Categorical Analysis of Restaurants in LA Downtown and their Online Delivery System

Roopashree Bhat
November 8th, 2019

1. Introduction

1.1. Background

Los Angeles is the most populous city of California, second populous city in the entire United states of America, after the New York city. Los Angeles is the cultural, financial and commercial center of southern California. Its known for its ethnic diversity and sprawling metropolis. Due to its ethnic diversity Los Angeles is known for diverse cuisines from all over the world. Though there are tons of restaurants to taste around, an increase in traffic and population in Los Angeles suburb has increased demand for online food delivery providers and apps.

By analyzing all restaurants around Los Angeles suburbs and their major delivery providers we can get valuable information about the current competitors in the region as well as those restaurants which currently don't have any delivery option and might need such applications in future. We can also study the major cuisines servicing the area currently.

1.2. Problem

Due to increasing population, traffic and drastic increase in demand for online food delivery both commercial as well as residential areas, the focus of this project is:

a) To analyze all restaurant venues and extract their major delivery providers if any, to get information on major competitor and at the same time to focus on those restaurants which don't have any delivery providers and might be potential customers for any new delivery providers in future

b) To analyze and categorize different cuisines which might provide valuable input for online delivery providers and to new startups planning to open new restaurants

1.3. Interest

This project will be of high interest to the latest trend of online delivery apps and providers who need in depth information of various restaurants and major competitors in the area.

It will also be beneficial to startups who want to start any new restaurants in the area to get information regarding which major cuisines are there in the area. It could give valuable input on which cuisines are not present in a particular neighborhood and starting those cuisines in the areas might prove highly profitable as Los Angeles is a city of diversity and people tend to look for diverse cuisines.

2. Data Source and Acquisition

2.1. Data Source

Geo spatial data of Los Angeles county was provided by datasets of LA county in this link Geo Spatial datasets of LA. From this link I took the latest Geojson file corresponding to neighborhoods of LA.

I have also referred to https://en.wikipedia.org/wiki/Los_Angeles#Demographics to analyze economy, cultural diversity, population distribution within LA. Since LA county has a large area distribution, I have used the above information to further filter and restrict the data to certain neighborhoods within central downtown region for my project.

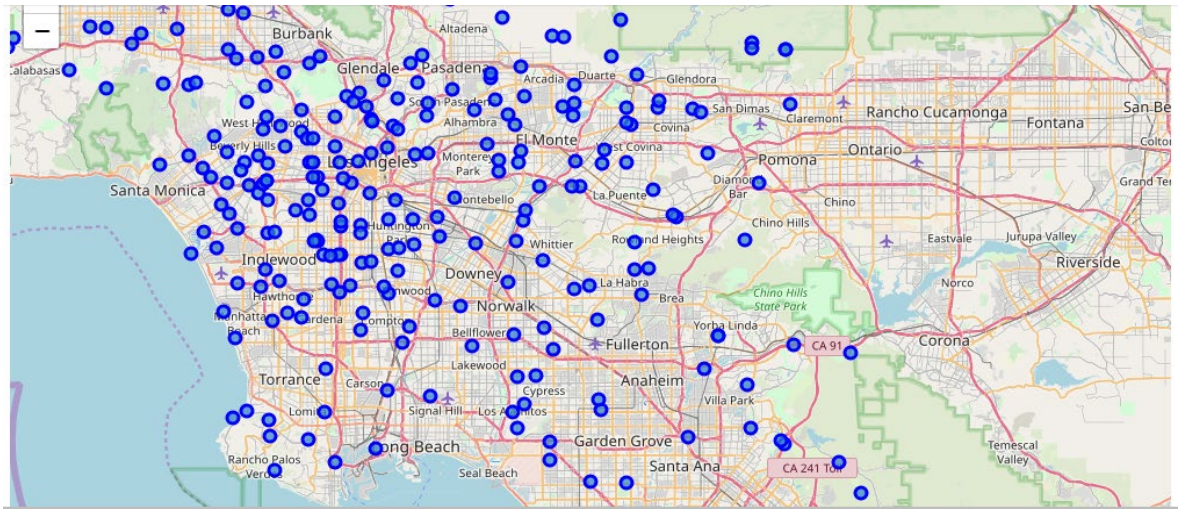
2.2. Data Acquisition and Cleaning

Using JSON library of Python, I have extracted the GeoJson file. Out of all the location features, regions of LA county, their corresponding Neighborhoods and their coordinates were extracted.

The coordinates given in the JSON file were for every Square Kilometer [km²] of neighborhood. Hence, I have taken the mean Latitude and Longitude for every neighborhood while converting it to a data frame. Hence the initial data frame consists of regions, neighborhood and coordinate information. Sample records of dataframe extracted is as below:

	Region	Neighborhood	Latitude	Longitude
0	antelope-valley	Acton	34.539023	-118.207034
1	south-la	Adams-Normandie	34.037396	-118.308002
2	santa-monica-mountains	Agoura Hills	34.168157	-118.776212
3	northwest-county	Agua Dulce	34.488109	-118.378224
4	san-gabriel-valley	Alhambra	34.10504	-118.121747

I used the Folium map to create a map of LA county. For this I took coordinates using Geopy Geolocator library. Folium map was created to show different regions and neighborhoods of LA county. From this map I could visualize that downtown region falls within central LA region of dataset extracted. Hence, only central-LA region was selected for all future analysis and processing.



The subset dataframe corresponding to central-la region is shown below. It has 26 neighborhoods in total.

	Region	Neighborhood	Latitude	Longitude
0	central-la	Arlington Heights	34.052504	-118.31672
1	central-la	Beverly Grove	34.085755	-118.37249
2	central-la	Carthay	34.051684	-118.36664
3	central-la	Chinatown	34.067573	-118.22472
4	central-la	Downtown	34.061247	-118.24462
5	central-la	East Hollywood	34.098356	-118.28739
6	central-la	Echo Park	34.095686	-118.24486
7	central-la	Elysian Park	34.093903	-118.24245
8	central-la	Elysian Valley	34.108424	-118.24973
9	central-la	Fairfax	34.088949	-118.35112
10	central-la	Griffith Park	34.15884	-118.30434
11	central-la	Hancock Park	34.083488	-118.3266
12	central-la	Harvard Heights	34.05282	-118.31329
13	central-la	Hollywood Hills	34.152118	-118.31782
14	central-la	Hollywood Hills West	34.1292	-118.36951
15	central-la	Hollywood	34.105244	-118.32686
16	central-la	Koreatown	34.068982	-118.2868
17	central-la	Larchmont	34.076204	-118.32062

18	central-la	Los Feliz	34.118408	-118.27321
19	central-la	Mid-City	34.034962	-118.36859
20	central-la	Mid-Wilshire	34.068991	-118.34619
21	central-la	Pico-Union	34.052497	-118.28145
22	central-la	Silver Lake	34.112812	-118.26483
23	central-la	West Hollywood	34.097724	-118.36817
24	central-la	Westlake	34.053889	-118.25907
25	central-la	Windsor Square	34.076231	-118.31359

This dataframe was further filtered and foursqaure API was used to get restaurant venues, cuisine information and delivery providers. The final dataframe was then categorized, clustered and modelled. In depth analysis is explained in methodology section.