Earthquake prediction using machine learning - PYTHON

PHASE 2 - INNOVATION

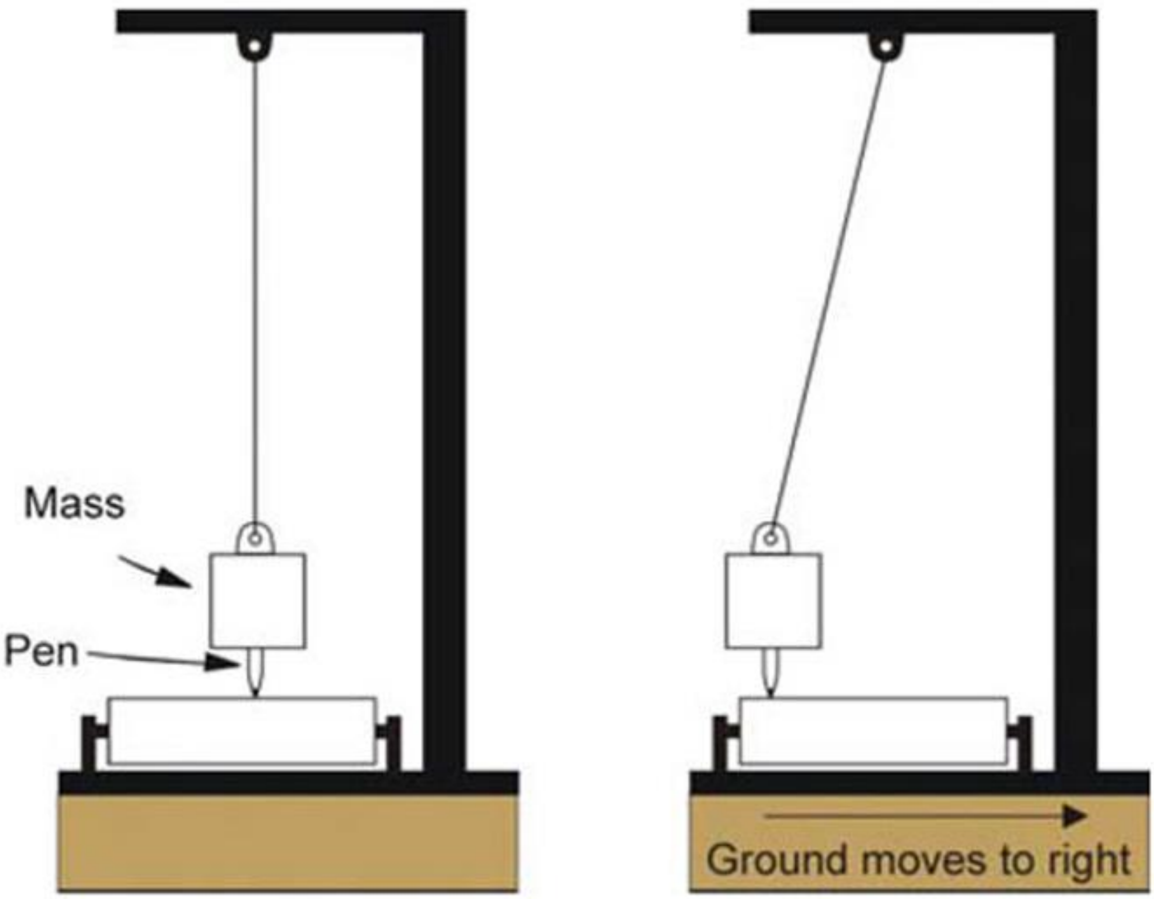
ABSTRACT :

Earthquake prediction has been a challenging research area, where a future occurrence of the devastating catastrophe is predicted. In this work, sixty seismic features are computed through employing seismological concepts, such as Gutenberg-Richter law, seismic rate changes, foreshock frequency, seismic energy release, total recurrence time. Further, Maximum Relevance and Minimum Redundancy (mRMR) criteria is applied to extract the relevant features. A Support Vector Regressor (SVR) and Hybrid Neural Network (HNN) based classification system is built to obtain the earthquake predictions. HNN is a step wise combination of three different Neural Networks, supported by Enhanced Particle Swarm Optimization (EPSO), to offer weight optimization at each layer. The newly computed seismic features in combination with SVR-HNN prediction system is applied on Hindukush, Chile and Southern California regions. The obtained numerical results show improved prediction performance for all the g regions, compared to previous prediction studies.

The following are the innovation tools used to predict the earthquake,

SEISMOMETER:

A seismogram is a record of the ground motions caused by seismic waves from an earthquake. A seismograph or seismometer is the measuring instrument that creates the seismogram. Almost all seismometers are based on the principle of inertia, that is, where a suspended mass tends to remain still when the ground moves.



