Existing company risk assessment caused by earthquakes in Sicily

Amine Ben Moussa

February 24, 2021

1. Introduction

1.1 Background

Evaluating the insurance risk for a bank is essential in order to prevent the risk of not re-entering the expenses due to those revenues. In all of Italy there are two regions with a high seismic risk.
One area of Lazio, one is the Sicilian area. Knowing in advance when a company is at risk of earthquake damage, not that knowing which companies at risk have not yet taken out an insurance policy is essential to obtain protection for the bank, but also a possible package of new promising customers.

1.2 Problem

The earthquake detection data can help to understand where and how much a particular earthquake is hit in the Sicilian region, possibly knowing how much a given activity is actually subject to risk

1.3 Interest

Obviously the banks that issue insurance, as well as the insurance companies, are very interested in knowing the history of the earthquakes and starting from the precise location of the activity, the possible risk that the same has risked, so as to be able to draw up a truthful plan, or if not present, the possibility of acquiring a new interested customer

2. Data acquisition and cleaning

2.1 Data sources

the data is taken from safe and insured channels:

- List of earthquakes: cnt.rm.ingv.it.
- List of critical earthquakes in Sicily: wikipedia.org.

2.2 Data cleaning

Once the data has been collected, we will obtain all the earthquakes that have occurred in Italy from 1985 to now. Obviously this is not of our interest, in addition to the huge amount of data, we will also have data not related to the region that interests us, Sicily.
We will start by creating a rectangle delimited to the island, with a good margin, as earthquakes 50-60 km away can cause damage if at the level of the earth's surface.
Having thus obtained the data relating to the region concerned, let's consider an earthquake that caused extensive damage, or the Amatrice earthquake on 24 August 2016.
We immediately see that the earthquake was about 8 km away from the earth's crust, and with a magnitude of 6. To reduce the exorbitant number of events, we remove all earthquakes with a depth greater than 80km, therefore double compared to Amatrice than at the epicenter

2.3 Feature selection

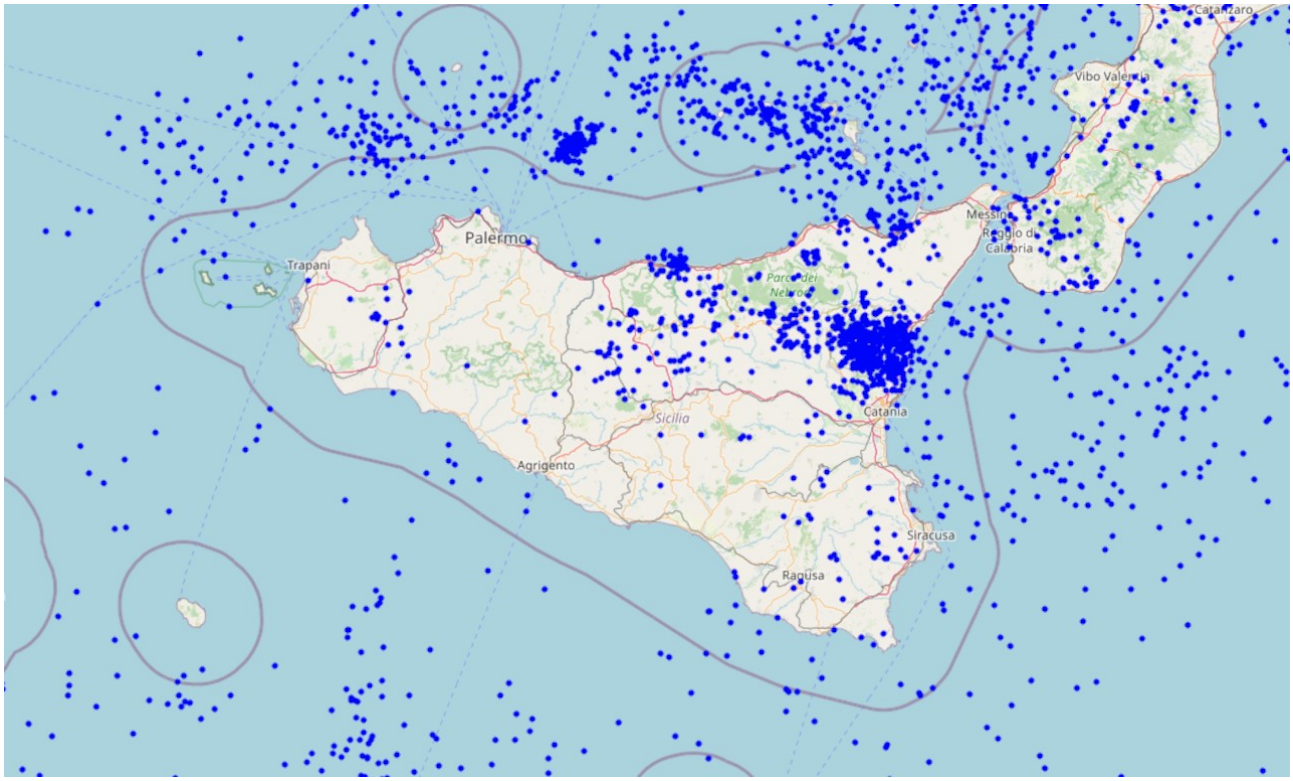
of the remaining 1793 events we say of great importance, we take only the information of our interest for the calculations.
Or:

