Final project IDS 2020/2021

Analysis and visualization of the earthquakes that occurred from 1965 to 2016

Group:
- Alex Di Stefano https://github.com/xDiste
- Christian Stingone https://github.com/ChriStingo

Language: Python

Libraries used:
- Charts: Matplotlib, plotly, GeoPandas, Wordcloud
- Data manipulation: Numpy, Pandas,
- Convert into nations: reverse_geocoder, pycountry_convert,
- Prediction: Basemap (also for a graph), Sklearn.

Link to notebook:
https://colab.research.google.com/drive/1nCCw0QBTf_i13HeYJRly1NeDAdD9OKbD?usp=sharing

Description:
The project we have carried out deals with visual and predictive analysis in relation to the dataset relating to the earthquakes that occurred between 1965 and 2016 focusing on analyzing the intensity of earthquakes, the depth and the place where they occurred to define a history and a trace of where they occurred.
The idea came to us by browsing the most popular datasets on the site [1], we agreed to deal with the analysis of earthquakes on those dates because in our opinion it is interesting to look at the frequency, with what intensity and where earthquakes occur to understand , from the extrapolated and graphed information, where are the most affected areas in order to try to understand where are the faults in the world or in any case which are the areas with high seismic density.
This analysis, in our opinion, can be useful in advancing the defense against earthquakes through, in fact, a deeper knowledge of the sources of these events.

Initially the dataset uploaded from the site [2] had extra columns describing information that was not useful to us and this was going to burden subsequent executions so we reworked it so that it contained only what served our purpose. We have also standardized some columns such as the magnitude which had two decimal digits that we, for simplicity, approximated them to obtain what you see in fig.1.1. We've also changed the date and time column to a familiar format. We then added a column to the dataframe relating to the country of an earthquake identified by the coordinates of latitude and longitude. To do this we used thelibrary reverse_geocoder which converts to iso alpha 2, i.e. each state is identified by two letters. In some operations performed subsequently we will be forced to convert this data also in iso alpha 3, to standardize the data to that of other libraries.
Furthermore, from the “Type” column we learn that the propagation of seismic waves is not caused only by the movement of the tectonic plates but also by other natural or man-made events as we can see from fig.1.2.

While in fig.1.3 we can see in what proportion each event led to a detection of the movement of the earth's mantle.

	Date	Time	Latitude	Longitude	Type	Depth	Magnitude	Magnitude Type	Source	Nation
0	1965-01-02	13:44:18	19.2460	145.6160	Earthquake	131.60	6.0	MW	ISCGEM	MP
1	1965-01-04	11:29:49	1.8630	127.3520	Earthquake	80.00	5.8	MW	ISCGEM	ID
2	1965-01-05	18:05:58	-20.5790	-173.9720	Earthquake	20.00	6.2	MW	ISCGEM	TO
3	1965-01-08	18:49:43	-59.0760	-23.5570	Earthquake	15.00	5.8	MW	ISCGEM	GS
4	1965-01-09	13:32:50	11.9380	126.4270	Earthquake	15.00	5.8	MW	ISCGEM	PH
...	...	...	...	...	...	...	...	...	...	...