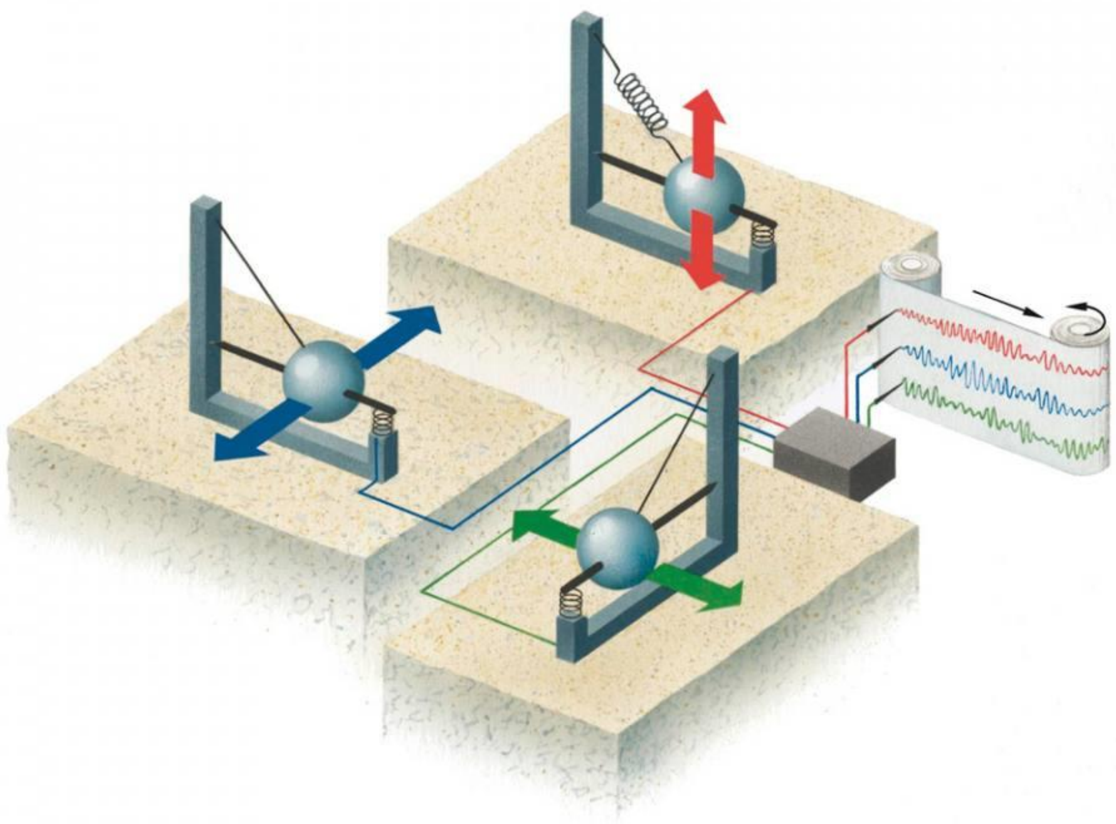
Seismometers allow us to detect and measure earthquakes by converting vibrations due to seismic waves into electrical signals, which we can then display as seismograms on a computer screen. Seismologists study earthquakes and can use this data to determine where and how big a particular earthquake is.

To record the actual motion of the ground in all three dimensions, seismologists need to use three separate sensors within the same instrument. Each sensor records the vibrations in a different direction:

The Z component measures up/down motion

The E component measures east/west motion

The N component measures north-south motion



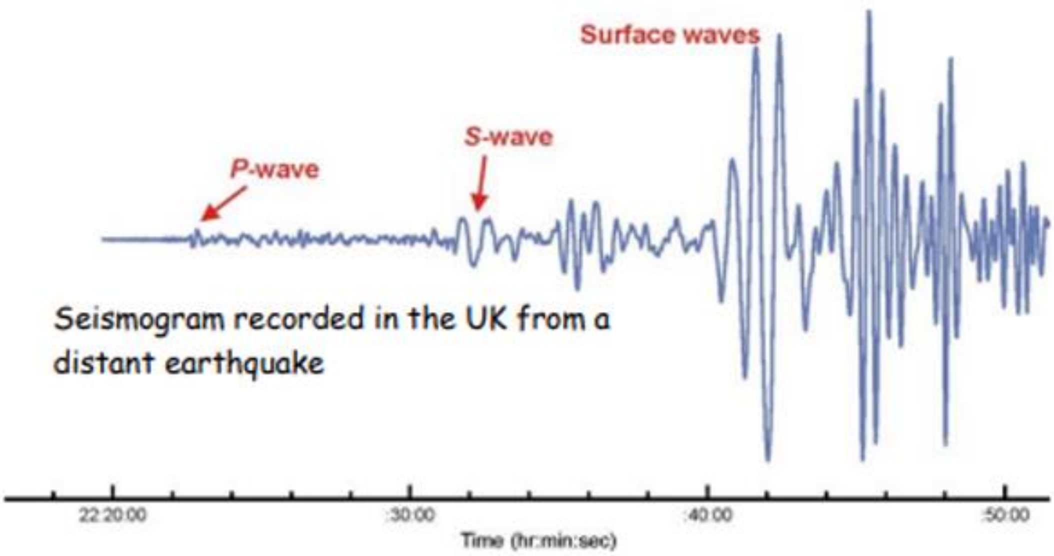
PARAMETERS:

Seismic waves

There are two basic types of seismic wave that travel through the body of the Earth: P-waves and S-waves. P-waves are longitudinal waves that consist of a series of compressions and dilations along the direction of travel. The P stands for primary because they travel the fastest. S-waves are transverse waves, whose motion is perpendicular to the direction of travel. The S stands for shear or secondary since they are slower than P-waves.