- Date earthquake
- Latitude
- Longitude
- Depth in km
- Magnitude
- Event location

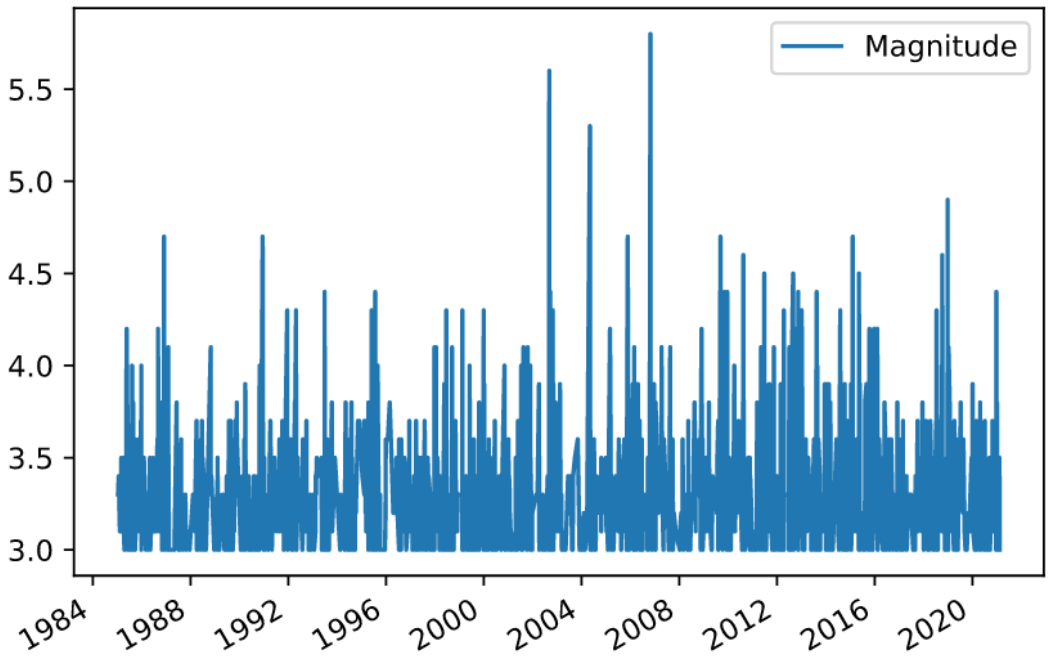
3. Exploratory Data Analysis

3.1 Data visualization and analysis

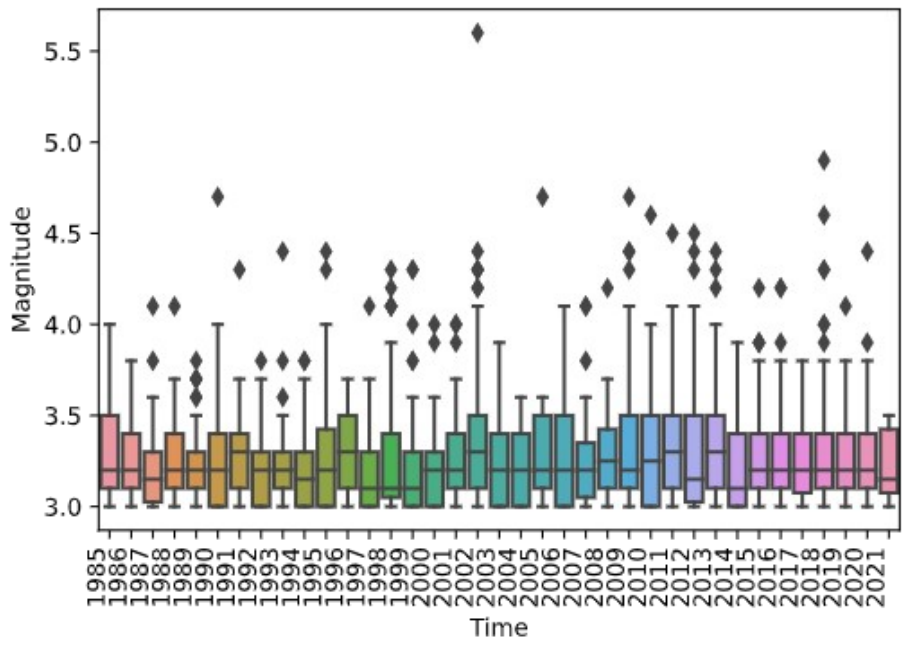
It is clearly seen from the distribution of the points that the epicenters of the earthquakes are mainly arranged around Sicily or grouped in the vicinity of the Etna volcano.



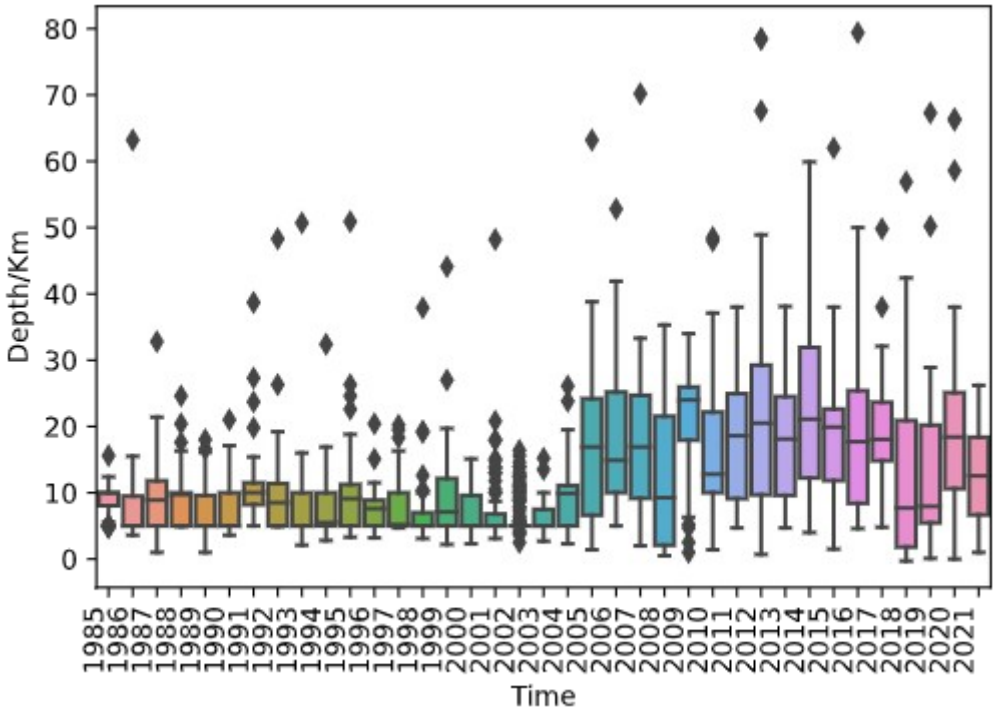
the temporal distribution of earthquakes is fairly regular over the years apart from some sporadic peaks, as can be seen from wikipedia, represent more unique than rare cases



Now we can see the plots for the earthquake averages, both for depth and intensity



It can be seen from the annual magnitude that over 75% of earthquakes do not exceed 4, while cases greater than 5 are extremely rare.
The average of earthquakes is around 3.28 in magnitude and has remained stable over the years



While the second graph shows that not only the average depth is increasing, compared to 1980-1990, but also that the out-line distribution is decreasing moving towards an average of 20km underground compared to 5-6k until the 2000s.

