

Setting Up a New Coffee Shop in Los Angeles, US

By Tien Nguyen



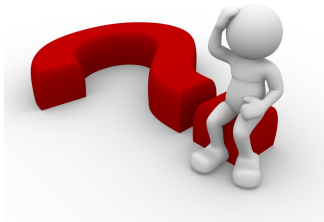
Introduction

- More than 450 million cups of coffee are consumed in the United States every day
- 63% of American adults drink coffee daily.
- Reasons people love to go to the coffee shop
- + getting a cup of coffee with the high-quality ingredients, best brewing recipes, consistency, fresh and appealing sweet & savory selections and
 - + meeting up or gossip with your friend,
 - + doing some work,
 - + reading a book,
 - + entertaining or simply passing some time.

Introduction



- Location is one of factors that make a successful coffee shop.
 - Los Angeles area is the largest city in California, United States => the most populated city in California
- => Los Angeles is considered as a good place to start this business.



Which neighborhoods in Los Angeles should a new coffee shop be located ?

Data Sets



WIKIPEDIA
The Free Encyclopedia

- Scraping neighborhood of Los Angeles from the Wikipedia website
- Using Python Geocoder package to get latitude and longitude coordinates of each neighborhoods.
- In total, 190 rows and 3 features.

	Neighborhood	Latitude	Longitude
0	Angelino Heights, Los Angeles	34.070278	-118.254722
1	Angeles Mesa, Los Angeles	33.994200	-118.313600
2	Angelus Vista, Los Angeles	34.046954	-118.317488
3	Arlington Heights, Los Angeles	34.241944	-118.425556
4	Arlington Heights, Los Angeles	34.042222	-118.318889



Source : https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles

Data Sets



FOURSQUARE

- Venue data from Foursquare API to get the top 100 venues that are within a radius of 2000 meters for those neighborhoods.
- As a result, the Foursquare venue dataset has 16909 observations where each observation consists of neighborhood latitude, neighborhood longitude, venue, and venue category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Angelino Heights, Los Angeles	34.070278	-118.254722	Guisados	Taco Place
1	Angelino Heights, Los Angeles	34.070278	-118.254722	Halliwel Manor	Performing Arts Venue
2	Angelino Heights, Los Angeles	34.070278	-118.254722	Eightfold Coffee	Coffee Shop
3	Angelino Heights, Los Angeles	34.070278	-118.254722	Subliminal Projects	Art Gallery
4	Angelino Heights, Los Angeles	34.070278	-118.254722	Button Mash	Arcade

Data Sets

- The population in a neighborhood also plays an essential role in the decision of opening a new coffee shop.
- **However**, no such service providing the population of each neighborhood
- Assumption : the more venues of a neighborhood, the more residents living there.
- Classifying the population into three categories depending on the number of venues that each neighborhood has as follows :
 - + Small population : the number of venues between 0 and 20 (exclusive)
 - + Medium population : the number of venues between 21 and 100 (exclusive)
 - + Large population : with the number of venues of 100 or more

	Neighborhood	Total Venue	Population Classification
0	Angeles Mesa, Los Angeles	75	Medium
1	Angelino Heights, Los Angeles	100	Medium
2	Angelus Vista, Los Angeles	100	Medium
3	Arleta, Los Angeles	52	Small
4	Arlington Heights, Los Angeles	100	Medium

Data Sets



Changing all categories that are synonyms of Coffee to Coffee Shop such as “Café”

