

# Cafe in California

Sai Thiha Zaw

June 25, 2020

## Introduction

California is a very popular place for innovation and startup. The 41 million population is big enough for market size and we think most people consume coffee daily. We discussed with our investor to open a coffee shop in California. So we explore a popular coffee shop and their location data using data science techniques.

## Data Description

We try to take data techniques advantage for making decisions for our business. We'll use California data from Kaggle datasets here and Foursquare API. We explore the following.

- The number of coffee shops and popular places.
- Most crowded areas of the city and rental prices.
- The popular place in California.

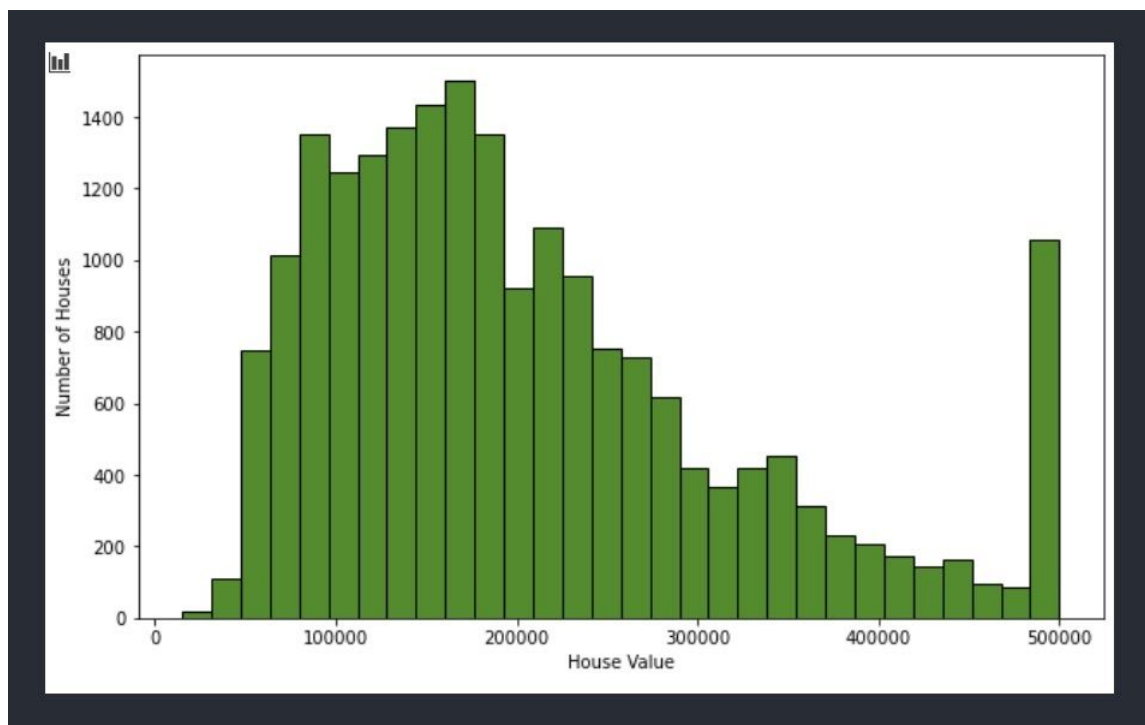
## Methodology

I explore California housing price datasets to estimate rental prices. Kaggle has a lot of datasets but some are not very updated. Cleaning the data to get house value, latitude, longitude, and population. Using the Folium library to visualize geographic details. Explore coffee shops in the target area using Foursquare API and geocoder library for California coordinates. After knowing venues, neighborhoods, and nearby places, we find the top ten places around the target. And then I use the k-means algorithm for finding neighborhood clusters.

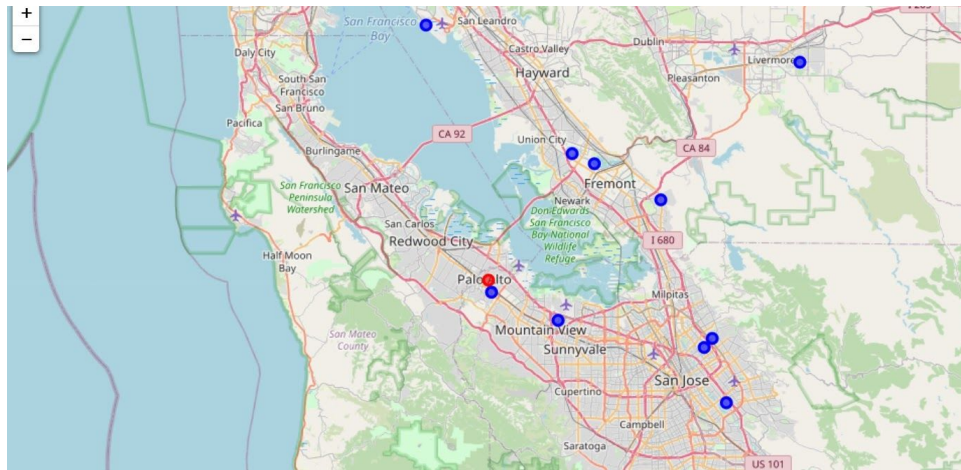
## Results section

I found the average California house prices are a maximum 500001, minimum 14999, median 179700. So we predict our room rental charges may be around 17000 per year. I also want to know which areas are most crowded so using population data for exploring the top hundred areas and placing markers by Folium library.

