Clustering the neigborhoods of California City

Victor Feunou

1 Introduction

Every family or business often has the aspiration to own its own house or building. While certainly a nice desire, financial situations often are a key barrier to overcome. For those having the financial power, an important choice to make is the neighborhood where they want to be owners. For small cities, the choice of the neighborhood can be solved by a quick tour of the city and survey from the internet. However, for a large city like California and a non-resident planning to be an owner, the choice of the neighborhood requires expertise knowledge of features proper to neighborhoods of the city. Features of the neighborhoods are important for a match with the future owner criteria. It is therefore important to understand the similarities and dissimilarities of the neighborhoods of the city of California. This is the purpose of this project.

1.1 Objective

We want to the build a segmentation and clustering of the neighborhoods of the city of California.The objective of the segmentation/Clustering is to provide guidance to investors or privates who want to buy a house in this city. Hence our aim is to identify neighborhood clusters in the city based on geographical and economic features of a neighborhood. Some include

1. Proximity to the ocean
2. number of popular venues
3. Median income
4. Median house value
5. Median age of houses
6. Number of households

1.2 Interest

The primary group that might be interested in the clustering of houses are real estate companies. They are the one investors or private often turn to when looking to buy a building. The clustering will enable them to quickly find a choice that fits the criteria of a home seeker.
A second group for which this project is of interest is made up of businesses such as Theater/Movies, Restaurants, Malls. Cluster will provide information on population class (medium, elites) in the neighborhoods which enable business to deploy targeted products.
The clustering of neighborhood is also important for the city council. Depending on the factors defining the clusters, some neighborhoods might be rebuild or appropriate plans developed.

2 Data Wrangling

As dataset, we use a modified version of the famous Califonia Housing prices data. It is a census data from 1990 for 20640 block group in the city of California. Each block group can be considered as a neighborhood of California and the data has the following metrics

1. longitude
2. latitude
3. housing median age
4. number of households
5. population
6. median house value
7. median house income
8. ocean proximity

Ocean proximity is represented by a categorical variable ocean_proximity with four values: *NEAR BAY, <1 H OCEAN, INLAND, NEAR OCEAN*. The other metrics are quantitative. Our version is taken from the Kaggle website and it is available here. For our analysis, we have omitted two other metrics from the dataset: total number of rooms, total number of bedrooms.

2.1 Data Cleaning

Fortunately for the above metrics, there were no missing values. However, we only choose neighborhoods with a population of more than 4000 people. With this cap on the population size, our dataset reduced to 586 rows.
The next step for our data cleaning has to do with the geographical coordinates. The longitude and latitude in the data were collected up to two decimal points. As a result, several neighborhoods had the same geographical coordinates. We group neighborhoods with same geographical coordinates together and took the mean value of their metrics. For the categorical value ocean_proximity, we took the mode. With this grouping, our dataset reduced to 575 rows.

2.2 Adding the metrics Number of venues using Foursquare API

We consider the number of popular venues in a neighborhood to be an important criteria for the choice of the neighborhood. Popular venues in a neighborhood determines activity level and the lifestyle in the neighborhood.
To determine the number of venues present in each neighborhood, we use the foursquare API calls. We found in total 4135 venues divided into 349 categories for the city of California. It turns out only 475 neighborhoods had at least one popular venue. Our final dataset is therefore reduced to 475 rows.
In principle, the venue categories should be used as features of our analysis since they described the neighborhoods in a more detailed way. However, the number of the categories is very high to be used as features. We will exploit the venue categories to describe the cluster we obtained later. We discuss how venue categories can be grouped together and use as features for the analysis in Section 4.

2.3 Dealing with the categorical variable ocean_proximity

As discussed above, the variable ocean_proximity has four values. We use *onehot encoding* for this categorical variables to obtain the four quantitative features taking values in the set $\{0, 1\}$.