So the situation is apparently improving, as the deeper the epicenter is, the greater the distance from any point on the surface

X < 2005	Depth/Km 17.840972 Magnitude 3.3059720
X >= 2005	Depth/Km 7.909506 Magnitude 3.263747

3.3 Using FourSquare

Now with the data we have collected we can find and classify all the activities adjacent to the place of the epicenter, also knowing how much it was hit and with what intensity

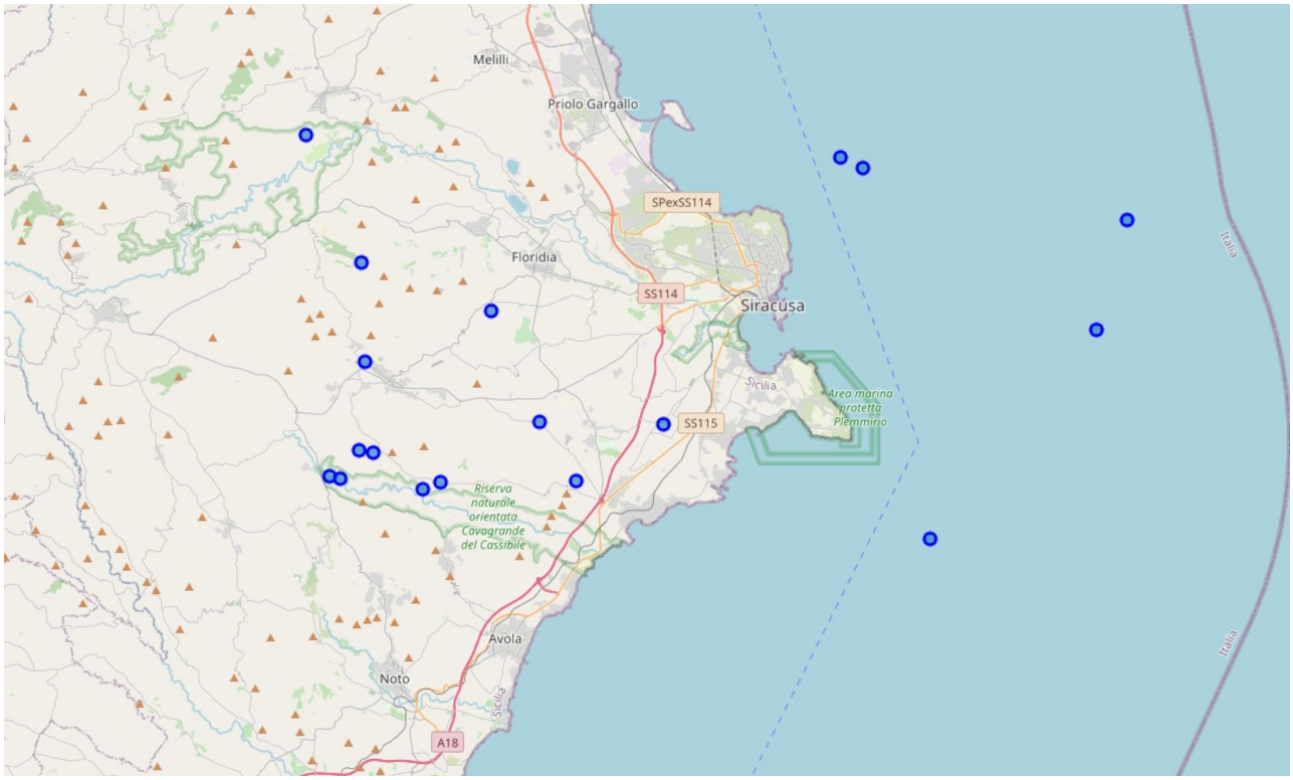
We will create a table based on the Amatrice earthquake for the extent of the damage and the relationship Zone -> Company

X <= 40km from the epicenter	RED
40km > X >= 60km from the epicenter	YELLOW
X >60km from the epicenter	GREEN

We will mostly open how many times a certain activity has been hit by an earthquake and how often (RYG)
Once we have obtained all the activities near or far from possible earthquakes, we can group them, at the beginning we will have more than 79 thousand records, through the grouping we go down to only about 1800 activities.
Now for each activity we check the distance from the epicenter of each earthquake, discarding the insertion if it is greater than 100 km since the average magnitude is 4, and we do not have values greater than 5.5

3.4 DataSet reduction

The data currently obtained are many, and crossing them with the results from ForSquare there are over 10 million results to manage.
It was thus decided to re-elaborate the data for individual portions of the region, now we will examine the region of Syracuse to obtain information only on the activities in that region



now we can manage the input data and process it fairly quickly.