Data Sets

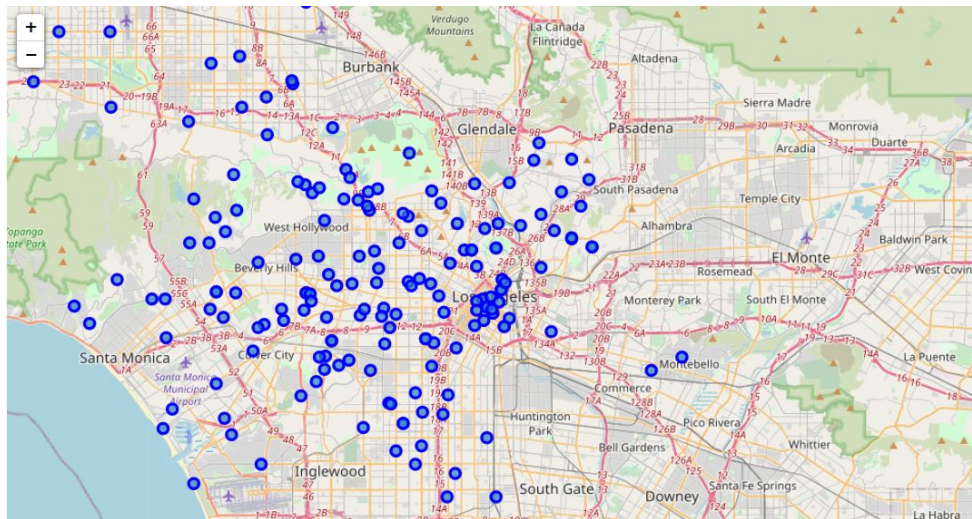
- The number of coffee shops in each neighborhood is one of the most important features in the model.
- However, we don't want to set up a new coffee shop in a neighborhood that has a lot of coffee shops but less population.

=> Calculating the percent of coffee shops located in each neighborhood.

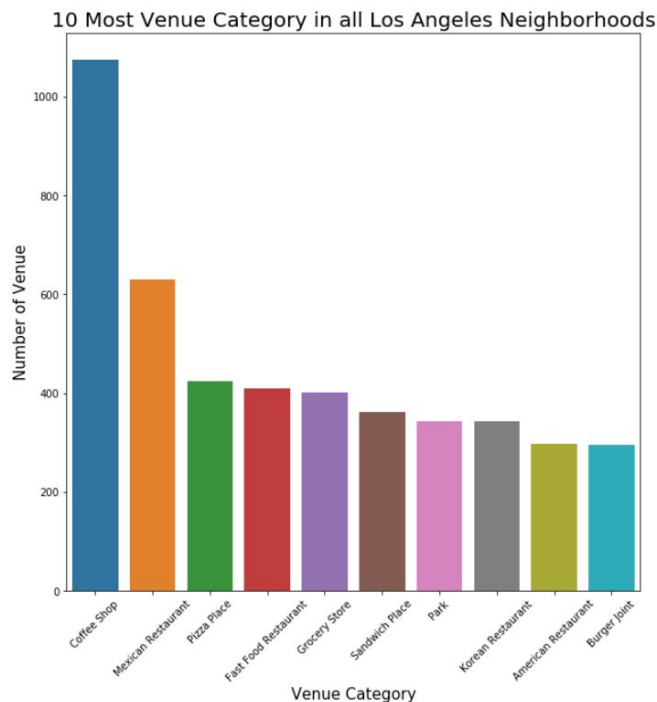
	Neighborhood	Coffee Shop	Total Venue	Population Classification	Percent
0	Angeles Mesa, Los Angeles	1	75	Medium	1.333333
1	Angelino Heights, Los Angeles	13	100	Large	13.000000
2	Angelus Vista, Los Angeles	6	100	Large	6.000000
3	Arleta, Los Angeles	0	52	Medium	0.000000
4	Arlington Heights, Los Angeles	6	100	Large	6.000000

