

Battle of the Neighborhoods

**Identification of the most suitable location to open a yoga studio in
the Sacramento County, California, USA**

Introduction

- Given the high cost of joining a yoga studio, it is reasonable to guess that it needs to be opened in an area with high population and median income.
- The population of the Sacramento County is 1,418,788 which is distributed in 54 zip codes.
- Manual search of an area most ideal for opening a yoga studio is a daunting task.
- Data analysis techniques can be used to reach the conclusion swiftly.

Acquiring the Data

- Following information was used for the analysis:
 - 1) Population, median age, median income and house price in each zip code
 - 2) Latitude and longitude of each zip code
 - 3) Latitude and longitude of gyms throughout the Sacramento County.
- Data 1 was scrapped from
<https://www.bestplaces.net/find/zip.aspx?st=CA&county=06067>
- Data 2, in the form of GeoJson file, was downloaded from
<https://geo.nyu.edu/catalog/stanford-cm480wn0393>
- Data 3 was acquired from Foursquare.

Methodology

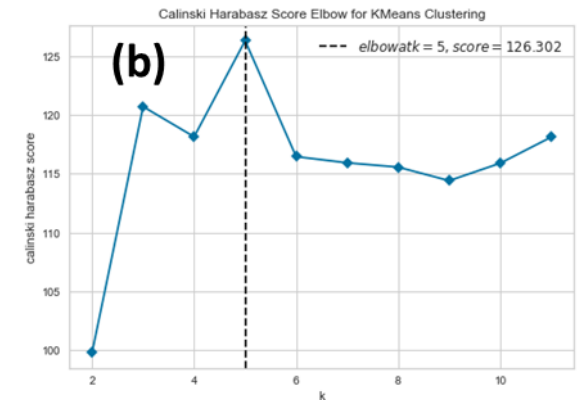
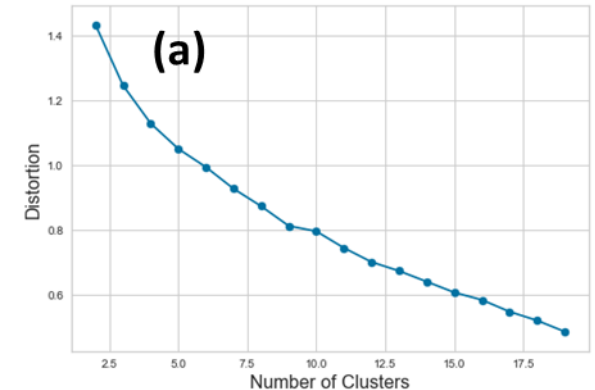
- The best way to obtain the ideal locations is to run the K-Mean clustering on standardized socio-economic database, i.e. data 1.
- After shortlisting a group of zip codes with required properties, the ideal zip code can be pinpointed by using location information of gyms and yoga centers from data 3.
- Zip code with high population and median income, not so high median age and no nearby gyms, yoga studios and fitness centers will be most ideal.
- It was assumed that gyms and fitness centers also offer yoga classes.