3 k-means clustering of the neighborhoods

To understand the similarities and dissimilarities of the neighborhoods, we use the *k-means clustering* algorithm. The key step of the algorithm is to choose the number of clusters. We achieve this by using the elbow approach which consists in exploring the error produced by the algorithm for a range of values of clusters and choose the value of k for which the increase of the number of clusters does not lead to a steep decrease of the error. In our case, we found 5 to be the best number of clusters to use. For the design of our algorithm, we use 150 initialization of the centroids and a maximum of 300 iterations per initialization. A visualization of the clusters is given by Figures 1 and 2. The labelling is as follows:

1. Cluster 1: red bullets
2. cluster 2: dark magenta bullets
3. cluster 3: dark midnight blue bullets
4. cluster 4: dark olive green bullets
5. cluster 5: dark orange bullets

Figure 1: Neighborhood clustering North California

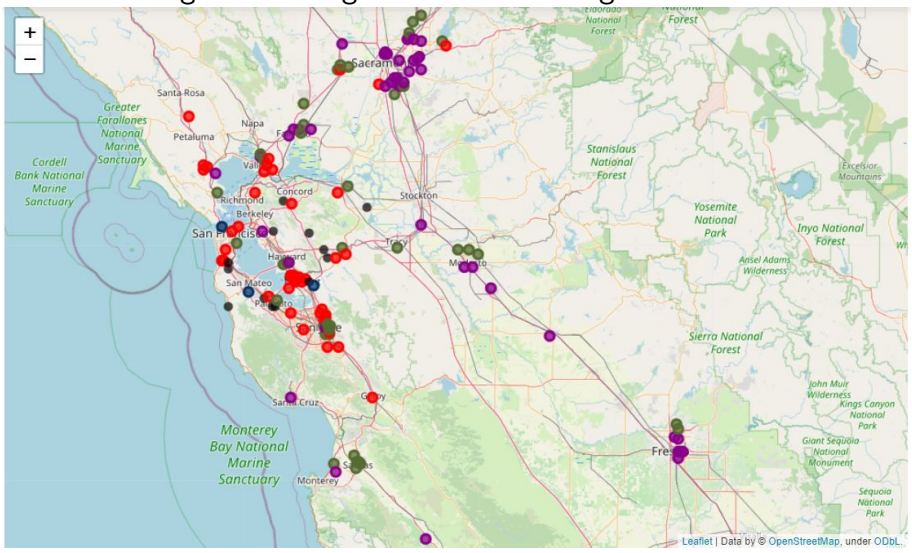
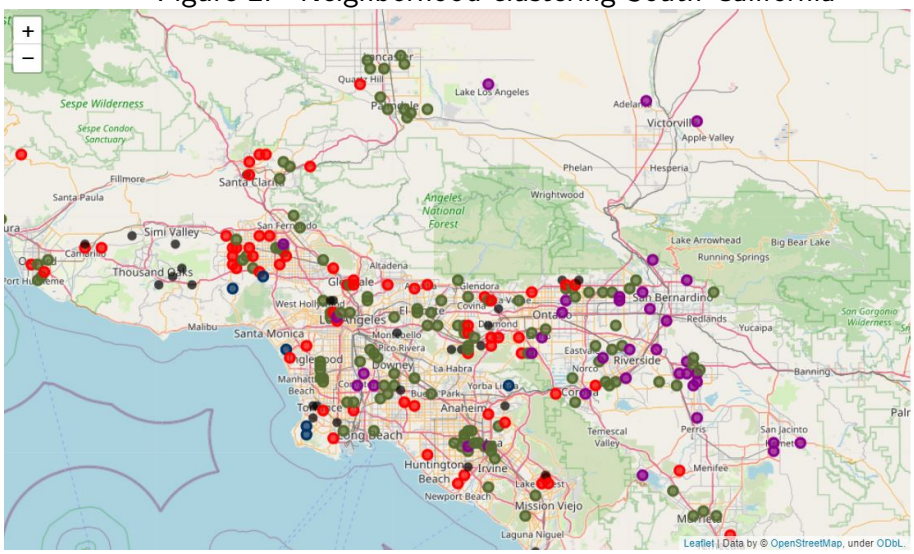


Figure 2: Neighborhood clustering South California



3.1 Cluster 1

Cluster 1 is made up of 122 neighborhoods. Figure 3 below gives a statistic description of the metrics while Figure 4 provides a description of the top 10 most common venue categories in each of the neighborhoods in Cluster 1. A key distinctive feature of this cluster is that the house prices are around 200,000 $ and the population living here can be considered as middle class: most have an income above 3,600 $. Another characteristic of the cluster is that most neighborhoods have 5 popular venues. This cluster is displayed in Figures 1 and 2 by the red bullets.