fig.1.1 First five records

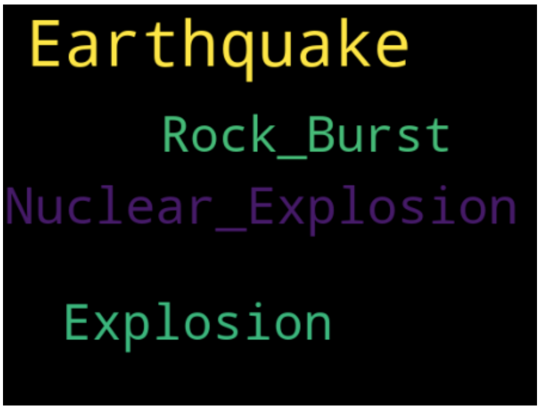


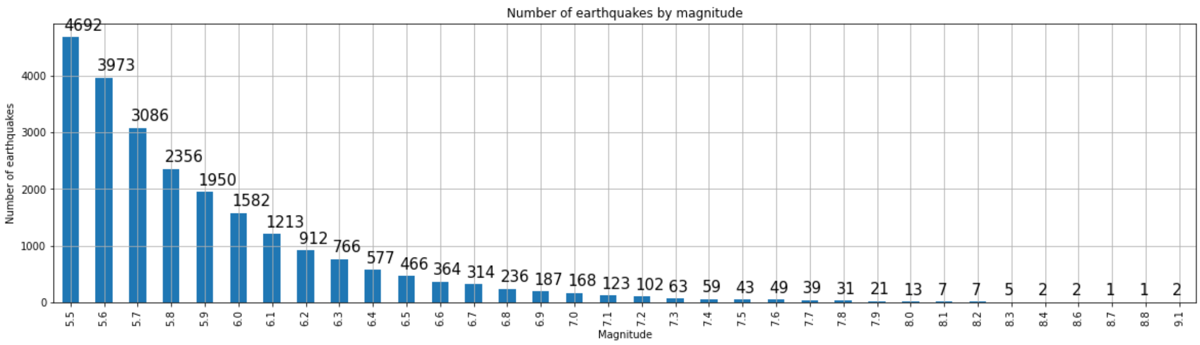
fig.1.2 Events that caused propagation of seismic waves

	Type	Numero eventi
0	Earthquake	23232
1	Explosion	4
2	Nuclear Explosion	175
3	Rock Burst	1

fig.1.3 Numbers of the types of events

Results:

As a first result we are going to extrapolate the information of how many earthquakes occur with a certain intensity. Graph 2.1 informs us that earthquakes with a lower magnitude are much more likely than high ones. We also note that the "curve" decreases exponentially and this gives us a confirmation that the magnitude scale was thought of as a logarithmic scale.



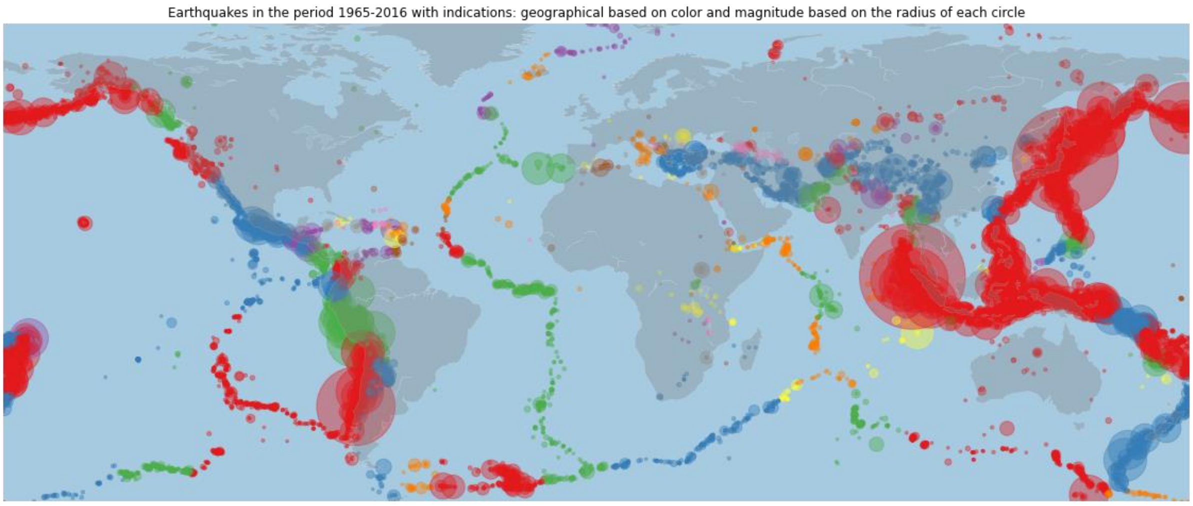
2.1 Number of earthquakes by magnitude

To confirm what has just been said, we can see how events with a certain magnitude are distributed over the time span of a year in fig.2.2. We can see that intensity 5 is far more likely than intensity 9. This estimate was extrapolated from our dataset, which contains earthquakes recorded between 1965 and 2016. If we had a longer range we would be able to determine this statistic with greater precision. Despite this we are very close to the statistics reported by wikipedia [3].

```
Magnitude 5: approximately 314.84313725490193 every year
Magnitude 6: approximately 129.7450980392157 every year
Magnitude 7: approximately 13.686274509803921 every year
Magnitude 8: approximately 1 every 1.3421052631578947 years
Magnitude 9: approximately 1 every 25.5 years
```

