Locating a Coffee shop in the Los Angeles Area

February 3, 2021

Todd S

1. **Introduction**
   a. **Background**

      Opening a business in the food service industry can be difficult in any year, let along during a pandemic. According to research done at the University of California Berkeley approximately 17% of restaurants fail within their first year and 19% of service business fail during their first year; from an investment perspective this makes opening a restaurant very risky. This can make opening a new coffee shop a stressful endeavor and an emphasizes the importance of preparation to ensure a successful business. One of the most critical factors to a coffee shops success is its location which can be difficult to determine

   b. **Problem**

      Data that can assist in helping select a good location might include the number of competitors in an area (number of coffee shops) as well as information that can assist with determining the potential customer base in the area (number of offices)

   c. **Interest**

      This information is of great interest to anyone looking to open a coffee shop because without a proper location, even the best coffee shops have a high chance of becoming a part of the 17% failure statistic.

2. **Data**
   a. **Data Sources**

      In order to help solve this problem data from the Foursquare API was used to gather information related to the locations of offices, coffee shops, cafes, etc. This information is constantly updated as users use the Foursquare service. To determine respective neighborhoods in the LA area, data from the UCLA department of GIS was used

   b. **Data Preparation**

      Initially the data was imported from the UCLA GIS website into a data frame that included the various geometric outlines of neighborhoods in the LA area. The data was imported from the Foursquare API of various places in the LA area

      There was significant amounts of information particularly in the Foursquare data and I decided to split the data into two data frames; one for coffee related places and one for office related places. This would facilitate for some initial independent analysis of the number of offices and the number of coffee shops in the LA area

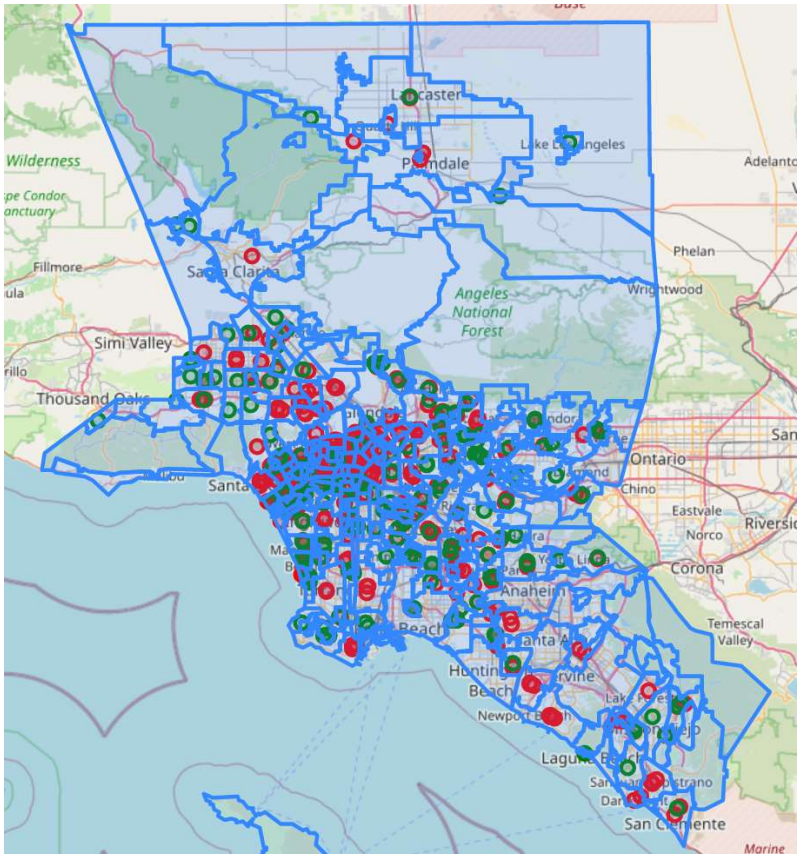
      After some initial analysis I combined the coffee and office data frames into a single data frame that included additional info such as the categorization of each row into either an office location or a coffee shop location. Then to ensure neighborhoods with low numbers of coffee shops and higher number of offices was selected the data was reduced to reflect just neighborhoods that have 1 or less coffee shops and 2 or more offices. This would allow for some more advanced analysis using clustering

3. Methodology

a. Exploratory Analysis

To begin the analysis, first the data needs to be more thoroughly understood. To do this, after retrieving the data and cleaning it, I first mapped the locations of various offices and coffee shops throughout the LA area along with the various neighborhood lines. This allowed me to gain some insights in the how they were geographically spread out.