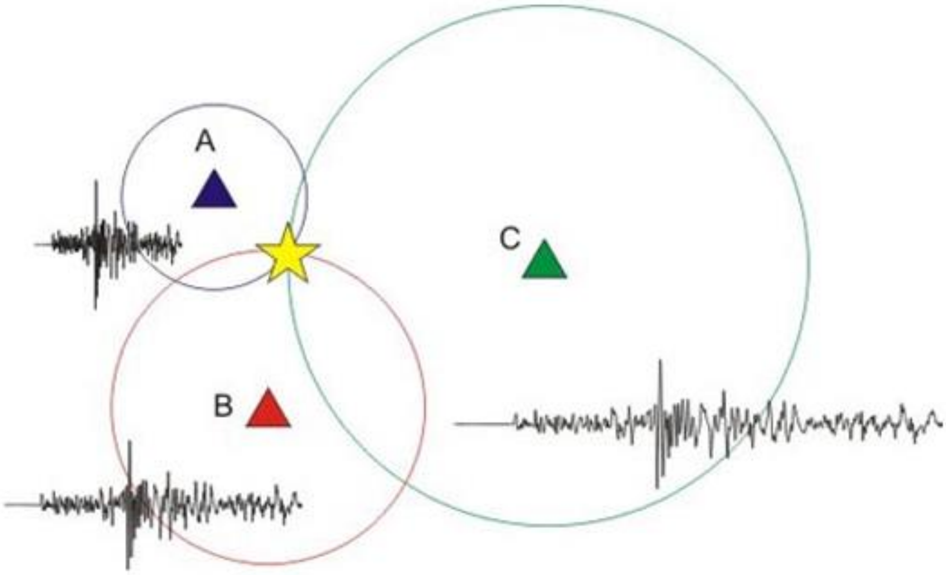
Where a free surface is present (like the Earth/air interface) these two types of motion can combine to form surface waves, which produce a type of shaking that causes buildings to fail and fall down. There are two types of surface waves: Rayleigh waves and Love waves. Rayleigh waves are generated by the interaction of P- and S-waves at the surface of the Earth, while Love waves are generated by interference of multiple shear waves. The ground motions from surface waves are often much larger than those motions from body waves.

Earthquakes generate different types of seismic waves and these travel at different speeds through the Earth. P-waves are fastest and are the first signal to arrive on a seismogram, followed by the slower S-wave, then the surface waves. The arrival times of the P- and S-waves at different seismometers are used to determine the location of the earthquake. Assuming that we know the relative speed of P- and S-waves, the time difference between the arrivals of the P- and S-waves determines the distance the earthquake is from the seismometer.



By looking at the seismograms from different recording stations, we can find out the epicentre of the earthquake. The signals arrive first at the closest station and last at the one furthest away. The time difference between the P- and S-waves tells us the distance the earthquake is from the seismometer. If we calculate the S minus P time to determine distance from the seismometer at three stations, we can work out where the epicentre of the earthquake is

Imagine A, B and C are three different seismometer stations at

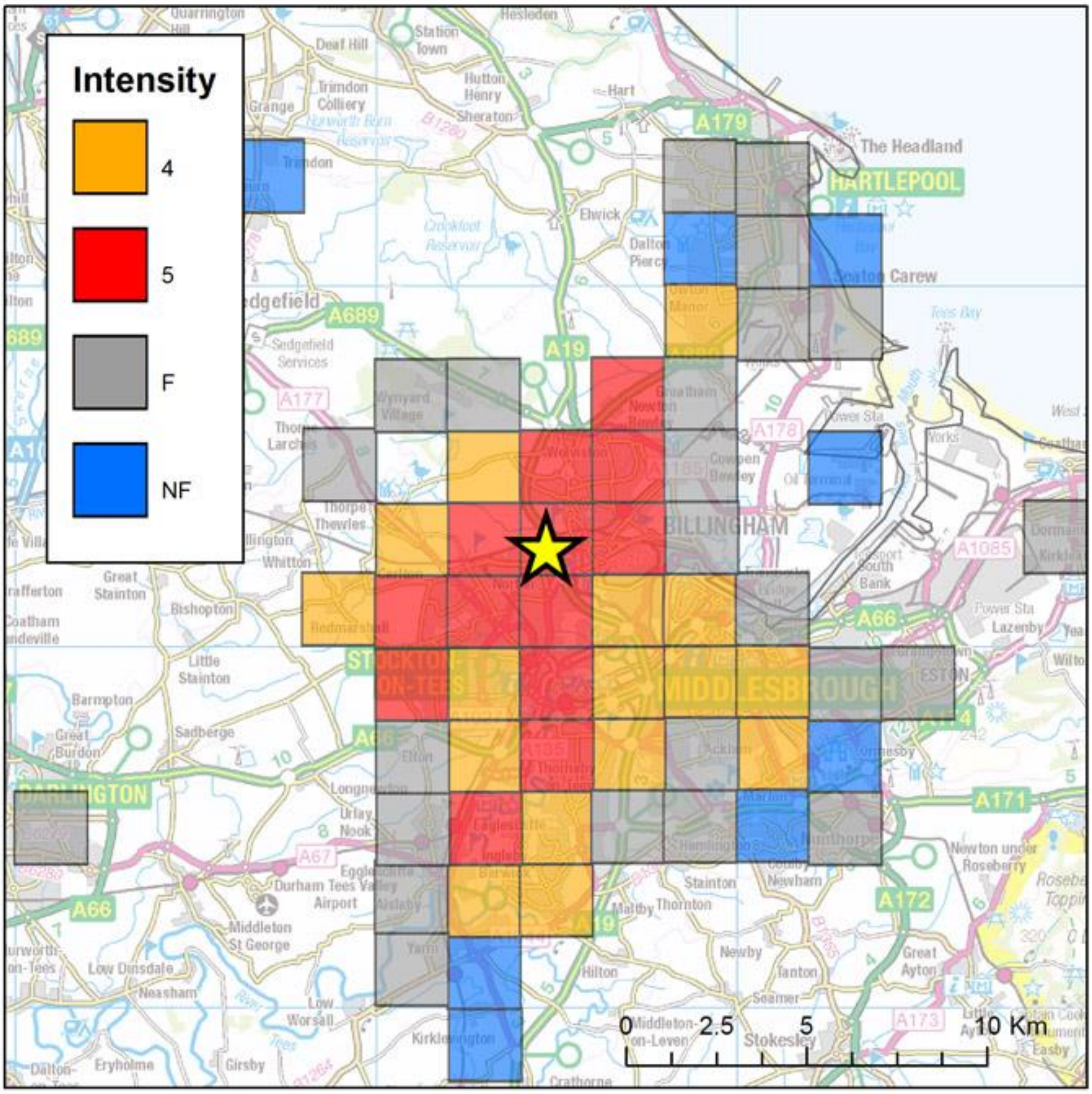


distant locations. Once we know the distance to an earthquake from three seismic stations, we can determine the location of the earthquake. Draw a circle around each station with a radius equal to its distance from the earthquake. The earthquake occurred at the point where all three circles intersect. BGS