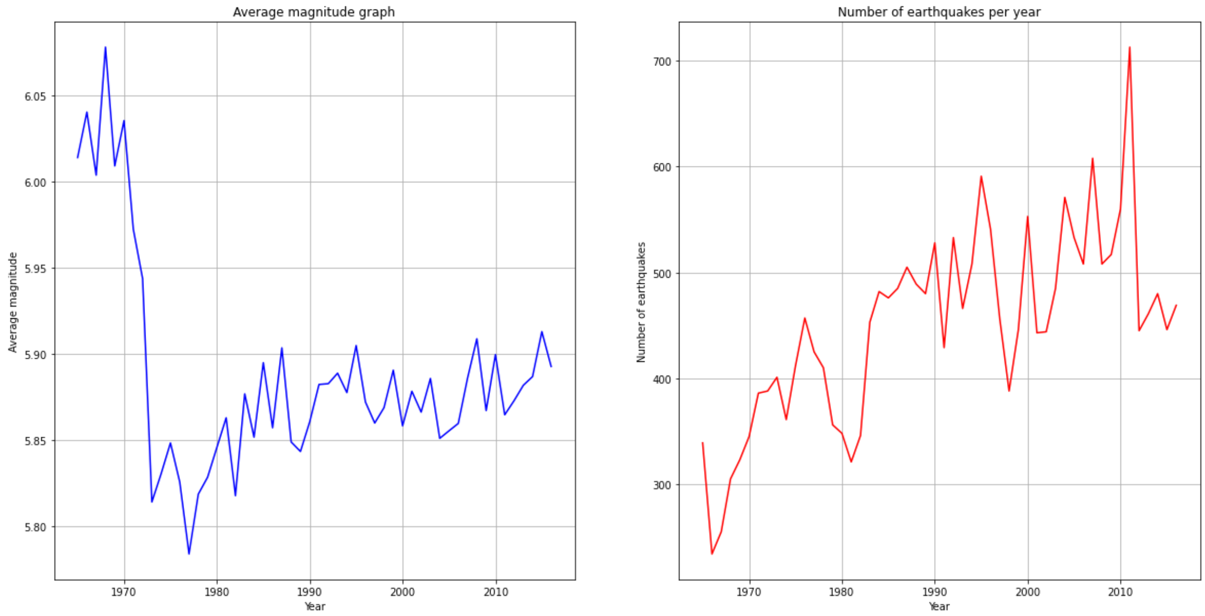
fig.2.2 Annual frequency by magnitude

Secondly we went to analyze where earthquakes occur and with what intensity, as seen in fig.2.3, this allows us to visualize where the most affected areas are and where earthquakes occur with a higher magnitude giving us information about faults and areas with high seismic density. In the notebook there is a similar graph, but with the peculiarity of being passable in the time axis.