Results of K-Means Clustering

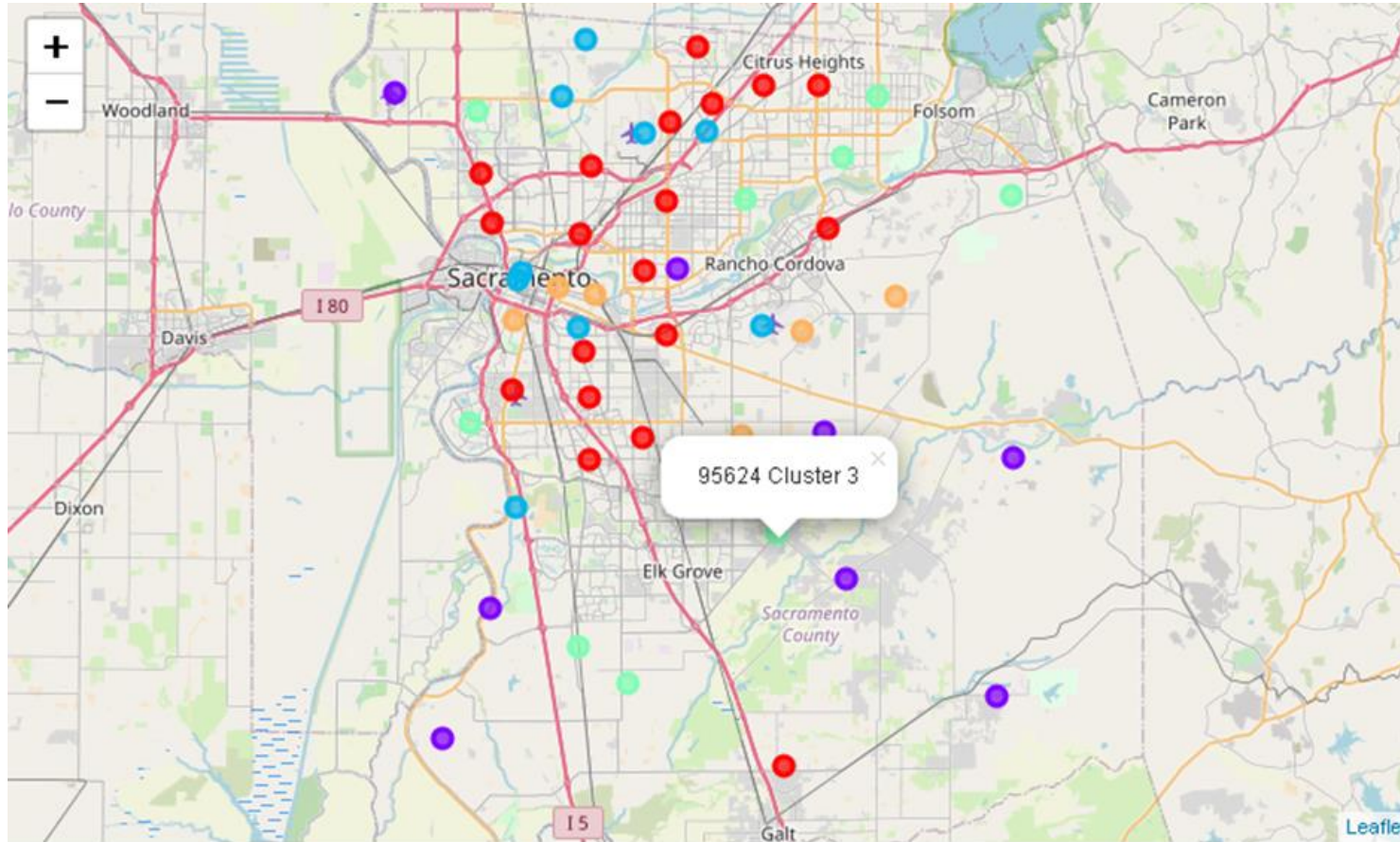
- Data 1, after cleaning, took the following form:

	Zipcode	City	Population	Median Age	Median Income(US\$)	Median Home Price(US\$)
0	95843	antelope	47088	33.6	64773	353600
1	95864	arden-arcade	23458	46.9	83467	577100
2	95821	arden-arcade	35734	38.0	38777	354600
3	95825	arden-arcade	34201	31.8	37382	313000
4	95608	carmichael	61365	43.4	55256	410400

- 'Zipcode' and 'City' columns were dropped.
- The resulting data frame was standardized.
- The best number of clusters was found to be 5, using the 'elbow' method.



Results of K-Means Clustering



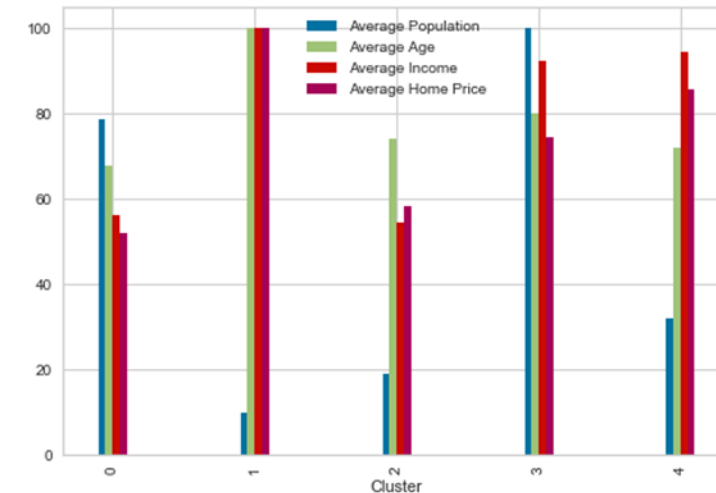
The distribution of clusters throughout the Sacramento County

Results of K-Means Clustering

- The average value of each column of each cluster was calculated to get the better picture of the cluster
- Each column was normalized to 100 so that all the columns can be visualized in one plot.

