

Capstone Project Notebook: Clustering venues by Earthquakes Magnitude in CA, US.

Author: Pamela Denisse Lopez Mata

This notebook is for the "Applied Data Science Capstone" course activities of the IBM-Data-Science-Specialization.

Table of contents

- [1. Introduction: Business Problem](#)
- [2. Data](#)
- [3. Methodology](#)
- [4. Results](#)
- [5. Discussion](#)
- [6. Conclusion](#)
- [7. References](#)

1. Introduction: Business Problem

Earthquake is a term used to describe both sudden slip on a fault, and the resulting ground shaking and radiated seismic energy caused by the slip, or by volcanic or magmatic activity, or other sudden stress changes in the earth. (1)

Earthquake hazard is anything associated with an earthquake that may affect the normal activities of people. This includes surface faulting, ground shaking, landslide, liquefaction, tectonic deformation, tsunamis, and seiches. (2)

Earthquake risk is the probable building damage, and number of people that are expected to be hurt or killed if a likely earthquake on a particular fault occurs. (3)

When you plan to invest in cities located along the Pacific Ocean, these are terms that should be seriously considered before making a decision, especially in cities with as much economic, tourist and cultural relevance as California. However, not only investors are interested in this issue, the government itself is concerned with mitigating as much as possible the greatest amount of risk under which its population is.

Considering both the government and investors and their need to have an updated knowledge of the latent risk of California, this work focuses on determining the risk under which the most important establishments are located in some of the most important areas of California.

2. Data

In order to achieve the analysis proposed, it will be necessary to consult the earthquake databases that the state institutions of the state of California have, such as the U.S. Geological Survey (USGS) portal of USGS Earthquake Hazards Program (<https://earthquake.usgs.gov/data/data.php>) or the Southern California Earthquake Data Center (SCEDC) (<https://scedc.caltech.edu/about/index.html>) for example. This information on earthquakes and risk areas could spatially locate the places of interest and analyze any would be at greater risk.

In addition, the first 100 venues information will be downloaded from Foursquare, 1.5 Km around each epicenter.

2.1. Load SM2.5+ Earthquakes Past 30 Days Data

127 earthquakes magnitude 2.5+ from last 30 days (Table 1).

Table 1

	id	mag	place	title	longitude	latitude
0	ci39120144	2.73	13km ENE of Ridgecrest, CA	M 2.7 - 13km ENE of Ridgecrest, CA	-117.551333	35.676333
1	nc73288720	2.94	16km NW of Pinnacles, CA	M 2.9 - 16km NW of Pinnacles, CA	-121.259499	36.638668
2	nc73288615	2.69	32km SSE of Somes Bar, CA	M 2.7 - 32km SSE of Somes Bar, CA	-123.294998	41.127998
3	ci39117528	2.59	22km ESE of Little Lake, CA	M 2.6 - 22km ESE of Little Lake, CA	-117.682667	35.850833
4	ci39116408	2.51	36km NW of Baker, CA	M 2.5 - 36km NW of Baker, CA	-116.316500	35.520833

2.2. Load venues in California, US from Foursquare

234 venues were returned by Foursquare (Table 2).