Figure 3: Description of metrics: Cluster 1

	count	mean	std	min	25%	50%	75%	max
Cluster Labels	122.0	0.000000	0.000000	0.000	0.00000	0.00000	0.000000	0.000
housing_median_age	122.0	15.143443	8.820403	2.000	7.00000	15.00000	21.000000	52.000
population	122.0	5826.450820	1801.706316	4014.000	4483.75000	4980.50000	7175.500000	13251.000
households	122.0	2066.885246	728.158134	962.000	1559.00000	1854.50000	2416.500000	4855.000
median_income	122.0	4.757689	1.237185	1.724	3.82415	4.95465	5.672625	6.992
median_house_value	122.0	244891.393443	23864.878489	209800.000	225000.00000	239750.00000	263900.000000	290500.000
<1H OCEAN	122.0	0.606557	0.490528	0.000	0.00000	1.00000	1.000000	1.000
INLAND	122.0	0.155738	0.364102	0.000	0.00000	0.00000	0.000000	1.000
NEAR BAY	122.0	0.163934	0.371743	0.000	0.00000	0.00000	0.000000	1.000
NEAR OCEAN	122.0	0.073770	0.262475	0.000	0.00000	0.00000	0.000000	1.000
Num Venues	122.0	8.827869	12.396142	1.000	2.00000	5.00000	9.750000	82.000

Figure 4: Neighborhood Description: Cluster 1

	0	1	2	3	6	8	9	13	15	18	...	433	434
Neighborhood	0	1	2	4	8	11	12	16	18	21	...	520	521
Cluster Labels	0	0	0	0	0	0	0	0	0	0	...	0	0
1st Most Common Venue	Farm	Baseball Field	Park	Fast Food Restaurant	Restaurant	Park	Grocery Store	Hotel	Park	BBQ Joint	...	Park	Athletics & Sports
2nd Most Common Venue	Indian Restaurant	Zoo Exhibit	Convenience Store	Mexican Restaurant	Snack Place	Hostel	Filipino Restaurant	Cocktail Bar	Island	Zoo Exhibit	...	Flower Shop	Zoo Exhibit
3rd Most Common Venue	Flower Shop	Garden Center	Construction & Landscaping	Nail Salon	Sushi Restaurant	Burmese Restaurant	Park	Boutique	Beach	Farm	...	Farmers Market	Farm
4th Most Common Venue	Farmers Market	Fast Food Restaurant	Flower Shop	Shoe Store	Chinese Restaurant	Dessert Shop	Fried Chicken Joint	Art Gallery	Tunnel	Fast Food Restaurant	...	Fast Food Restaurant	Fast Food Restaurant
5th Most Common Venue	Fast Food Restaurant	Filipino Restaurant	Farmers Market	Bakery	Sandwich Place	Farmers Market	Flea Market	Café	Zoo Exhibit	Filipino Restaurant	...	Filipino Restaurant	Filipino Restaurant
6th Most Common Venue	Filipino Restaurant	Fish & Chips Shop	Fast Food Restaurant	Paper / Office Supplies Store	Sporting Goods Shop	Fast Food Restaurant	Farmers Market	Italian Restaurant	Flea Market	Fish & Chips Shop	...	Fish & Chips Shop	Fish & Chips Shop
7th Most Common Venue	Fish & Chips Shop	Fish Market	Filipino Restaurant	Clothing Store	Bowling Alley	Filipino Restaurant	Fast Food Restaurant	Gym / Fitness Center	Fast Food Restaurant	Fish Market	...	Fish Market	Fish Market
8th Most Common Venue	Fish Market	Fishing Spot	Fish & Chips Shop	Coffee Shop	Café	Fish & Chips Shop	Frozen Yogurt Shop	Church	Filipino Restaurant	Fishing Spot	...	Fishing Spot	Fishing Spot
9th Most Common Venue	Fishing Spot	Fishing Store	Fish Market	Lake	Filipino Restaurant	Fish Market	Fish & Chips Shop	Spa	Fish & Chips Shop	Fishing Store	...	Fishing Store	Fishing Store
10th Most Common Venue	Fishing Store	Flea Market	Fishing Spot	Kids Store	Liquor Store	Fishing Spot	Fish Market	Kitchen Supply Store	Fish Market	Flea Market	...	Flea Market	Flea Market

12 rows × 122 columns

3.2 cluster 2

Here we have 144 neighborhoods. Figures 5 and 6 give a description of the metrics and venue categories of the neighborhoods. This cluster has 3 distinctive features:

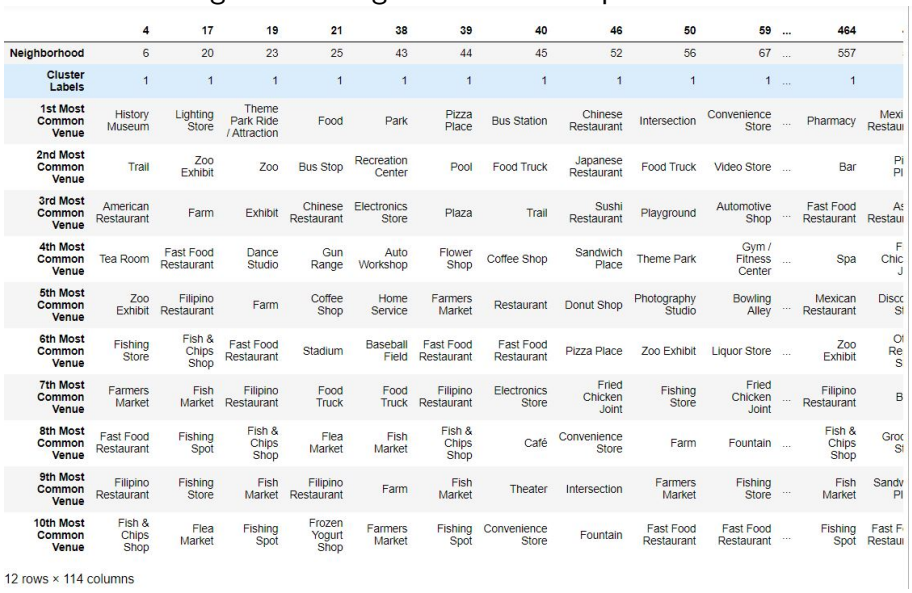
- They are situated inland
- The population in this cluster can be classified as low class: 75% earned below 4,000 $.
- Houses also have a very low value compare to the other clusters.

Neighborhoods in this cluster are displayed with magenta bullets on the map in Figures 1 and 2.

Figure 5: Description of metrics: Cluster 2

	count	mean	std	min	25%	50%	75%	max
Cluster Labels	114.0	1.000000	0.000000	1.0000	1.000000	1.000000	1.000000	1.0000
housing_median_age	114.0	15.381579	7.940961	4.0000	10.000000	15.000000	19.000000	43.0000
population	114.0	5715.657895	2574.103492	4002.0000	4559.250000	5033.000000	6116.750000	28566.0000
households	114.0	1859.057018	756.686956	348.0000	1431.250000	1691.500000	2114.250000	6082.0000
median_income	114.0	2.917262	0.826155	1.2065	2.317675	2.805525	3.597525	5.0012
median_house_value	114.0	108519.736842	24108.948526	35000.0000	90525.000000	114400.000000	129525.000000	139800.0000
<1H OCEAN	114.0	0.157895	0.366252	0.0000	0.000000	0.000000	0.000000	1.0000
INLAND	114.0	0.780702	0.415598	0.0000	1.000000	1.000000	1.000000	1.0000
NEAR BAY	114.0	0.035088	0.184814	0.0000	0.000000	0.000000	0.000000	1.0000
NEAR OCEAN	114.0	0.026316	0.160779	0.0000	0.000000	0.000000	0.000000	1.0000
Num Venues	114.0	6.035088	7.167935	1.0000	2.000000	4.000000	6.750000	46.0000