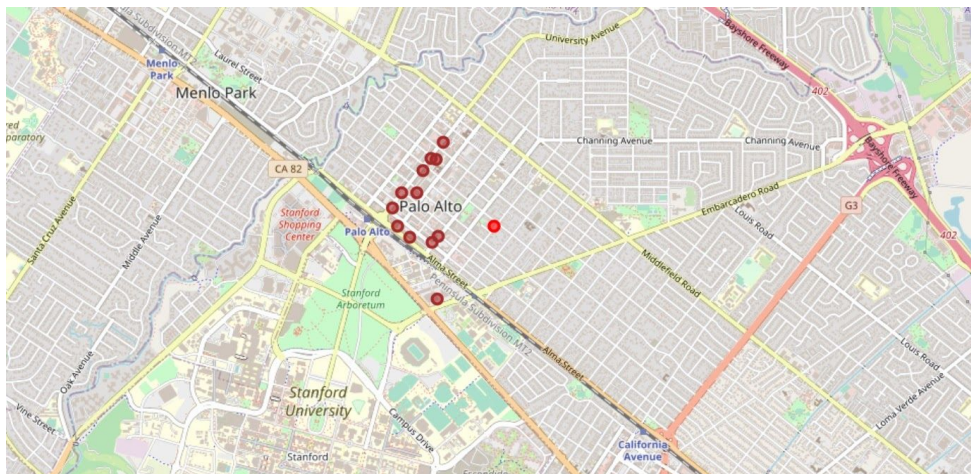
	longitude	latitude	population	median_income	median_house_value
count	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000
mean	-119.569704	35.631861	1425.476744	3.870671	206855.816909
std	2.003532	2.135952	1132.462122	1.899822	115395.615874
min	-124.350000	32.540000	3.000000	0.499900	14999.000000
25%	-121.800000	33.930000	787.000000	2.563400	119600.000000
50%	-118.490000	34.260000	1166.000000	3.534800	179700.000000
75%	-118.010000	37.710000	1725.000000	4.743250	264725.000000
max	-114.310000	41.950000	35682.000000	15.000100	500001.000000



Maker on Population and finding Silicon Valley coordinates using the Geolocator library, they are latitude 37.4429964, longitude -122.1545229.



Explore popular coffee shops and nearby cafes using the Foursquare API. I have got Peet's Coffee & Tea, Philz Coffee, Blue Bottle Coffee, WFM Coffee Bar, Verve Coffee, JM Coffee, etc.. "Blue Bottle Coffee" got the highest score (9.1) by Foursquare's rating.



It has 96 unique categories including Historic Site, Park, Greek Restaurant, Gym / Fitness Center, Spa, etc.,

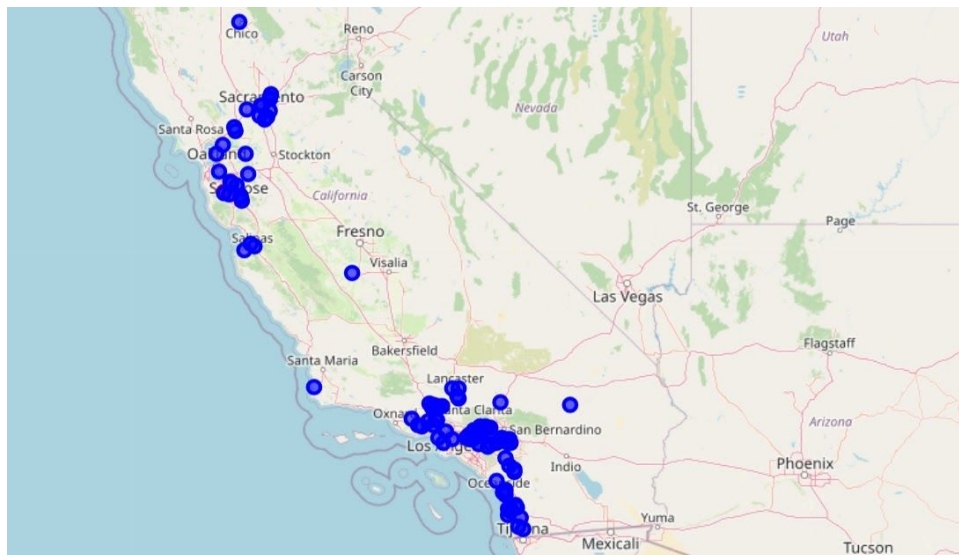
After analyzing I have got the top five neighborhoods. They are American Restaurant, Antique Shop, Construction & Landscaping, Diner, and Elementary School. Each neighbor has ten common venues.

Running the K-means algorithm for clustering neighborhoods, I got knowledge of the best neighbor for coffee shops. Park, Greek Restaurant, Gym/Fitness Center, Spa, Grocery Store, Diner, American Restaurant, Antique, they are best neighbors for coffee shops in Silicon Valley, California.



## Discussion

As population data results, We should also consider opening a coffee shop in Los Angeles city near Palo Alto city. There is similar socioeconomics.



## Conclusion

In this study, Palo alto city, University Avenue Road, and Alma street are the best places for opening a coffee shop. We should also consider American Restaurant, Gym/Fitness Center and spa. Thank you.