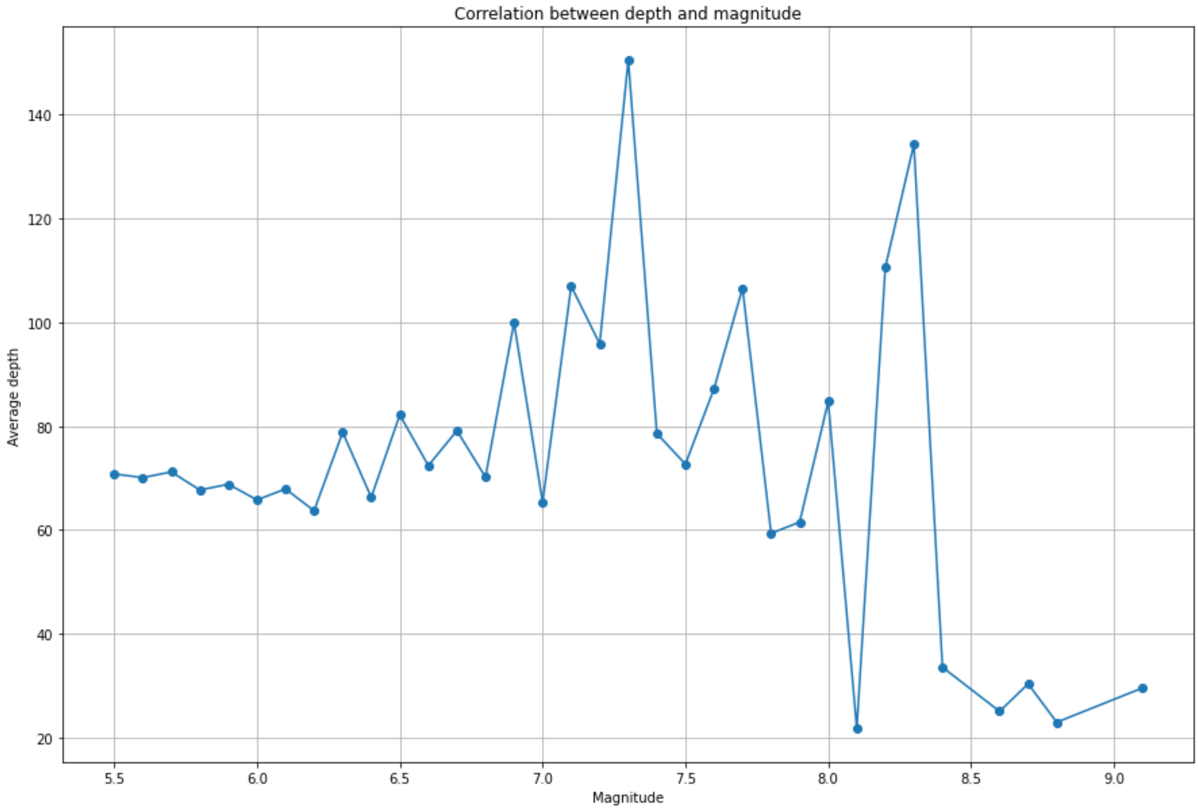
2.3 Earthquakes in the world from 1965 to 2016

We also went to extract information regarding the average annual magnitude and the number of earthquakes per year that inform us of the situation by giving a history and making known the years with the greatest number and that with the least number of earthquakes, as can be seen in fig.2.4. After that we can say that there is no correlation between the number of earthquakes and the annual average. Another thing that we can notice is a confirmation of the observations made in the first result in fig.2.1 that is that when we have recorded more earthquakes the average tends to normalize around a value of average magnitude, or about 5.87.



2.4 Average annual magnitude (left); Annual number of earthquakes (right)

As the penultimate result reported, we wanted to point out that there is no correlation between the depth at which an earthquake occurs and extends and its intensity in terms of magnitude. In fact, from fig.2.5 it is clear that there is no relationship between the power of an earthquake and its depth.



2.5 Correlation between depth and magnitude (which is not there)

In conclusion, we went to make a predictive analysis to try to understand, given a position in terms of latitude and longitude, with what magnitude and with what depth an earthquake could occur. The model, inspired by [4], works with an accuracy of about 88% managing to have a satisfactory precision about the data that we can analyze, which are those used to train the model. To create the model we used the library [5] and the result is shown in fig.2.6.

```
Elements we train on: 23177
Elements we test on: 235
Accuracy 88.2962918580128 %
```

2.6. Result from training the model

Comments:
In the results above we showed how we managed to realize our idea by centering what we wanted to point out in terms of information, i.e. where most earthquakes occur and how intense trying to graphically deduce where the earthquakes are located. faults in our planet and any temporal correlations or with other data, such as depth.
Furthermore, in the notebook there are other studies that we preferred not to include in the report so as not to dwell too much and, therefore, we recommend viewing them (in the notebook there are also animated views, impossible to report here).

Breakdown of the work:
All work was done by both of us, every piece of code and every output was written, edited and supervised by both. This report was also drawn up together.
The choice of the dataset was made by mutual agreement.

Bibliography:
[1] Kaggle: Your Machine Learning and Data Science Community
[2] Significant Earthquakes, 1965-2016
[3] Richter Scale
[4] https://www.kaggle.com/mahadevmm9/earthquake-prediction
[5] scikit -learn: machine learning in Python - scikit-learn 0.24.2 documentation