Figure 6: Neighborhood Description: Cluster 2



	4	17	19	21	38	39	40	46	50	59	...	464	
Neighborhood	6	20	23	25	43	44	45	52	56	67	...	557	
Cluster Labels	1	1	1	1	1	1	1	1	1	1	...	1	
1st Most Common Venue	History Museum	Lighting Store	Theme Park Ride / Attraction	Food	Park	Pizza Place	Bus Station	Chinese Restaurant	Intersection	Convenience Store	...	Pharmacy	Mexi Restau
2nd Most Common Venue	Trail	Zoo Exhibit	Zoo	Bus Stop	Recreation Center	Pool	Food Truck	Japanese Restaurant	Food Truck	Video Store	...	Bar	Pi Pl
3rd Most Common Venue	American Restaurant	Farm	Exhibit	Chinese Restaurant	Electronics Store	Plaza	Trail	Sushi Restaurant	Playground	Automotive Shop	...	Fast Food Restaurant	As Restau
4th Most Common Venue	Tea Room	Fast Food Restaurant	Dance Studio	Gun Range	Auto Workshop	Flower Shop	Coffee Shop	Sandwich Place	Theme Park	Gym / Fitness Center	...	Spa	F Chic J
5th Most Common Venue	Zoo Exhibit	Filipino Restaurant	Farm	Coffee Shop	Home Service	Farmers Market	Restaurant	Donut Shop	Photography Studio	Bowling Alley	...	Mexican Restaurant	Disc St
6th Most Common Venue	Fishing Store	Fish & Chips Shop	Fast Food Restaurant	Stadium	Baseball Field	Fast Food Restaurant	Fast Food Restaurant	Pizza Place	Zoo Exhibit	Liquor Store	...	Zoo Exhibit	O Re S
7th Most Common Venue	Farmers Market	Fish Market	Filipino Restaurant	Food Truck	Food Truck	Filipino Restaurant	Electronics Store	Fried Chicken Joint	Fishing Store	Fried Chicken Joint	...	Filipino Restaurant	B
8th Most Common Venue	Fast Food Restaurant	Fishing Spot	Fish & Chips Shop	Flea Market	Fish Market	Fish & Chips Shop	Café	Convenience Store	Farm	Fountain	...	Fish & Chips Shop	Groc St
9th Most Common Venue	Filipino Restaurant	Fishing Store	Fish Market	Filipino Restaurant	Farm	Fish Market	Theater	Intersection	Farmers Market	Fishing Store	...	Fish Market	Sandv Pl
10th Most Common Venue	Fish & Chips Shop	Flea Market	Fishing Spot	Frozen Yogurt Shop	Farmers Market	Fishing Spot	Convenience Store	Fountain	Fast Food Restaurant	Fast Food Restaurant	...	Fishing Spot	Fast F Restau

12 rows × 114 columns

3.3 Cluster 3

Here we have 10 neighborhoods. The description of the metrics and venue categories are given by Figures 7 and 8. The distinctive features of this neighboorhood are as follows:

1. Population have high income. More than 75% earned above 7,000 $. They are the elite class
2. Houses have a minimum price of 400,000 $. It is the cluster with the highest set of prices.
3. Houses turn to be very old and neighborhoods have a lot of venues.

The neighborhoods are displayed in Figures 1 and 2 with dark blue bullets

Figure 7: Description of metrics: Cluster 3

	count	mean	std	min	25%	50%	75%	max
Cluster Labels	10.0	2.00000	0.000000	2.0000	2.00000	2.00000	2.0000	2.0000
housing_median_age	10.0	20.00000	12.373807	4.0000	10.00000	20.50000	22.7500	45.0000
population	10.0	5907.10000	2444.188866	4076.0000	4525.75000	5083.50000	6102.7500	12203.0000
households	10.0	2269.15000	1217.958630	1165.0000	1606.00000	1791.25000	2682.7500	4930.0000
median_income	10.0	6.80733	2.008536	3.4051	5.36715	7.48865	8.3193	9.1232
median_house_value	10.0	473915.45000	29347.618051	415300.0000	453112.62500	481600.00000	500001.0000	500001.0000
<1H OCEAN	10.0	0.50000	0.527046	0.0000	0.00000	0.50000	1.0000	1.0000
INLAND	10.0	0.00000	0.000000	0.0000	0.00000	0.00000	0.0000	0.0000
NEAR BAY	10.0	0.10000	0.316228	0.0000	0.00000	0.00000	0.0000	1.0000
NEAR OCEAN	10.0	0.40000	0.516398	0.0000	0.00000	0.00000	1.0000	1.0000
Num Venues	10.0	17.40000	22.207106	1.0000	3.00000	9.50000	16.0000	64.0000