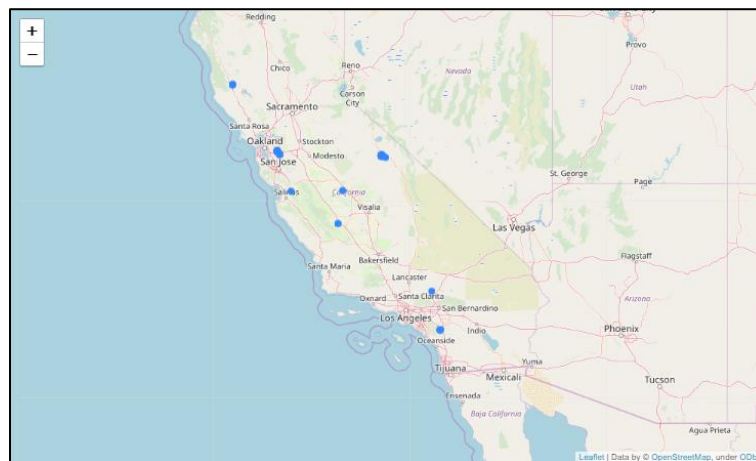
Table 2

	Earthquake	Earthquake Magnitude	Earthquake Place	Earthquake Title	Earthquake Latitude	Earthquake Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	nc73287360	2.54	3km E of Mammoth Lakes, CA	M 2.5 - 3km E of Mammoth Lakes, CA	37.640999	-118.950333	SKADI	37.637522	-118.965933	Modern European Restaurant
1	nc73287360	2.54	3km E of Mammoth Lakes, CA	M 2.5 - 3km E of Mammoth Lakes, CA	37.640999	-118.950333	Roberto's Cafe	37.641956	-118.966163	Mexican Restaurant
2	nc73287360	2.54	3km E of Mammoth Lakes, CA	M 2.5 - 3km E of Mammoth Lakes, CA	37.640999	-118.950333	Mammoth Rock-N-Bowl	37.636578	-118.964965	Bowling Alley
3	nc73287360	2.54	3km E of Mammoth Lakes, CA	M 2.5 - 3km E of Mammoth Lakes, CA	37.640999	-118.950333	Mammoth Tavern	37.637532	-118.965910	Pub
4	nc73287360	2.54	3km E of Mammoth Lakes, CA	M 2.5 - 3km E of Mammoth Lakes, CA	37.640999	-118.950333	Mimi's Cookie Bar	37.637504	-118.966654	Bakery

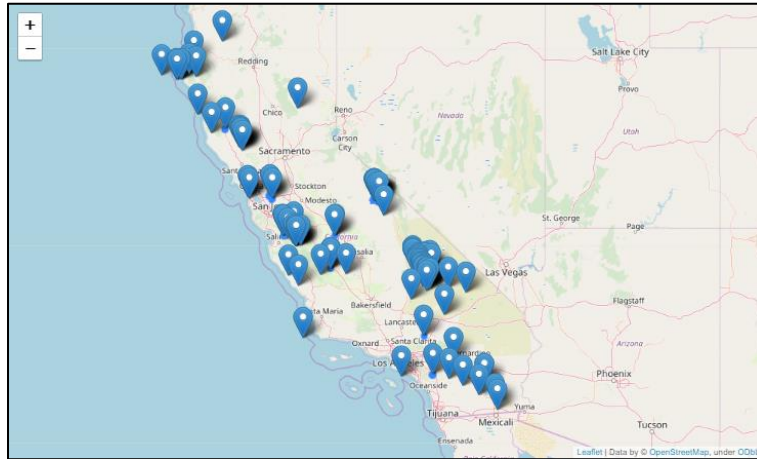
3. Methodology

3.1. Data Visualization

3.1.1. Venues data map



3.1.2. Earthquakes data map



3.2. California's venue data preprocessing

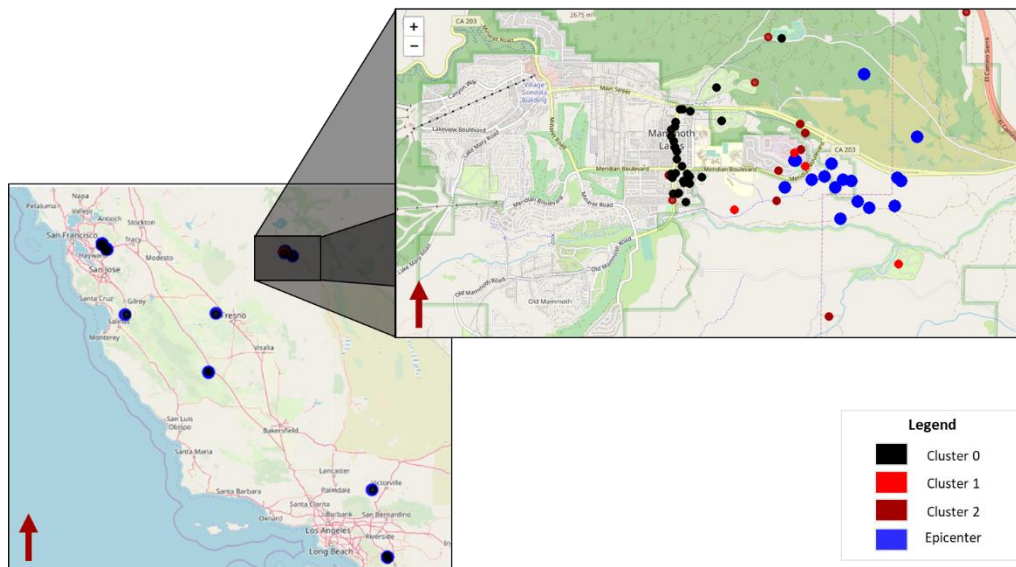
Manipulate data in order to know by how many earthquakes epicenters a venue is surrounded on 1.5 Km. 5th most common Earthquake Magnitude by Venue (Table 3).