Visualization World

Los Angeles map containing all the neighborhoods

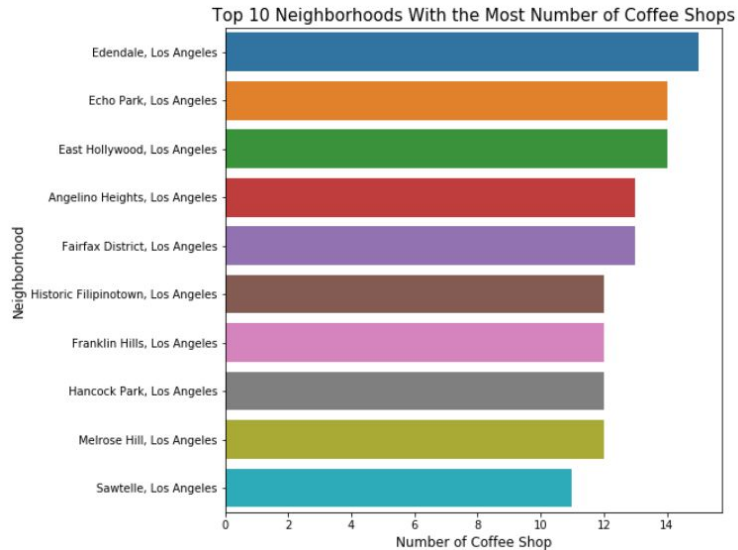


Top 10 most venue categories in the Los Angeles area



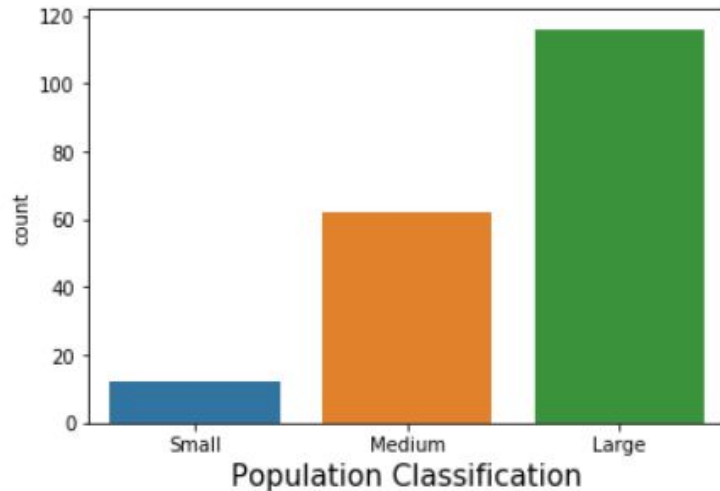
- About 416 unique venue categories in Los Angeles
- Coffee shops are at the top of the charts
- Number of coffee shops is about 1.5 times the second most common venue category (Mexican Restaurant)

Neighborhoods that have the most number of coffee shops



- Edendale, which has 15 coffee shops (neighborhood that has the most number of coffee shops.)
- Among the top 10 positions, Historic Filipinotown, Franklin Hills, Hancock Park, and Melrose Hill have the same number of coffee houses.

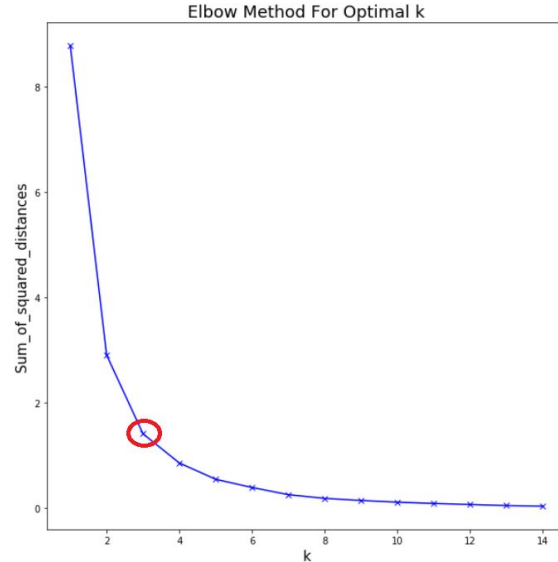
Number of the neighborhood in each population classification



- About 18 neighborhoods are having fewer than 60 venues which are considered a low population
- About 110 neighborhoods that have more than 100 venues which implies that these neighborhoods have many residents.

Clustering model

- Performing K-mean clustering with the feature “Percent” to cluster the neighborhood.
- K-mean algorithm is one of the most common cluster methods of unsupervised learning.
- To specify the number of clusters K, we used the elbow method.