For each earthquake we will carry out checks to find the distance from the epicenter and thus classify three labels [red, yellow, green] for each earthquake.

	name	red	yellow	green
0	A Putia Do Calabrisi	1.0	15.0	2.0
1	Agua Beach	11.0	7.0	0.0
2	Antica Dolceria Bonajuto	9.0	5.0	4.0
3	Apple Centro Sicilia	1.0	14.0	3.0
4	Augustus Birreria Pizzeria	16.0	2.0	0.0
...	...	...	...	...
123	Unesco - Late Baroque Towns of the Val di Noto...	15.0	3.0	0.0
124	a putia	18.0	0.0	0.0
125	i Sapori degli Iblei	9.0	5.0	4.0
126	movimentocentrale bike café	18.0	0.0	0.0
127	pasticceria Artale siracusa	18.0	0.0	0.0

128 rows × 4 columns

Grouping the results by commercial activity, we will find in the bands the number of earthquakes and the severity of it to which it has been subjected from 1985 to today.

3.4 DataSet Classification K-mean

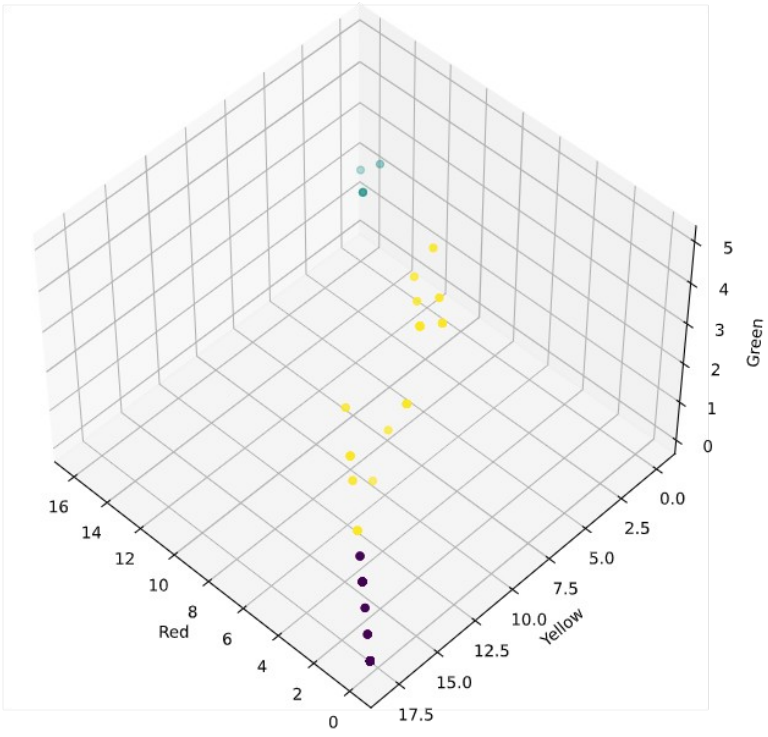
To better understand how to group activities by risk, given that there are even complex cases, where it is not clear how to classify a certain result, we will use the K-means classification.

we will set up the subdivision of the results in 3 groups so as to have a classification divided into:

- High
- Average
- Low

4. Result

4.1 Result visualization



now our dataset is ready to be searched with a clear business risk situation for each returned asset.

if an activity is located in a high risk area, it will be possible, if not present, to offer an adequate situation to protect the activity.

4.1 Future

For the future, as the average shows, the decrease in the depth of earthquakes also indicates a lower impact on activities and therefore a possible improvement in the seismic situation in Sicily.
By also having the current insurance information and any associated problems available, it would be possible to improve the risk rate and have a clearer and safer insurance status relating to the customer and neighbors for the insurer.