Earthquake intensity

Intensity is a qualitative measure of the strength of shaking caused by an earthquake determined from the observed effects on people, objects and buildings. For a given earthquake, the intensity normally decreases with distance from the epicentre. There are a number of different intensity scales in use around the world that are all based on the shaking people experience and the effects it has on objects and buildings. It is also possible to estimate intensity from recordings of ground motions.

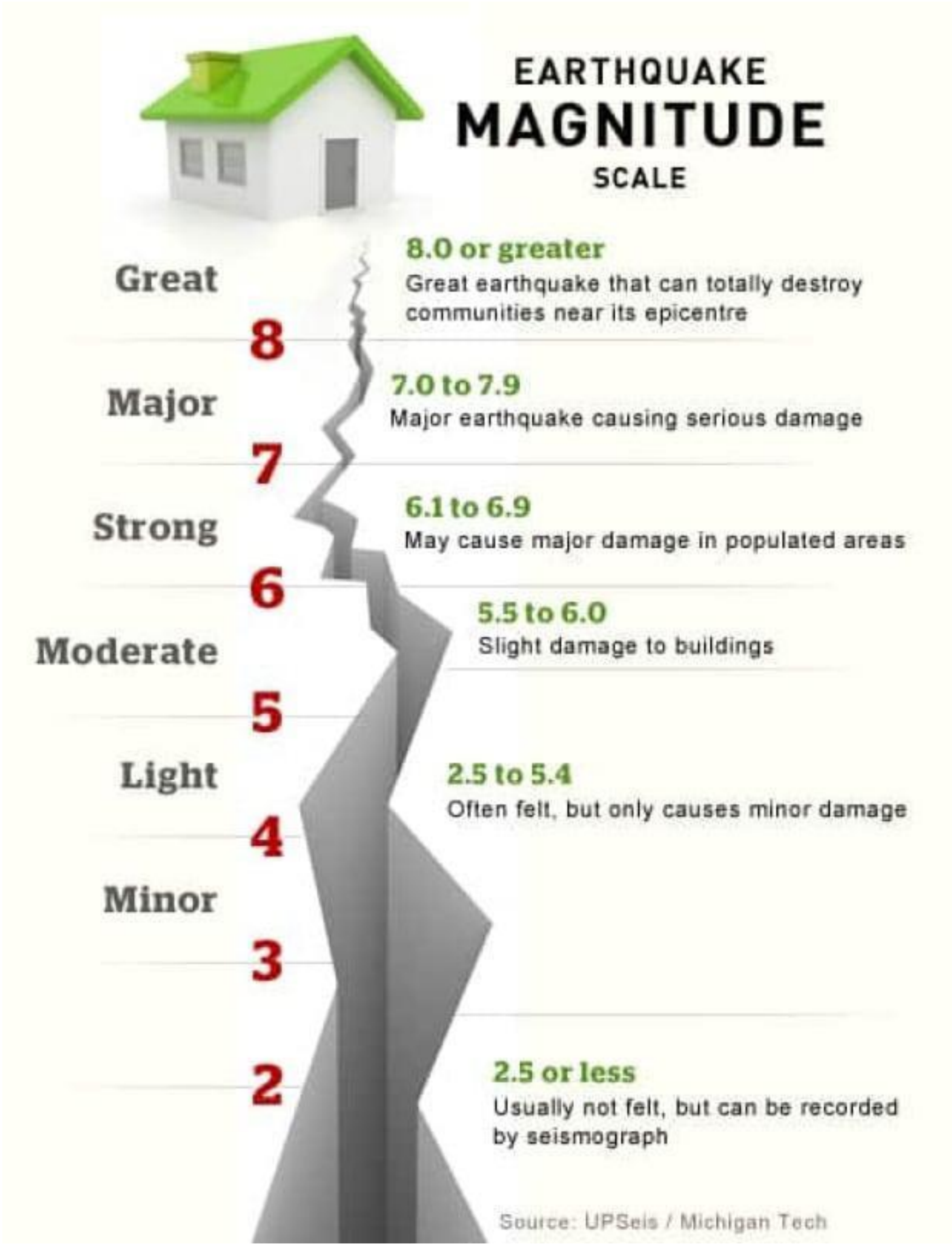
Macroseismic intensities (EMS) for the magnitude 3.1 ML earthquake on 23 January 2020, near Stockton-on-Tees, UK. The yellow star shows the earthquake epicentre. Intensities are calculated in 2 km grid squares from over 840 reports from people who felt the earthquake. A minimum of five observations are needed in any grid square to calculate a value of intensity, otherwise the value is recorded as ‘Felt’, but no intensity is calculated (shown by grey squares). Blue squares indicate that reports from these locations suggest that the earthquake was not felt.