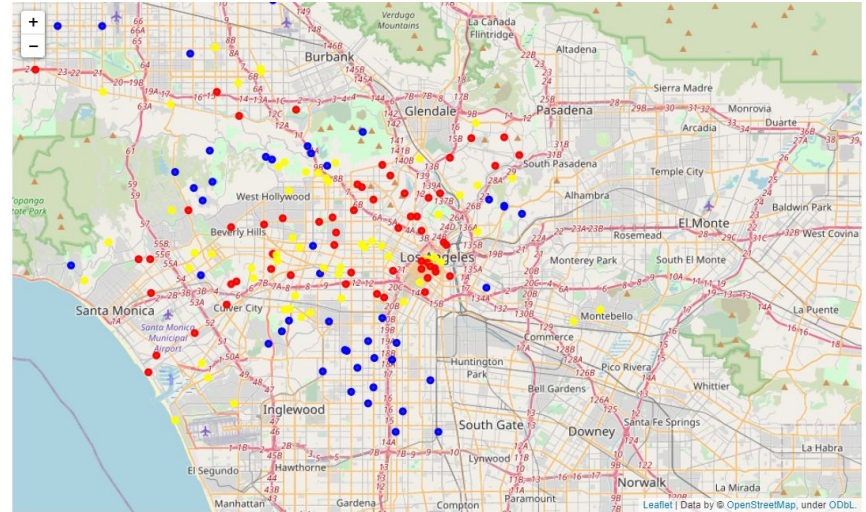


Number of cluster = 3

Results

Cluster 0

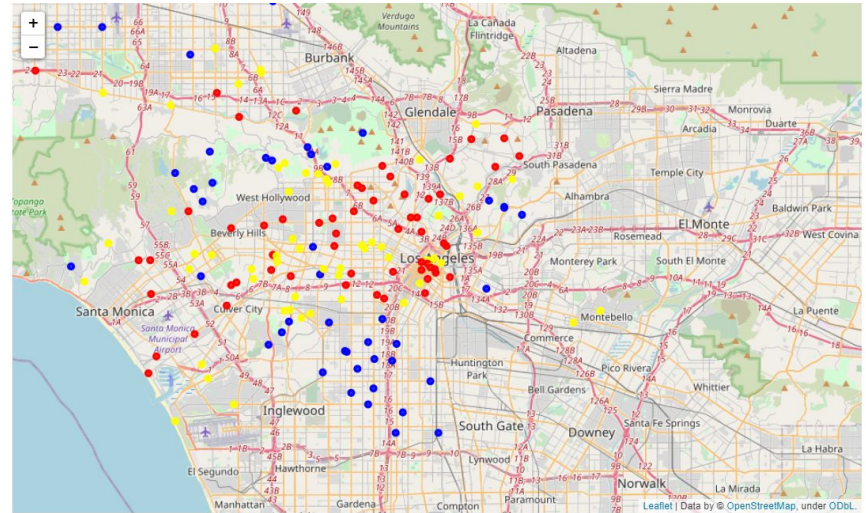
- Has 78 neighborhoods in the cluster.
- Neighborhoods that have a moderate number of coffee shops.
- Each neighborhood has between 1 to 8 coffee shops.
- It takes 4% to 8% of the total of those neighborhoods' venues.
- Cluster 0 is colored in red.