Figure 8: Neighborhood Description: Cluster 3

	7	16	60	179	197	208	214	215	340	431
Neighborhood	9	19	68	223	243	256	263	264	399	516
Cluster Labels	2	2	2	2	2	2	2	2	2	2
1st Most Common Venue	Fishing Spot	Park	Coffee Shop	Park	Mobile Phone Shop	Harbor / Marina	Theater	Spa	Bank	Pool
2nd Most Common Venue	Scenic Lookout	Trail	Gym	Flower Shop	Yoga Studio	Boat or Ferry	Park	Gym	Japanese Restaurant	Arcade
3rd Most Common Venue	Trail	Athletics & Sports	Park	Farmers Market	Sushi Restaurant	Pizza Place	Fast Food Restaurant	Nightlife Spot	Sandwich Place	Zoo Exhibit
4th Most Common Venue	Gift Shop	Flea Market	Church	Fast Food Restaurant	Bank	Spa	Trail	Zoo Exhibit	Coffee Shop	Flower Shop
5th Most Common Venue	Tourist Information Center	Farmers Market	Optical Shop	Filipino Restaurant	Chinese Restaurant	Coffee Shop	Performing Arts Venue	Flea Market	Restaurant	Farmers Market
6th Most Common Venue	Park	Fast Food Restaurant	Flower Shop	Fish & Chips Shop	Japanese Restaurant	Beach	Shopping Mall	Farmers Market	Mobile Phone Shop	Fast Food Restaurant
7th Most Common Venue	Beach	Filipino Restaurant	Sandwich Place	Fish Market	Sandwich Place	ATM	Video Store	Fast Food Restaurant	Juice Bar	Filipino Restaurant
8th Most Common Venue	Historic Site	Fish & Chips Shop	Art Gallery	Fishing Spot	Frozen Yogurt Shop	Burger Joint	Carpet Store	Filipino Restaurant	Pizza Place	Fish & Chips Shop
9th Most Common Venue	Harbor / Marina	Fish Market	Taiwanese Restaurant	Fishing Store	American Restaurant	Bakery	Fish & Chips Shop	Fish & Chips Shop	Seafood Restaurant	Fish Market
10th Most Common Venue	Bus Station	Fishing Spot	Convenience Store	Flea Market	Coffee Shop	Bank	Event Space	Fish Market	Ice Cream Shop	Fishing Spot

3.4 Cluster 4

Here we have 187 neighborhoods.Figures 9 and 10 give a description of the metrics and venue categories of the neighborhoods. The distinctive features are as follows:

1. The population has median income distribution slightly higher to those in cluster 2. They are thus from the medium class.
2. Here house prices lies between those of cluster 2 and 1.
3. The neighborhoods are all situated less than 1 hour from the ocean
4. Every neighboorhood has either a Filipino restaurant or a mexican restaurant. This suggest inhabitants in this region may surely be foreigners.

Figure 9: Description of metrics: Cluster 4

	count	mean	std	min	25%	50%	75%	max
Cluster Labels	187.0	3.000000	0.000000	3.0000	3.00000	3.0000	3.0000	3.0000
housing_median_age	187.0	16.015152	9.626535	2.0000	7.00000	16.0000	23.0000	47.0000
population	187.0	5574.563280	1680.634965	4010.0000	4452.50000	5008.0000	6123.0000	16305.0000
households	187.0	1757.674688	718.053191	275.0000	1297.50000	1581.0000	1964.5000	5358.0000
median_income	187.0	3.675156	1.056732	1.4133	2.82755	3.6348	4.5426	6.4865
median_house_value	187.0	171790.196078	18569.840292	140800.0000	156650.00000	170700.0000	187500.0000	208000.0000
<1H OCEAN	187.0	0.556150	0.498171	0.0000	0.00000	1.0000	1.0000	1.0000
INLAND	187.0	0.336898	0.473919	0.0000	0.00000	0.0000	1.0000	1.0000
NEAR BAY	187.0	0.037433	0.190330	0.0000	0.00000	0.0000	0.0000	1.0000
NEAR OCEAN	187.0	0.069519	0.255017	0.0000	0.00000	0.0000	0.0000	1.0000
Num Venues	187.0	9.395722	12.986908	1.0000	3.00000	4.0000	10.0000	99.0000