Earthquake magnitude

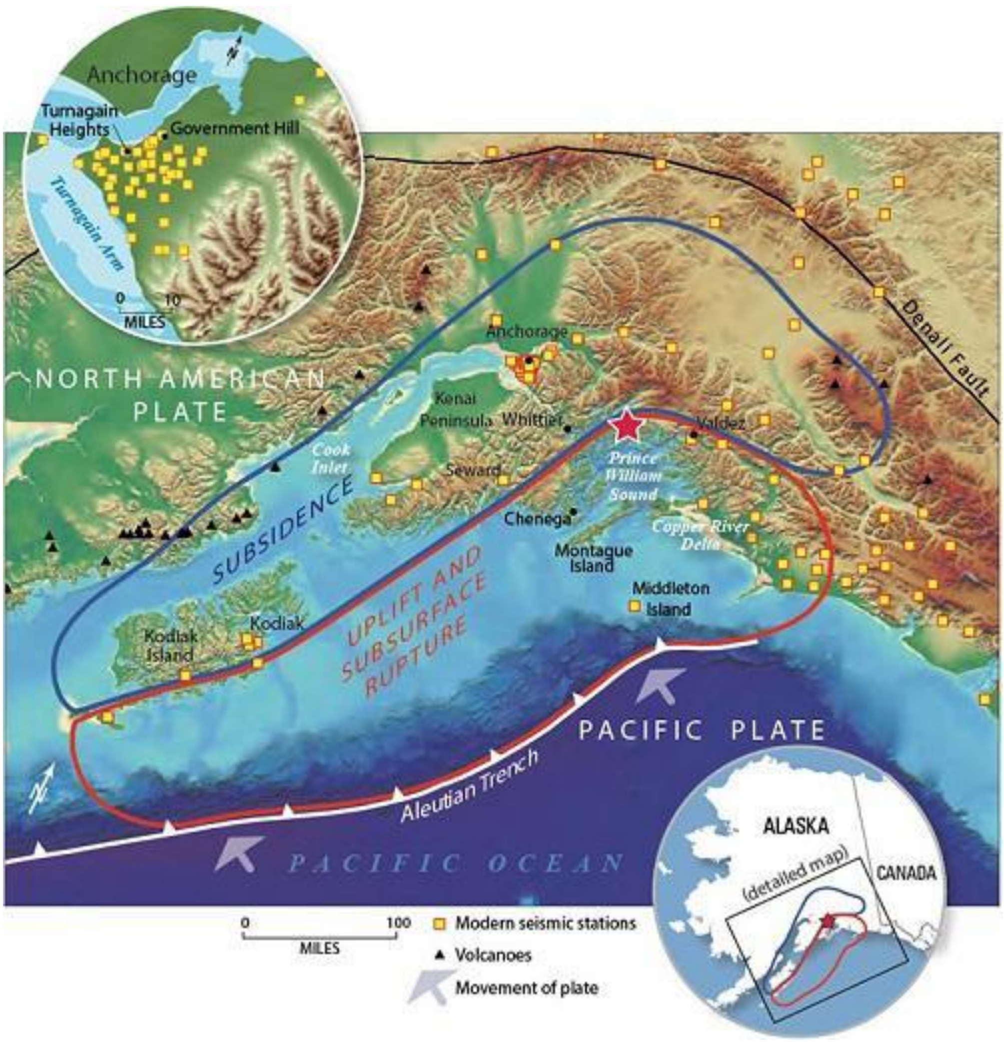
Magnitude is a measure of the amount of energy released during an earthquake and can be estimated from the amplitude of ground motions recorded by seismometers. It is independent of distance from the epicentre.

A number of different magnitude scales have been developed based on the amplitude of different parts of the observed record of ground motion with specific corrections for distance. Earthquake magnitude scales are logarithmic, i.e. a one unit increase in magnitude corresponds to a tenfold increase in amplitude.



Moment magnitude

Nowadays, the most standard and reliable measure of earthquake size is moment magnitude (Mw), which is based on seismic 'moment'. Moment is related to the area of the earthquake fault rupture and the amount of slip on the rupture, as well as the strength of the rocks themselves. Richter's original magnitude scale underestimates the size of large events, so the constants used in the definition of Mw were chosen so that the magnitude numbers for Richter and moment magnitudes match for smaller events.



FEATURE ENGINEERING:

Feature Engineering helps to derive some valuable features from the existing ones. These extra features sometimes help in increasing the performance of the model significantly and certainly help to gain deeper insights into the data.

```
Splitted = df['Origin Time'].str.split(' ', n=1,  Expand=True)
 Df['Date'] = splitted[0]
Df['Time'] = splitted[1].str[:-4]
 Df.drop('Origin Time',Axis=1   Inplace=True)
Df.head()
```

Output:

	Latitude	Longitude	Depth	Magnitude	Location	Date	Time
0	29.06	77.42	5.0	2.5	53km NNE of New Delhi, India	2021-07-31	09:43:23
1	19.93	72.92	5.0	2.4	91km W of Nashik, Maharashtra, India	2021-07-30	23:04:57
2	31.50	74.37	33.0	3.4	49km WSW of Amritsar, Punjab, India	2021-07-30	21:31:10
3	28.34	76.23	5.0	3.1	50km SW of Jhajjar, Haryana	2021-07-30	13:56:31
4	27.09	89.97	10.0	2.1	53km SE of Thimphu, Bhutan	2021-07-30	07:19:38

Now, let's divide the date column into the day, month, and year columns respectively.

```
Df['day'] = splitted[2].astype('int')
Df['month'] = splitted[1].astype('int')
Df['year'] = splitted[0].astype('int')
 Df.drop('Date', axis=1,  Inplace=True)
Df.head()
```

Output:

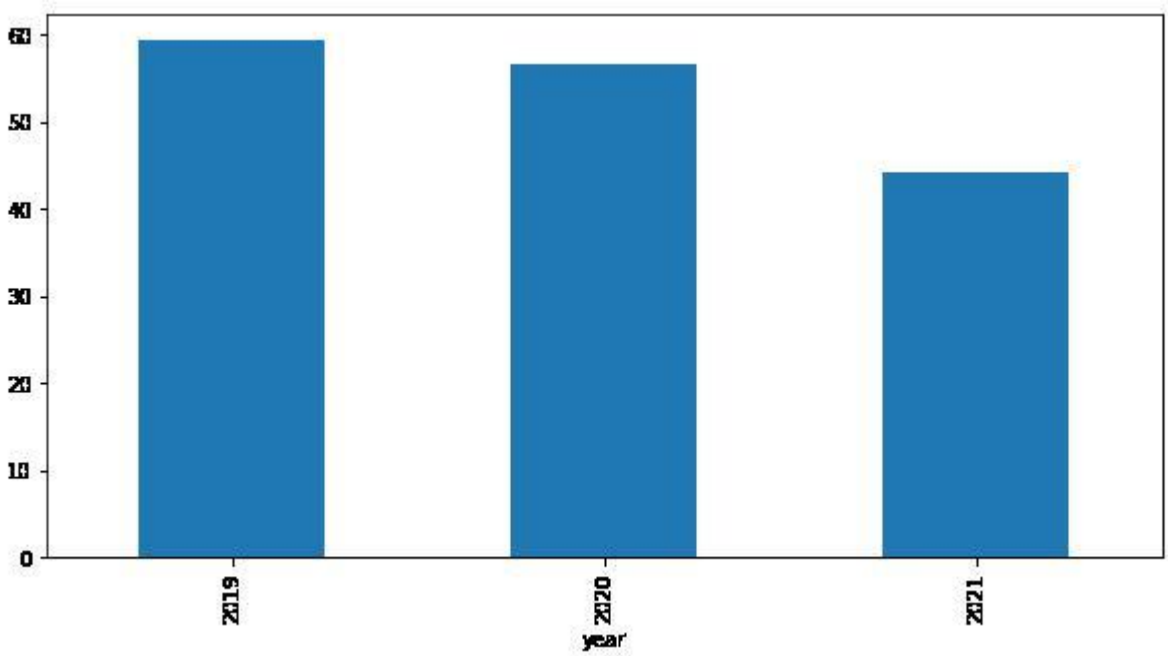
	Latitude	Longitude	Depth	Magnitude	Location	Time	day	month	year
0	29.06	77.42	5.0	2.5	53km NNE of New Delhi, India	09:43:23	31	7	2021
1	19.93	72.92	5.0	2.4	91km W of Nashik, Maharashtra, India	23:04:57	30	7	2021
2	31.50	74.37	33.0	3.4	49km WSW of Amritsar, Punjab, India	21:31:10	30	7	2021
3	28.34	76.23	5.0	3.1	50km SW of Jhajjar, Haryana	13:56:31	30	7	2021
4	27.09	89.97	10.0	2.1	53km SE of Thimphu, Bhutan	07:19:38	30	7	2021

Exploratory Data Analysis

EDA is an approach to analyzing the data using visual techniques. It is used to discover trends, and patterns, or to check assumptions with the help of statistical summaries and graphical representations

Plt.f
Plt.

plt.figure(figsize=(10, 5))



x = df.groupby('year').mean()['Depth']

x.plot.bar()

plt.show()

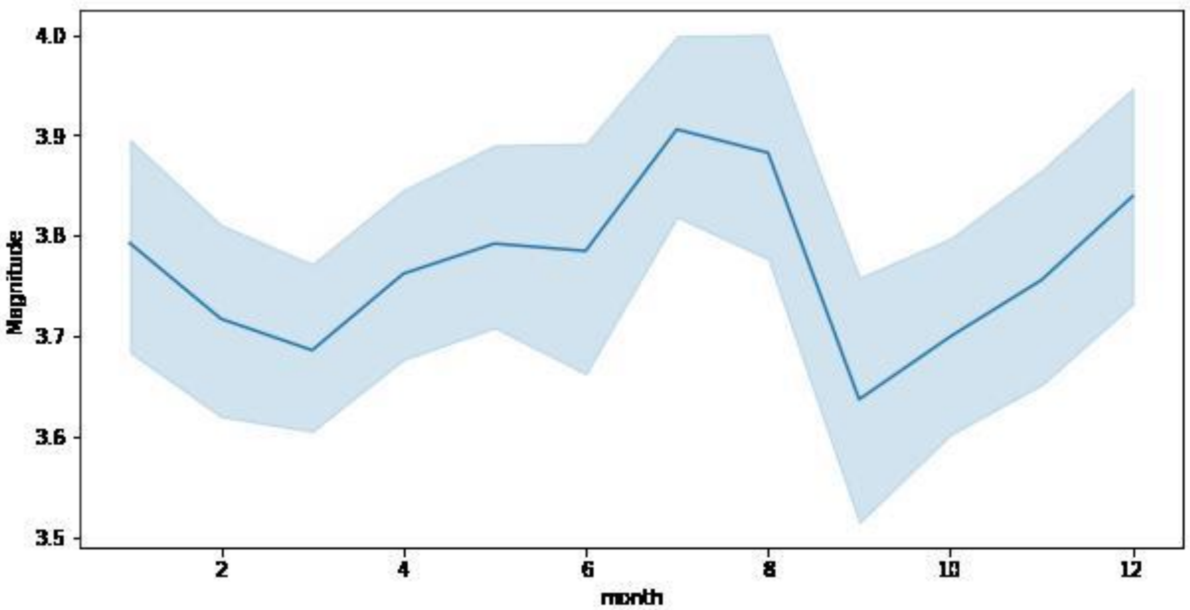
The depth from which earthquakes are starting is reducing with every passing year.

