Battle of Neighborhoods in LA

Introduction

In this capstone project, I will analyze districts in the city of Los Angeles: I will group similar districts in terms of their venue distribution. By doing so, a future entrepreneur should be able to tell most favorable locations in launching his or her business. For example, when a businessman desires to open up a cafe, it would be more beneficial if he does so in the urban cities than rural areas. Hence, the essential problem of this project is an entrepreneur is trying to launch a business(Tea Room) in LA—but he or she does not know where to open since each district has differentiated venues. Furthermore, the **target audience for this analysis are** entrepreneurs who are trying to launch a business, in this case, Tea Room. But this filtering method itself can be applied to anyone willing to open a business in LA

Data

The data used for the analysis was extracted from "opendatasoft.com .“ It contains coordinates based on different zip codes in Los Angeles. zip codes and the names of the neighborhood were matched manually via Excel using the data from the website “lacounty.gov.” Thereafter, I will drop the rows with missing values, eventually analyzing 93 neighborhoods in LA.

```
In [10]: # drop rows with any missing values
         df = df.dropna()
         df.head()
```

Out[10]:

	Zip	City	State	Latitude	Longitude	DistrictName
0	90001	Los Angeles	CA	33.972914	-118.248780	Acton
1	90002	Los Angeles	CA	33.948315	-118.248450	Agoura Hills
2	90003	Los Angeles	CA	33.962714	-118.276000	Alhambra
3	90091	Los Angeles	CA	33.786594	-118.298662	Alhambra
4	90004	Los Angeles	CA	34.077110	-118.307550	Alondra Park

```
In [11]: df.shape
```

Out[11]: (93, 6)

Figure 1: This is the data dimension which I employed for analysis.

Data Analysis Process in Brief

The following is the step-by-step process of the data analysis.

I. extract data showing coordinates based on different of LA from "opendatasoft.com" later on the name of the neighborhoods were given manually from Excel.
II. based on the zip code coordinates, retrieve the necessary data—district, its name, latitude, longitude, and zip code— using FourSquare API
III. retrieve the district's neighborhood information (e.g. the number of restaurants and parks)
IV. lastly, sort by venue frequency and filter to remove unfavorable neighborhoods. In essence, the purpose of the analysis is to analyze different distributions of venues in districts of LA and suggest some ideas where to open up a new business
V. 93 neighborhoods wil be analyzed

Methodology

This section covers the detailed process of step 1 - 4 stated above.

I. **Retrieve necessary data from FourSquare**
   a. We retrieved about hundred venues within the radius of 500—and we further extracted categories of the venues. Having done so, I implemented One Hot encoding and analyzed the frequency of the venue categories.
II. **Check the frequency distribution by processing the data to return K top venues in each neighborhood.**
III. **Find the most favorable location to open a business by filtering the neighborhood's top venues**
   a. Such step is significant as a business cannot survive in excessively competitive working environment, with similar shops surrounding them.

Results

To contrive the results for the analysis, I need filter neighborhood that has similar types of business in top 5 venues to avoid possible competition. These similar or the same type of business include “Coffee Shop”, “Ice Cream Shop”, “Juice Bar”, “Café” and “Tea Room” Afterwards, most appropriate location(neighborhood) will be chosen with favorable surroundings as one of top popular venue. Having done so the locations reduced to 35 as shown below in *figure 2*. From here, I can decide which location is the best to open a "Tea Room" Business.

	Neighborhood	1st Popular Venues	2nd Popular Venues	3rd Popular Venues	4th Popular Venues	5th Popular Venues
0	Acton	Donut Shop	Mexican Restaurant	Pizza Place	Burger Joint	Pharmacy
1	Agoura Hills	Park	Women's Store	Event Space	Food Stand	Food Service
2	Alhambra	Fast Food Restaurant	Sandwich Place	Burger Joint	Mexican Restaurant	Pizza Place
5	Arcadia	Sandwich Place	Mexican Restaurant	Fast Food Restaurant	Food Truck	Pizza Place
7	Avalon	Sandwich Place	Fast Food Restaurant	Mexican Restaurant	Southern / Soul Food Restaurant	Department Store
10	Baldwin Park	Fast Food Restaurant	Pizza Place	Fried Chicken Joint	Restaurant	Diner
11	Bell	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Bakery	Sandwich Place
12	Bell Canyon	Japanese Restaurant	Sushi Restaurant	Bar	Bubble Tea Shop	Pizza Place
15	Claremont	Dance Studio	Fried Chicken Joint	Sandwich Place	Latin American Restaurant	Liquor Store
21	Diamond Bar	Mexican Restaurant	Fast Food Restaurant	Pizza Place	Pharmacy	Donut Shop
22	Downey	Mexican Restaurant	Video Game Store	Food Truck	Shoe Store	Sandwich Place
23	Duarte	Hotel	Pool	Hotel Bar	Gym / Fitness Center	Park
24	El Monte	Gym	Japanese Restaurant	Garden Center	Sushi Restaurant	Pizza Place
28	Glendora	Bakery	Pizza Place	Asian Restaurant	Pharmacy	Restaurant
30	Hawthorne	Taco Place	Pharmacy	Light Rail Station	Photography Lab	Bank

Figure 2 Neighborhood's by popular venues

```
index                          38
Neighborhood          La Habra Heights
1st Popular Venues       Grocery Store
2nd Popular Venues    Mexican Restaurant
3rd Popular Venues         Music Venue
4th Popular Venues                Park
5th Popular Venues      Scenic Lookout
Name: 21, dtype: object
```

Figure 3 My choice of final locations

Discussion

Considering the aforementioned results, “La Habra Heights” was the most ideal location for the business to open as, first, there are no similar shops/stores as top venues, meaning no rigorous competition in the area; are venues such as Music Venue and Scenic Lookout which gathers lots of population; and are Grocery Store as top venue indicates there are number of residents in the area, which we hope to attract later to the shop.

Conclusion

In conclusion, as the first fifth venues are not related to "Tea Room," we can say that opening this in LaHabra Heights, a music venue, would be the most profitable choice.