Fig 1: Map of Various Coffee Shop and Office Locations



Next I attempted some simple sorting to identify the top 25 neighborhood for the most offices and for the least coffeeshops. This will allow myself with a variety of neighborhoods that may be a good candidate as the analysis continues

Fig 2, 3: Neighborhoods with the Most Offices and Limited Number of Coffee Shops

Neighborhood	Number of Offices
Bellflower	4
El Monte	4
Harbor Gateway	4
Manhattan Beach	4
Temple City	4
Culver City	3
Hawthorne	3
Paramount	3
Placentia	3
Rancho Santa Margarita	3
Rolling Hills	3
San Marino	3
Westminster	3
Alhambra	2
Azusa	2
Baldwin Hills/Crenshaw	2
Broadway-Manchester	2
Buena Park	2
Canoga Park	2
Cheviot Hills	2
Cypress	2
Diamond Bar	2
El Segundo	2
Emerald Bay	2
Fullerton	2

Neighborhood	Number of Coffee Shops
Altadena	1
Athens	1
Avocado Heights	1
Baldwin Hills/Crenshaw	1
Beverlywood	1
Boyle Heights	1
Chatsworth	1
Chesterfield Square	1
Cudahy	1
Cypress	1
East Pasadena	1
East Whittier	1
El Sereno	1
Elysian Park	1
Fountain Valley	1
Gramercy Park	1
Granada Hills	1
Harbor Gateway	1
Harvard Park	1
Hawaiian Gardens	1
Hawthorne	1
Hollywood Hills West	1
La Crescenta-Montrose	1
La Palma	1
Laguna Hills	1

After reviewing the various lists and looking at the map it is clear that further analysis will need to be conducted before any recommendations can be made.

b. Analysis including Clustering and Heat Maps

After some exploratory analysis to understand the geographic layout and various neighborhoods in the area. I decided to be more specific in the list of potential neighborhoods. To understand particular areas that have high numbers of coffee shops and high numbers offices, I used heat maps of the LA area for each respective category