	Average Population	Average Age	Average Income	Average Home Price
Cluster				
0	78.778193	67.953627	56.376584	52.124399
1	9.792645	100.000000	100.000000	100.000000
2	19.152519	74.022716	54.387279	58.211608
3	100.000000	80.057981	92.314688	74.613966
4	31.979972	71.951831	94.583521	85.717101

Data frame with normalized average values.



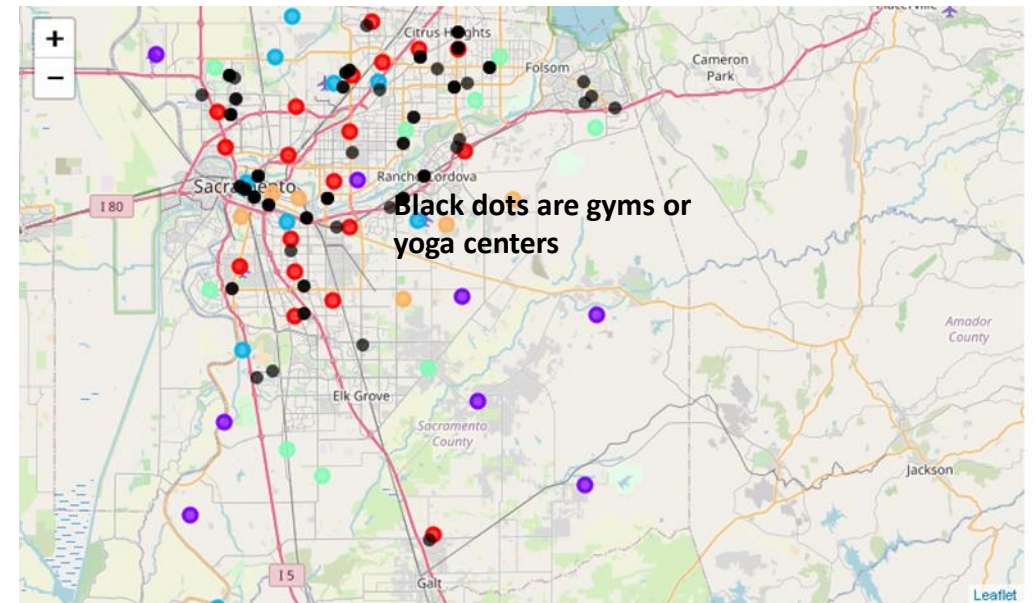
A bar graph of average values of population, age, income and home price of each cluster.

Zip codes belonging to cluster 3 have the most desirable properties

The use of the Foursquare Data

- The venue information was gathered within 5000 m radius of each zip code location.
- The number of yoga centers and gyms was calculated for each zip code.
- The yoga centers and gym locations were superimposed to the cluster map.

	Zipcode	Total Yoga Places
0	95843	2
1	95864	2
2	95821	2
3	95825	1
4	95608	3



Merging of Socio-Economic and Foursquare Data

- The total number of yoga and gyms were calculated for each zip codes of cluster 3.

	Cluster Labels	Zipcode	City	Population	Median Age	Median Income(US\$)	Median Home Price(US\$)	Longitude	Latitude	Total Yoga Places
0	3	95608	carmichael	61365	43.4	55256	410400	-121.324950	38.626384	3
1	3	95624	elk	65447	36.8	80233	435900	-121.301845	38.431007	0
2	3	95757	elk	48242	34.7	87782	492400	-121.412077	38.342434	0
3	3	95758	elk	63977	36.0	70003	400000	-121.449334	38.364344	0
4	3	95628	fair	40882	45.5	72739	438200	-121.252645	38.651726	4
5	3	95630	folsom	75864	40.9	100308	554200	-121.125443	38.628682	4
6	3	95662	orangevale	31738	40.9	65122	390800	-121.225283	38.686473	4
7	3	95831	sacramento	42218	44.2	68679	452800	-121.531206	38.495147	1
8	3	95835	sacramento	38607	36.6	80449	400900	-121.525621	38.677556	4

