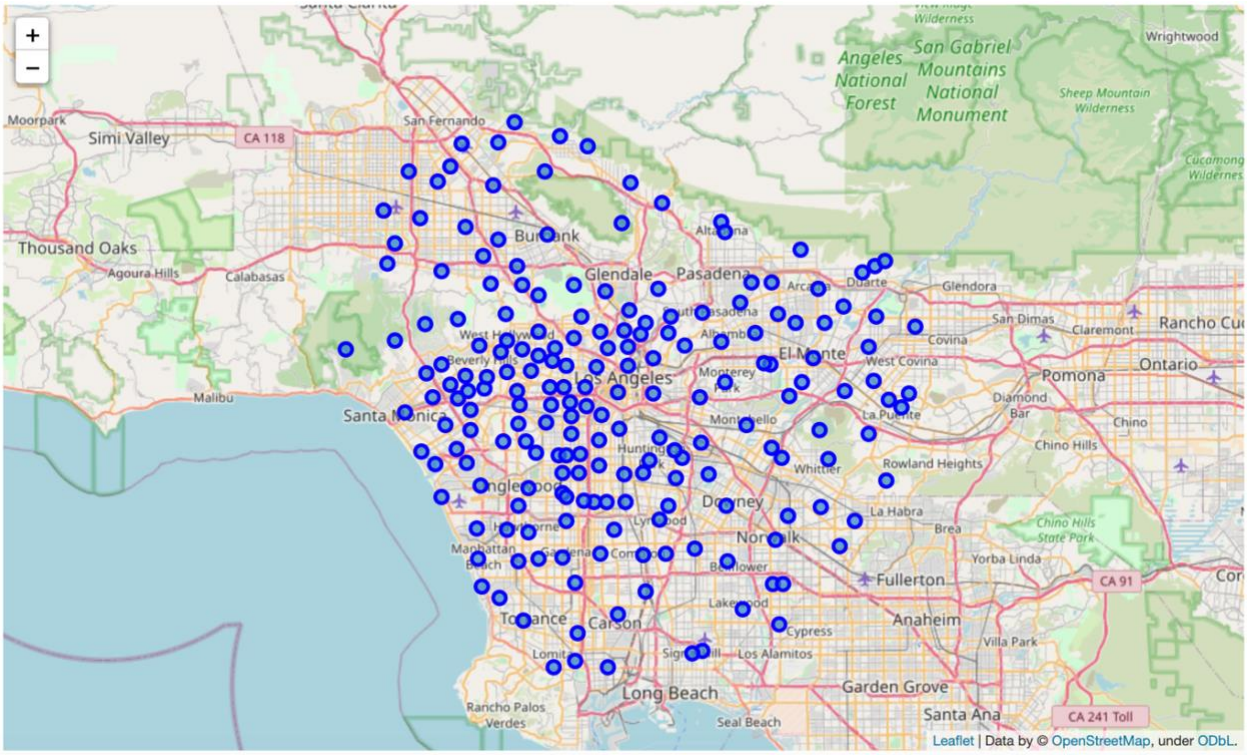
I acquired the location data from this website. There is an API endpoint link for the json file in this website as well.

2.1 Data cleaning/wrangling

Data cleaning was pretty straightforward. Lot of redundant columns were dropped. Some columns like the 'type' or the geometry of the boundaries ('the_geom') of a neighborhood were also removed. One major problem which I came to realize after plotting the folium map is that latitude and longitude values are swapped with each other. Once that is fixed along with renaming of some columns, the final table looked as shown below:

	Neighbourhood	sqmi	Longitude	Latitude
0	Acton	39.3391089485	-118.16981019229348	34.497355239240846
1	Adams-Normandie	0.805350187789	-118.30020800000011	34.031461499124156
2	Agoura Hills	8.14676029818	-118.75988450000015	34.146736499122795
3	Agua Dulce	31.4626319451	-118.3171036690717	34.504926999796837
4	Alhambra	7.62381430605	-118.13651200000021	34.085538999123571

The data consisted of 272 unique neighborhoods. To make the analysis slightly easier, I used a distance function from the LA central coordinates to reduce the number of rows to 199. This is just so that I can reduce the number of calls to the foursquare location app. The final map with all the neighborhoods is shown below –


Angeles National Forest
San Gabriel Mountains National Monument
Wrightwood
Sheep Mountain Wilderness
Moorpark
Simi Valley
CA 118
San Fernando
Thousand Oaks
Agoura Hills
Calabasas
Burbank
Glendale
Pasadena
Glendora
San Dimas
Claremont
Covina
West Covina
El Monte
Ontario
Pomona
Diamond Bar
Chino
Chino Hills
Malibu
Santa Monica
Los Angeles
Rowland Heights
Whittier
Downey
Norwalk
La Habra
Brea
Yorba Linda
Fullerton
CA 91
Anaheim
Villa Park
Torrance
Carson
Cypress
Los Alamitos
Long Beach
Garden Grove
Santa Ana
Seal Beach
Rancho Palos Verdes
CA 241 Toll
Leaflet | Data by © OpenStreetMap, under ODbL.