Figure 10: Neighborhood Description: Cluster 4

	5	14	20	24	32	33	35	51	54	55	...	447	453
Neighborhood	7	17	24	28	36	37	40	57	60	62	...	537	544
Cluster Labels	3	3	3	3	3	3	3	3	3	3	...	3	3
1st Most Common Venue	Park	Coffee Shop	Farm	Park	Food Truck	Fast Food Restaurant	Chinese Restaurant	Snack Place	Park	Park	...	Ice Cream Shop	Pool
2nd Most Common Venue	Harbor / Marina	Park	Music Venue	Theater	Coffee Shop	Chinese Restaurant	Zoo Exhibit	Park	Electronics Store	Flower Shop	...	Zoo Exhibit	Child Care Service
3rd Most Common Venue	Farm	Motorcycle Shop	Baseball Field	Home Service	Fast Food Restaurant	Intersection	Exhibit	Nightlife Spot	Exhibit	Farmers Market	...	Flower Shop	Football Stadium
4th Most Common Venue	Farmers Market	Bookstore	Flower Shop	Fishing Store	Soup Place	ATM	Farmers Market	Farm	Farmers Market	Fast Food Restaurant	...	Farmers Market	Trail
5th Most Common Venue	Fast Food Restaurant	Zoo Exhibit	Fast Food Restaurant	Farm	Gym	Pharmacy	Fast Food Restaurant	Fast Food Restaurant	Fast Food Restaurant	Filipino Restaurant	...	Fast Food Restaurant	Zoo Exhibit
6th Most Common Venue	Filipino Restaurant	Filipino Restaurant	Filipino Restaurant	Farmers Market	Grocery Store	Grocery Store	Filipino Restaurant	Filipino Restaurant	Filipino Restaurant	Fish & Chips Shop	...	Filipino Restaurant	Fishing Store
7th Most Common Venue	Fish & Chips Shop	Fish & Chips Shop	Fish & Chips Shop	Fast Food Restaurant	Big Box Store	Tennis Court	Fish & Chips Shop	Fish & Chips Shop	Fish & Chips Shop	Fish Market	...	Fish & Chips Shop	Farmers Market
8th Most Common Venue	Fish Market	Fish Market	Fish Market	Filipino Restaurant	Taco Place	Liquor Store	Fish Market	Fish Market	Fish Market	Fishing Spot	...	Fish Market	Fast Food Restaurant
9th Most Common Venue	Fishing Spot	Fishing Spot	Fishing Spot	Fish & Chips Shop	Supermarket	Farmers Market	Fishing Spot	Fishing Spot	Fishing Spot	Fishing Store	...	Fishing Spot	Filipino Restaurant
10th Most Common Venue	Fishing Store	Fishing Store	Fishing Store	Fish Market	Sporting Goods Shop	Sandwich Place	Fishing Store	Fishing Store	Fishing Store	Flea Market	...	Fishing Store	Fish & Chips Shop

12 rows × 187 columns

3.5 Cluster 5

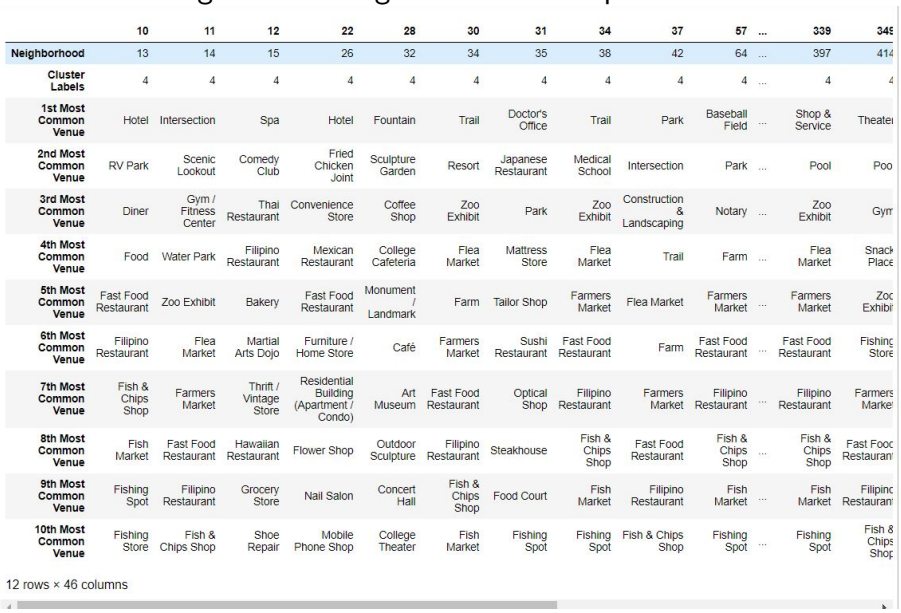
Here we have 46 neighborhoods. Figures 11 and 12 provide a description of the metrics and venue categories. The neighborhoods are displayed in Figures 1 and 2 with dark orange bullets. The distinctive features are :

1. House prices lies between those of cluster 1 and cluster 3. They are expensive
2. 75% of the population has median income above 4,000 $. This population can be qualified as high class.