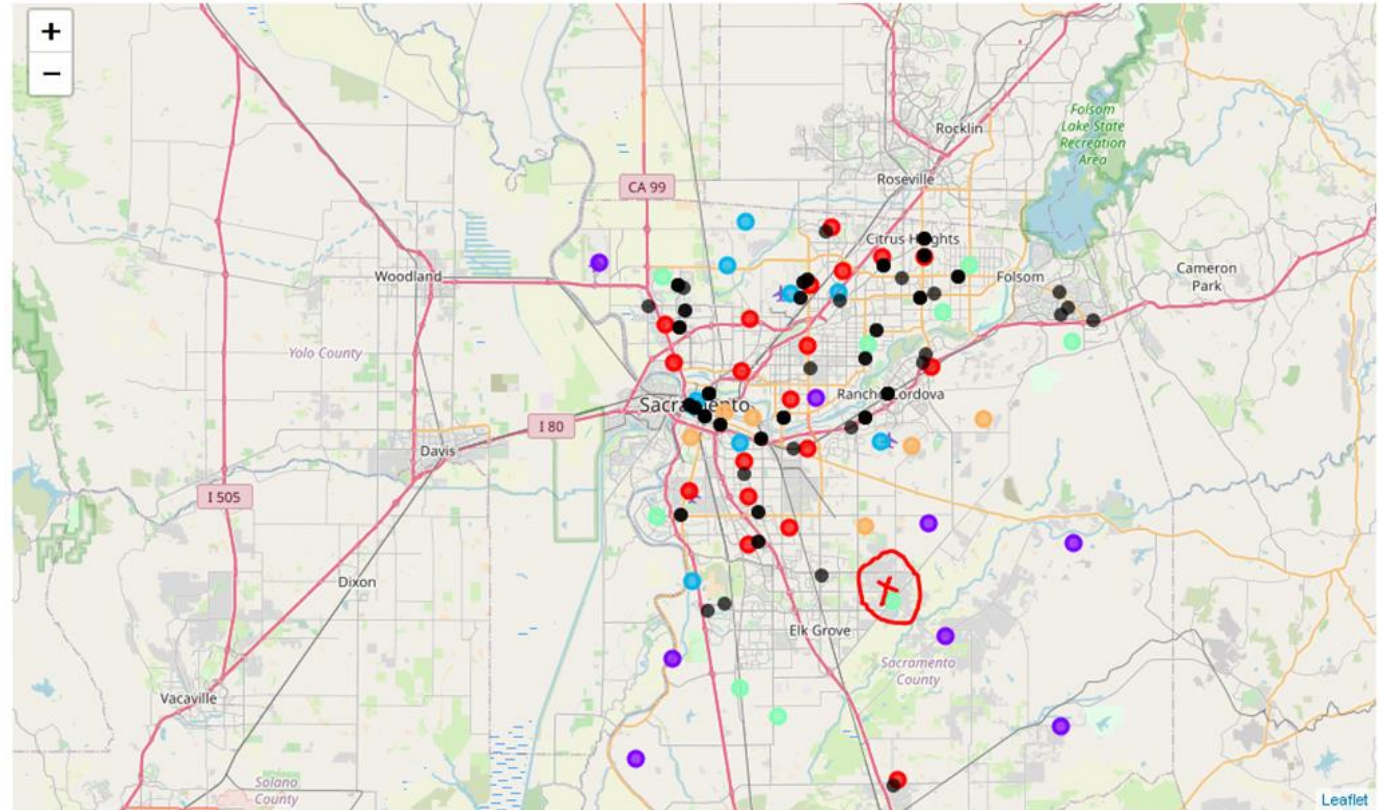
- This table reveals that the city of Elk Groove has no yoga studios and gyms.

Final Result

➤ Based on the data of previous slide, it can be said that the best zip code to open a yoga studio is 95624 in Elk Grove.

➤ This particular location is most ideal of the three because it has the highest population and the second highest median income. The median age is also ideal for people who are most likely to join a yoga studio.

The zip code is marked with a red cross inside a red circle in the adjacent figure



Conclusion

- The aim of this project was to suggest the best area to open a yoga studio.
- Using the K-Means Clustering, a group of zip codes was identified with high population and high median income.
- Using the Foursquare data, total number of yoga studios, gyms and fitness centers were calculated for each zip code
- **Based on these results, 95624 zip code in Elk Groove was found to be most suitable to open a yoga studio.**
- The analysis can be further improved by identifying which gyms and fitness centers really provide yoga classes and which zip codes has proper infrastructures.