Fig 4: Heat Map of the Number of Offices

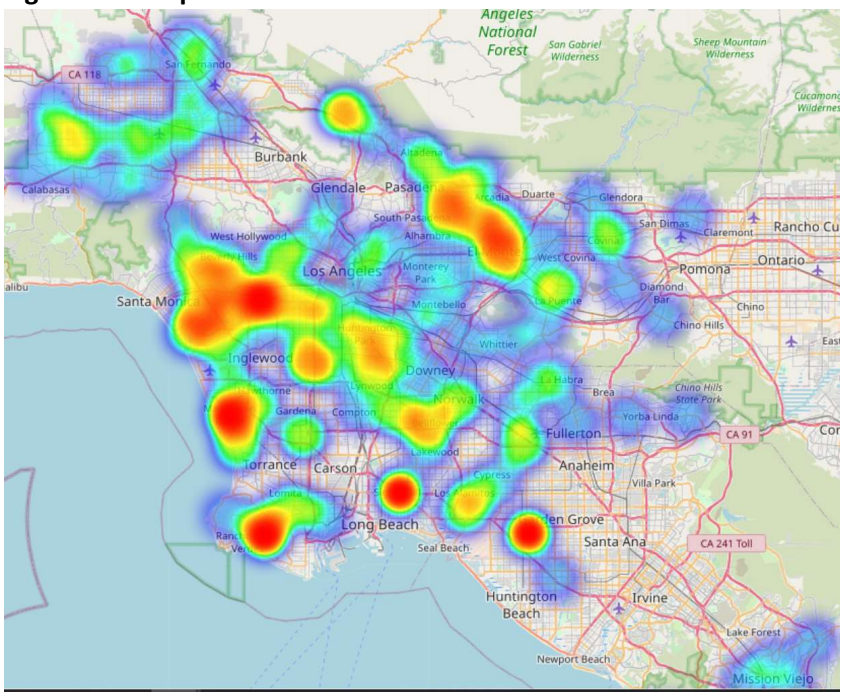
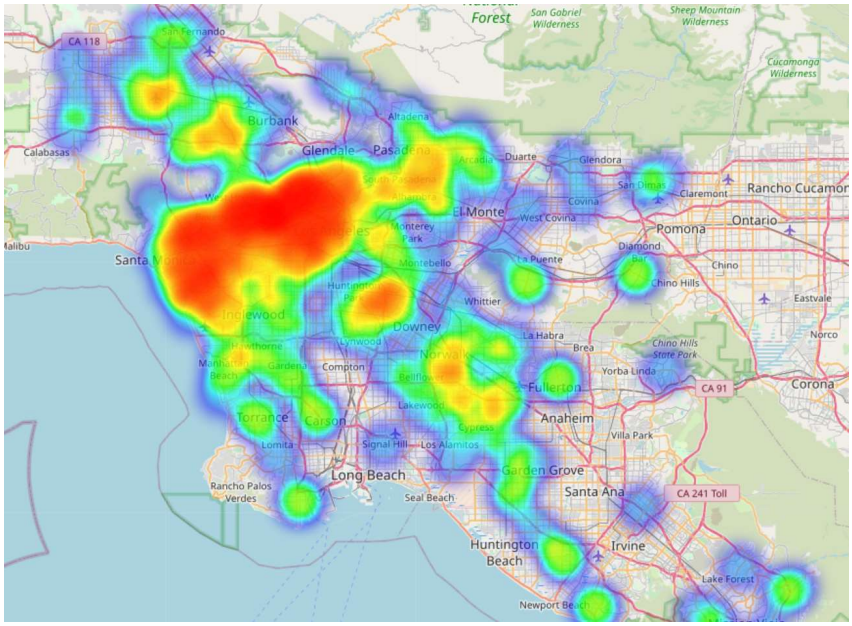
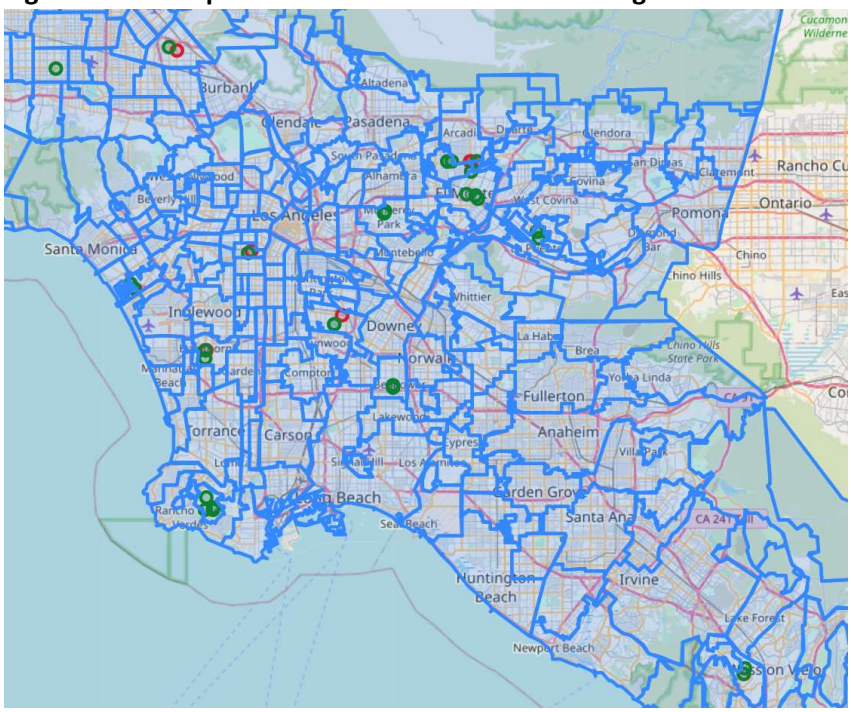