Table 3

	Venue	1st Most Common Magnitude	2nd Most Common Magnitude	3rd Most Common Magnitude	4th Most Common Magnitude	5th Most Common Magnitude
0	7-Eleven	2.91	3.96	2.73	2.47	2.48
1	9110 Nail Salon	2.74	2.91	2.73	2.47	2.48
2	Amador Lakes Gym	2.91	3.96	2.73	2.47	2.48
3	Amador Rancho Park	2.74	2.91	2.73	2.47	2.48
4	Angelina's Cafe	2.47	3.96	3.14	2.48	2.5

Until now, it is possible to identify how many of each earthquake magnitudes have had effect on each venue. So, the next step is to know the top 5 for each one.

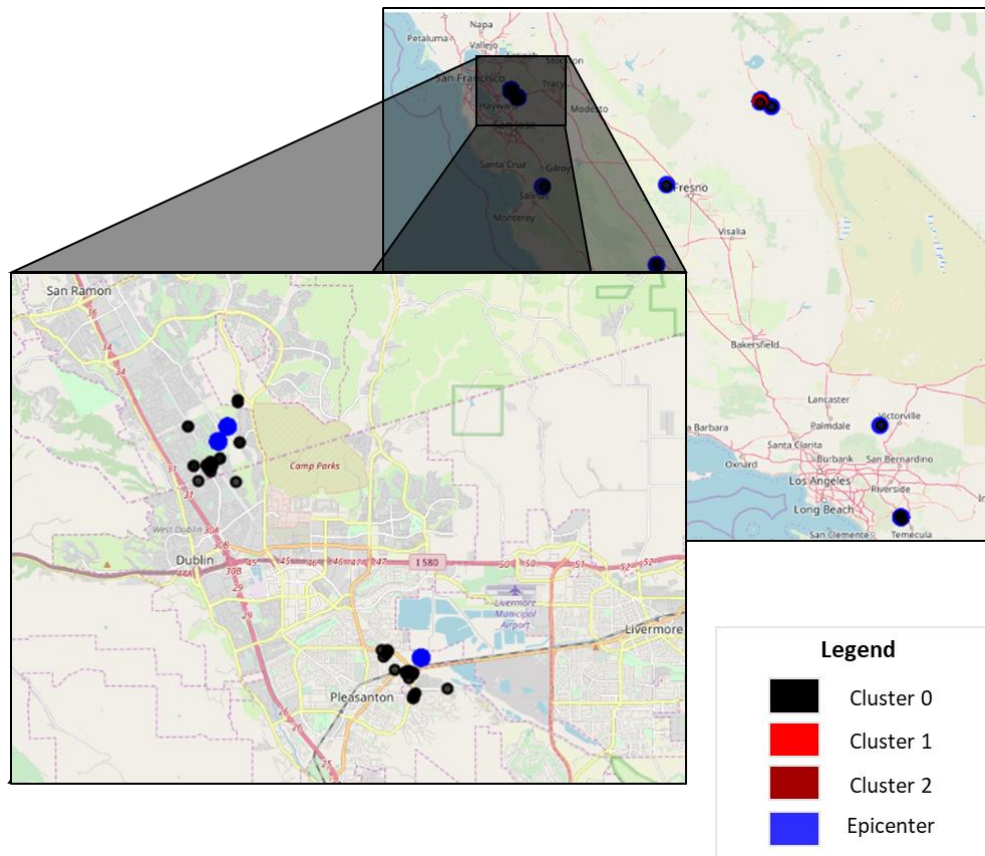
4. Results



5. Discussion

During the execution of this code there were many connection issues, so from the 50 Km considered in the beginning, the venues consulting was restricted to only 1.5 KM, that's why at first sight, the venues analyzed are almost over the epicenter's points.

Zooming into then, it's clearly that the black points are closest to the epicenters (in most cases to just one). Only to the East it can be seen more diversity of clusters (black, dark red and light red).



6. Conclusion

The k-means clustering identified on the map that the venues nearest to the epicenters are in the Cluster 0, while the less common are Cluster 1 and Cluster 2.

It is proposed that for better results, the venues consulting should include a greater radius value and it can also be included other earthquake magnitudes like 3+ or 5+, to have a better understanding of Earthquake Hazards.

7. References

1. <https://earthquake.usgs.gov/learn/glossary/?term=earthquake>
2. <https://earthquake.usgs.gov/learn/glossary/?term=earthquake%20hazard>
3. <https://earthquake.usgs.gov/learn/glossary/?term=earthquake%20risk>
4. <https://earthquake.usgs.gov/earthquakes/feed/v1.0/geojson.php>