Plt.figure(figsize=(10, 5))

Sb.lineplot(data=df, X='month'   Y='Magnitude')

Plt.show()

Output:



Here we can observe that the changes of an earthquake with higher magnitude are more observed during the season of monsoon.

Plt.subplots(figsize=(15, 5))

Plt.subplot(1, 2, 1)

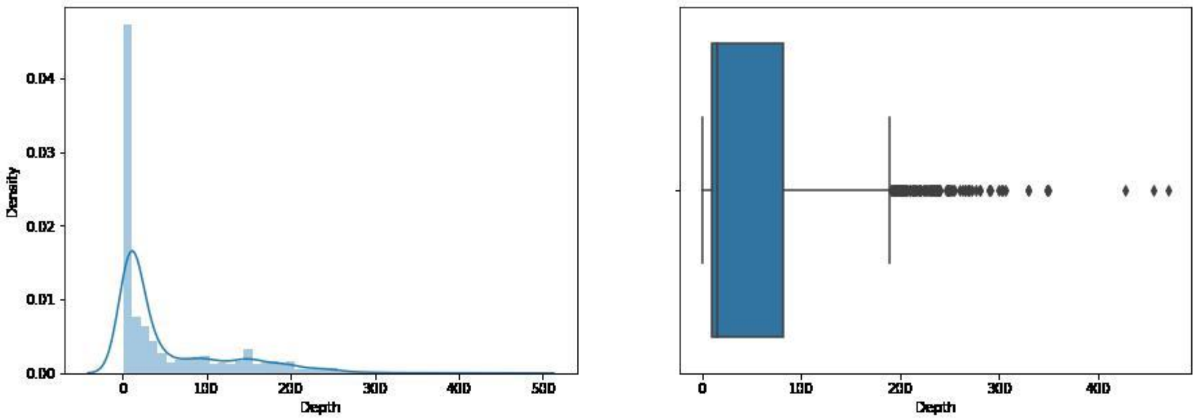
Sb.distplot(df['Depth'])

Plt.subplot(1, 2, 2)

Sb.boxplot(df['Depth'])

Plt.show()

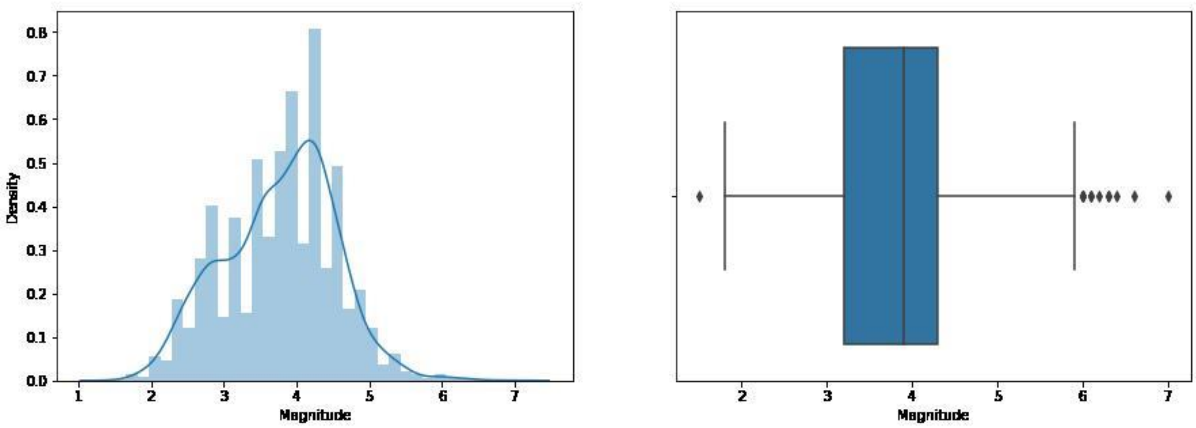
Output:



From the distribution graph, it is visible that there are some outliers that can be confirmed by using the boxplot. But the main point to observe here is that the distribution of the depth at which the earthquake rises is left-skewed.

```
Plt.subplots(figsize=(15, 5))
 Plt.subplot(1, 2, 1)
Sb.distplot(df[‘Magnitude’])
 Plt.subplot(1, 2, 2)
Sb.boxplot(df[‘Magnitude’])
 Plt.show()
```

Output:



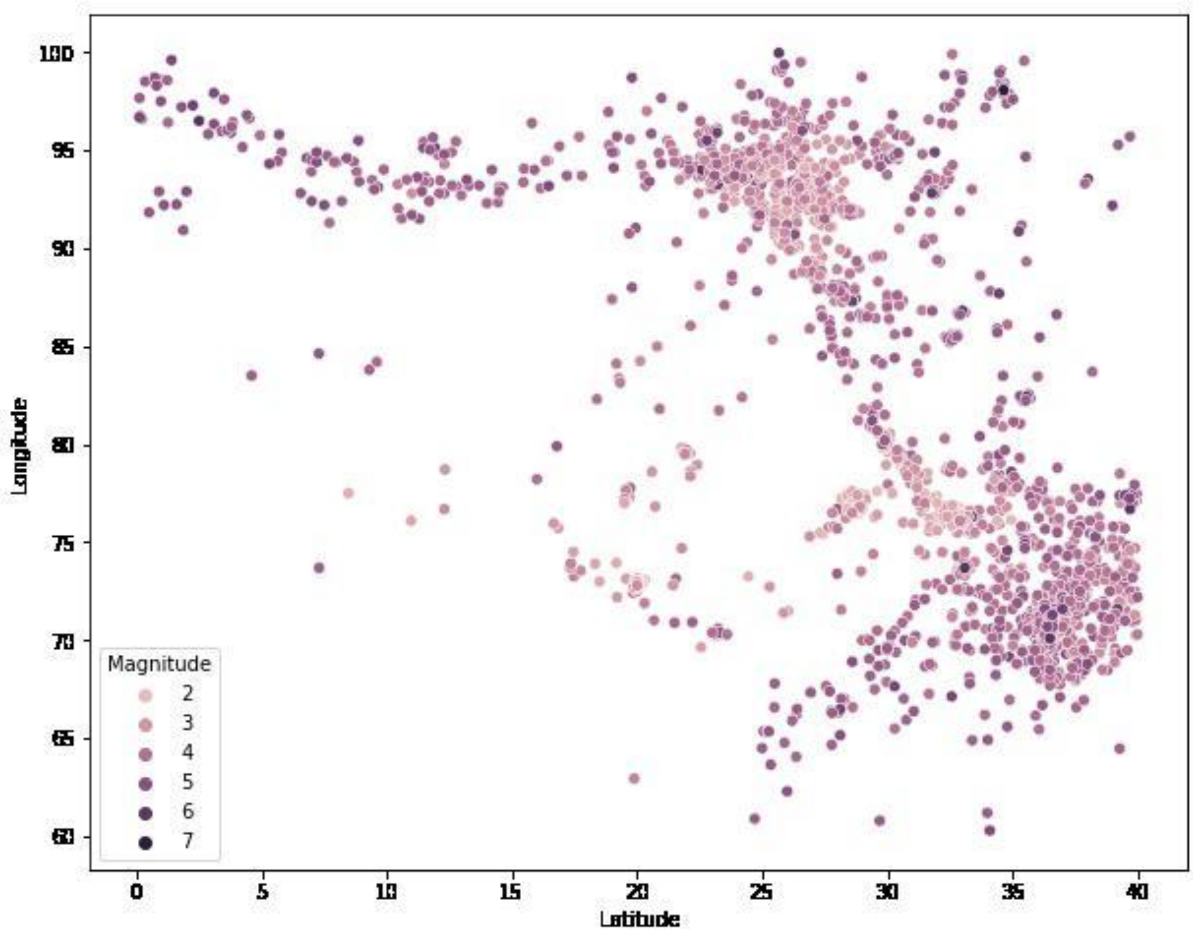
As we know that many natural phenomena follow a normal distribution and here we can observe that the magnitude of the earthquake also follows a normal distribution.

Plt.figure(figsize=(10, 8))

Sb.scatterplot(data=df, X='Latitude'   Y='Longitude', Hue='Magnitude')

Plt.show()

Output:



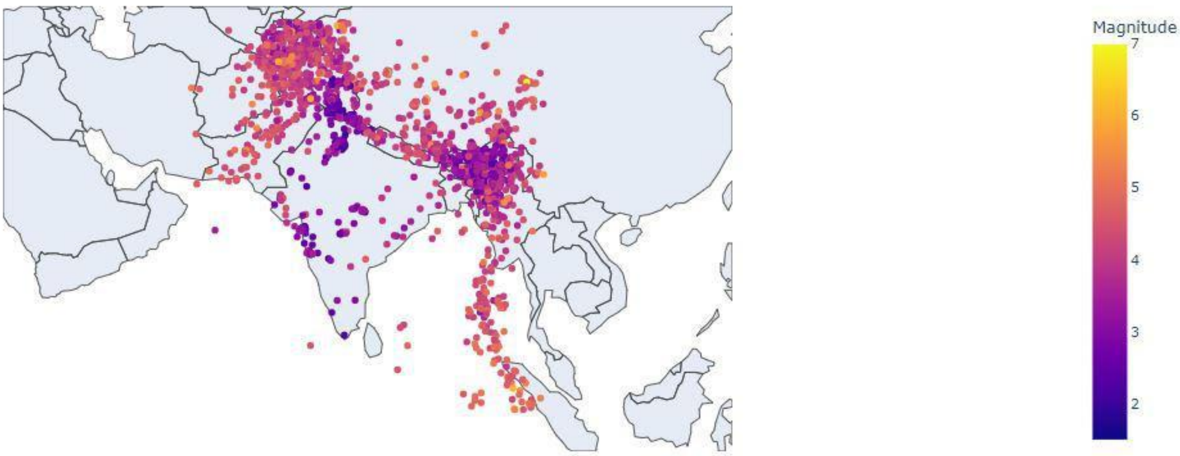
Now by using Plotly let's plot the latitude and the longitude data on the map to visualize which areas are more prone to earthquakes.

Import plotly.express as px

Import pandas as pd

```
Fig = px.scatter_geo(df, lat='Latitude',

                     Lon='Longitude',

                     Color="Magnitude",

                     Fitbounds='locations',

                     Scope='asia')
Fig.show()
```

Output:



CONCLUSION:

Earthquakes occur mainly as a result of plate tectonics, which involves blocks of the Earth moving about the Earth's surface. The blocks of rock move past each other along a fault. Smaller earthquakes, called foreshocks, may precede the main earthquake, and aftershocks may occur after the main earthquake. Earthquakes are mainly confined to specific areas of the Earth known as seismic zones, which coincide mainly with ocean trenches, mid-ocean ridges, and mountain ranges.