Fig 5: Heat Map of the Number of Coffee Shops



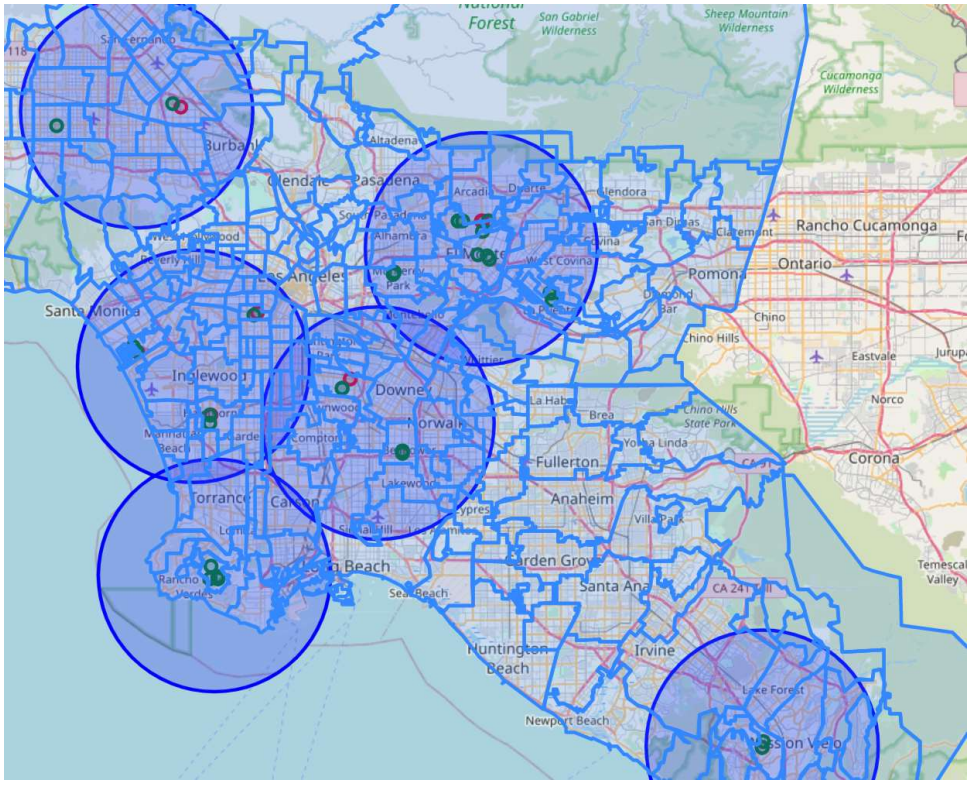
Looking at these two maps together a couple of potential areas stand out: Rolling Hills and Signal Hill. To narrow down the list of potential neighborhoods though I first needed to determine what defined a potential area; to do this I decided that a neighborhood with 1 or less coffee shops and 2 or more offices would be a good criterion. I then mapped the coffee shops and offices in the remaining neighborhoods to get an initial understanding of the remaining neighborhoods

Fig 6: Coffee Shops and Offices Narrowed Down Neighborhoods



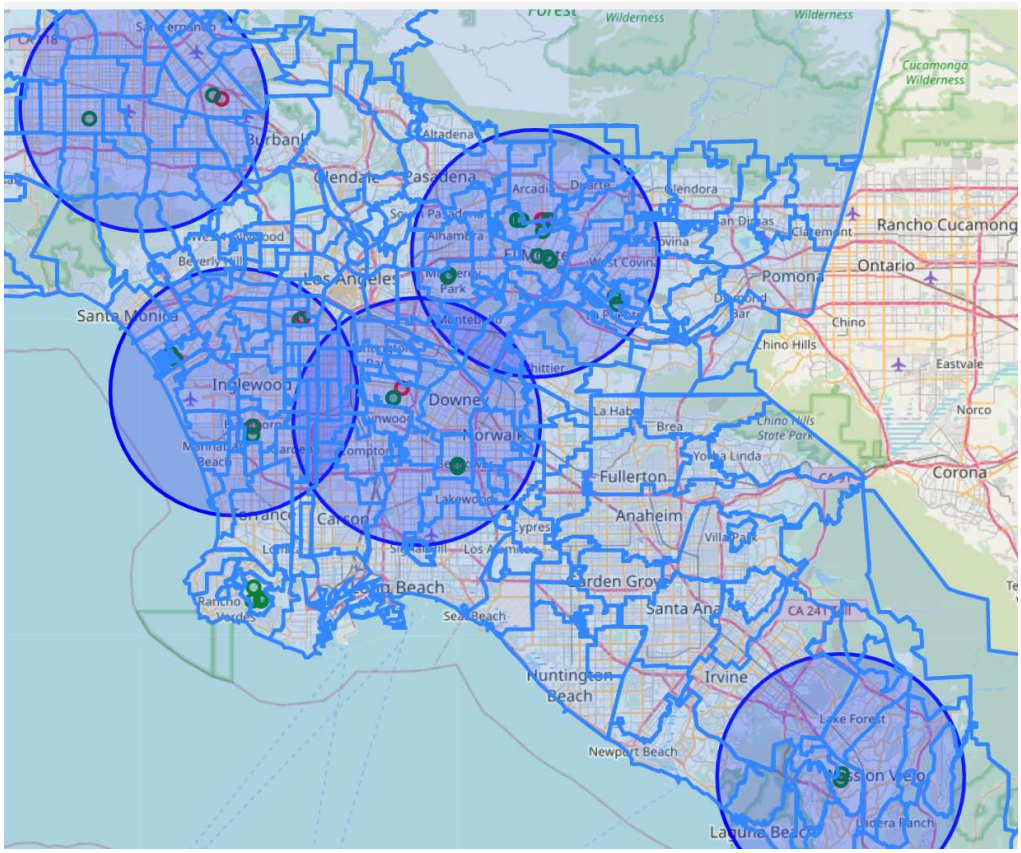
After narrowing down the list I wanted to understand if there were any similarities between neighborhoods by clustering them using both K-Means clustering and then DBSCAN clustering. First I clustered the data using k-means clustering based on simply the number of offices and coffee shops