Figure 11: Description of metrics: Cluster 5

	count	mean	std	min	25%	50%	75%	max
Cluster Labels	46.0	4.000000	0.000000	4.0000	4.00000	4.00000	4.00000	4.0000
housing_median_age	46.0	14.195652	8.898488	3.0000	6.00000	13.00000	21.00000	40.0000
population	46.0	5866.739130	2063.488475	4005.0000	4304.25000	5002.00000	6851.50000	12873.0000
households	46.0	2092.554348	755.696016	950.0000	1504.25000	1947.00000	2521.00000	4012.0000
median_income	46.0	5.700546	1.570974	1.6903	4.91825	6.12045	6.60945	8.3931
median_house_value	46.0	338433.695652	28986.050389	293900.0000	315325.00000	331900.00000	357375.00000	402500.0000
<1H OCEAN	46.0	0.521739	0.505047	0.0000	0.00000	1.00000	1.00000	1.0000
INLAND	46.0	0.130435	0.340503	0.0000	0.00000	0.00000	0.00000	1.0000
NEAR BAY	46.0	0.086957	0.284885	0.0000	0.00000	0.00000	0.00000	1.0000
NEAR OCEAN	46.0	0.260870	0.443961	0.0000	0.00000	0.00000	0.75000	1.0000
Num Venues	46.0	9.543478	12.648595	1.0000	2.00000	4.00000	12.50000	51.0000

Figure 12: Neighborhood Description: Cluster 5



	10	11	12	22	28	30	31	34	37	57	...	339	345
Neighborhood	13	14	15	26	32	34	35	38	42	64	...	397	414
Cluster Labels	4	4	4	4	4	4	4	4	4	4	...	4	4
1st Most Common Venue	Hotel	Intersection	Spa	Hotel	Fountain	Trail	Doctor's Office	Trail	Park	Baseball Field	...	Shop & Service	Theater
2nd Most Common Venue	RV Park	Scenic Lookout	Comedy Club	Fried Chicken Joint	Sculpture Garden	Resort	Japanese Restaurant	Medical School	Intersection	Park	...	Pool	Poo
3rd Most Common Venue	Diner	Gym / Fitness Center	Thai Restaurant	Convenience Store	Coffee Shop	Zoo Exhibit	Park	Zoo Exhibit	Construction & Landscaping	Notary	...	Zoo Exhibit	Gym
4th Most Common Venue	Food	Water Park	Filipino Restaurant	Mexican Restaurant	College Cafeteria	Flea Market	Mattress Store	Flea Market	Trail	Farm	...	Flea Market	Snack Place
5th Most Common Venue	Fast Food Restaurant	Zoo Exhibit	Bakery	Fast Food Restaurant	Monument / Landmark	Farm	Tailor Shop	Farmers Market	Flea Market	Farmers Market	...	Farmers Market	Zoo Exhibit
6th Most Common Venue	Filipino Restaurant	Flea Market	Martial Arts Dojo	Furniture / Home Store	Café	Farmers Market	Sushi Restaurant	Fast Food Restaurant	Farm	Fast Food Restaurant	...	Fast Food Restaurant	Fishing Store
7th Most Common Venue	Fish & Chips Shop	Farmers Market	Thrift / Vintage Store	Residential Building (Apartment / Condo)	Art Museum	Fast Food Restaurant	Optical Shop	Filipino Restaurant	Farmers Market	Filipino Restaurant	...	Filipino Restaurant	Farmers Market
8th Most Common Venue	Fish Market	Fast Food Restaurant	Hawaiian Restaurant	Flower Shop	Outdoor Sculpture	Filipino Restaurant	Steakhouse	Fish & Chips Shop	Fast Food Restaurant	Fish & Chips Shop	...	Fish & Chips Shop	Fast Food Restaurant
9th Most Common Venue	Fishing Spot	Filipino Restaurant	Grocery Store	Nail Salon	Concert Hall	Fish & Chips Shop	Food Court	Fish Market	Filipino Restaurant	Fish Market	...	Fish Market	Filipino Restaurant
10th Most Common Venue	Fishing Store	Fish & Chips Shop	Shoe Repair	Mobile Phone Shop	College Theater	Fish Market	Fishing Spot	Fishing Spot	Fish & Chips Shop	Fishing Spot	...	Fishing Spot	Fish & Chips Shop

12 rows × 46 columns

4 Conclusion and Future directions

In this work, we have obtained a clustering of the neighborhoods of the city of California. We find five clusters which correspond to the five classes of the society : low class (cluster 2), medium class (cluster 4), middle class (cluster 1), high class (cluster 5) and elite class (cluster 3). The housing prices in the clusters reflect the class category of the clusters.
The clustering can be improved by considering the venue categories as features. Since there are over 349 categories, the best possible step is to group the categories into further small groups. This can be achieve using PCA or classification. For example, we can consider all restaurant types as a single category, all distraction venues as entertainment, all markets as simply market.