Results

Cluster 1

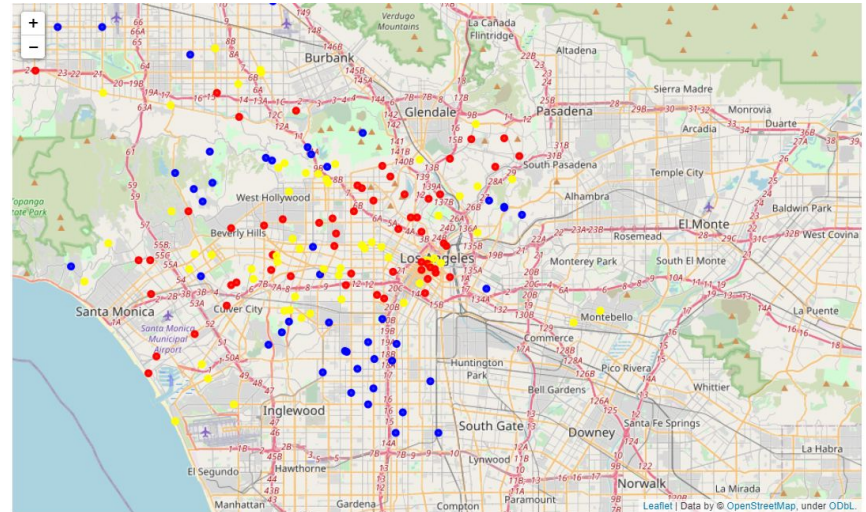
- Has 60 neighborhoods in the cluster
- Neighborhoods that have the fewest number of coffee shops.
- There are less than 8 coffee shops in each neighborhood.
- It takes 0% to 4% of the total of those neighborhoods' venues.
- Cluster 1 is colored in blue.



Results

Cluster 2

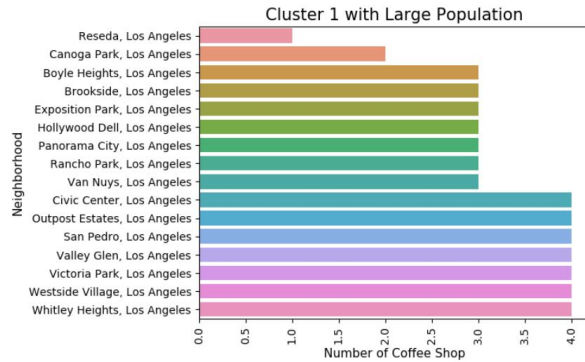
- Has 52 neighborhoods in the cluster
- Neighborhoods that have a high number of coffee shops.
- The number of coffee shops in each neighborhood is between 2 to 15.
- It takes 9% to 16% of the total of those neighborhoods' venues.
- Cluster 2 is colored in yellow.



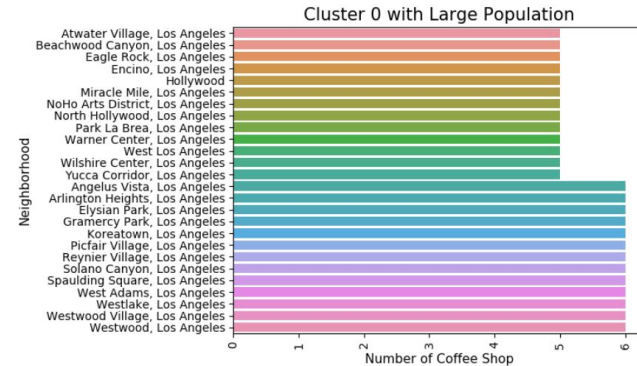
Discussion

- The moderate and high percent number of coffee shops are concentrated in the center of Los Angeles.
- Cluster 1 (blue) has a very low percentage of coffee shops.
=> a great opportunity and high potential location to open a new coffee shop because there is very little to no competition to open a new coffee shop.
- Cluster 0 can be chosen to open a new coffee shop if the population is large but the number of coffee shops is not too high.
- Cluster 2 neighborhoods should be avoided setting up a new coffee shop since they already have a high concentration of coffee shops and may suffer from intense competition.

Discussion (Recommendation)



Reseda is a good neighborhood to set up a new coffee shop as it only has one coffee shop but it has a large population.



Atwater Village has a large population but only 5 coffee shops.

Conclusion & Future Work

- Built a useful model to assist our contractors/clients to find a neighborhood that their coffee shop should be set up in.
- Recommended potential neighborhoods that have a few coffee shops but have a large population.
- Considering other factors such as the cost of rent, the population of the neighborhood, etc.
- Classified the population of each neighborhood based on the number of venues in each neighborhood.
- Using the free FourSquare API that came with limitations as to the number of API calls and results returned.
- For future work, can build a better recommendation if we have more fields and more data for our K-means clustering algorithm.