Fig 7: K-Means Clustering Coffee Shops and Offices Only



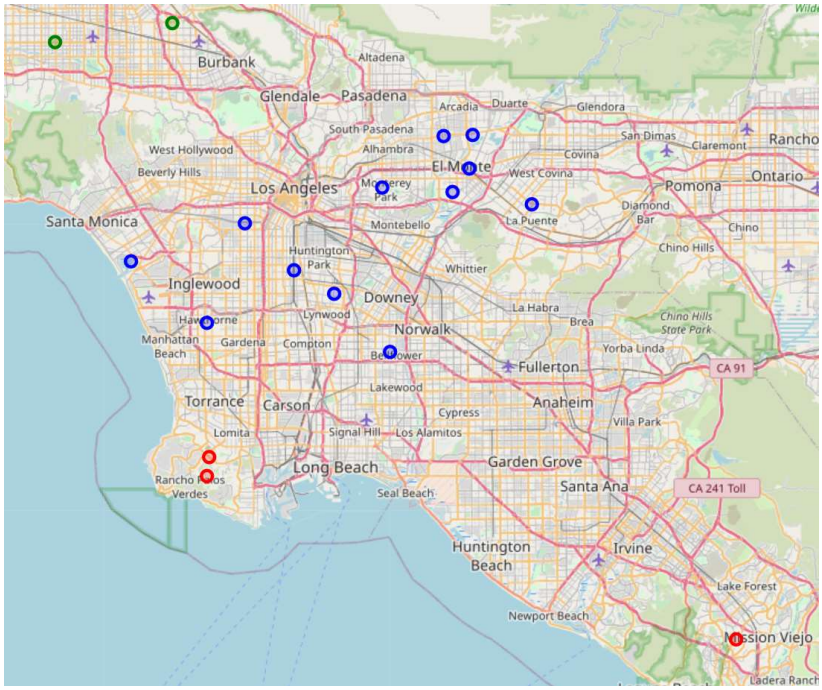
Then to understand if the results would change with additional locations used in areas, I clustered using a larger foursquare dataset of the various venues in the neighborhood

Figure 8: K-Means Clustering all Venues



I then wanted to explore clustering using the DBSCAN method so I could gain further insights into the similarities between various neighborhoods

Figure 9: DBSCAN Clustering



Looking at the relationship between the various neighborhoods will allow me to further narrowing down potential locations by eliminated similar neighborhoods that are not a top candidate. The results of this analysis will be discussed in the next sections along with a recommendation of a few neighborhoods that could potentially be a good location for a coffee shop.

4. Results

While going through the analysis, I was consistently narrowing down the potential locations for a coffee shop. First beginning by using heat maps to determine where the are high concentrations of coffee shops and offices. Looking at the heat maps, it appears that the area surrounding the city of LA itself have high concentrations of coffee shops, particularly as you head towards Santa Monica. The number of coffees shops however, appear to be more spread out throughout the LA area

To help understand how the areas are related I used two types of clustering. This analysis further re-enforced that the central LA area had a lot in common whereas down in the southern region of the LA area (Long Beach and Orange County) have a lot in common. In looking at these areas one of the things that separates them is the higher concentration of coffee shops relative to the number of offices in the area. It is important to note that both Rolling Hills and Mission Viejo have a relatively low number of coffee shops compared with other parts of the LA area.

5. Discussion

Looking at the results, I believe the Rolling Hills or Mission Viejo neighborhood could be a very good location for a coffee shop. This is because there is a very low concentration of coffee shops (little or none in the Foursquare data) and there are multiple offices which means a significant potential customer base. Also, by looking at the clustering data I was able to determine that there are not many neighboring neighborhoods that have coffee shops that

could attract customers due to the proximity. It is important to understand that competition can come from a neighborhood next to you, not just your particular neighborhood.

6. **Conclusion**

By conducting this analysis, I was able to determine two good potential areas to open a coffee shop based on the data rather than word of mouth. Of course, the location is one of many factors that need to be considered whenever opening a coffee shop though. Hopefully this analysis helps any potential coffee shop entrepreneurs that are looking to kick of their business, especially in these difficult times. I recommend that anyone considering opening a coffee shop in the LA area check out these two areas and then after reviewing other important things such as financing, type of coffee shop, etc. make their informed decision.