Earthquake is a natural phenomenon whose occurrence predictability is still a hot topic in academia. This is because of the destructive power it holds. In this project, we’ll learn how to analyze and visualize earthquake data with Python and Matplotlib.

**Importing Libraries and Dataset**

Python libraries make it very easy for us to handle the data and perform typical and complex tasks with a single line of code .The libraries we are going to use in this earthquake prediction model are pandas and matplotlib.

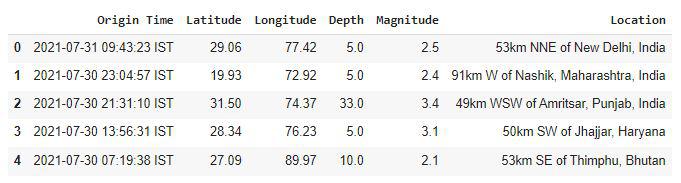
* [**Pandas**](https://www.geeksforgeeks.org/python-pandas-dataframe/)– This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.
* [**Matplotlib**](https://www.geeksforgeeks.org/matplotlib-tutorial/)/**[Seaborn](https://www.geeksforgeeks.org/introduction-to-seaborn-python/)**– This library is used to draw visualizations.

|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sb    import warnings  warnings.filterwarnings('ignore') |

Now, let’s load the dataset into the pandas data frame for easy analysis.

|  |
| --- |
| df = pd.read\_csv('dataset.csv')  df.head() |

**Output:**



The dataset we are using here contains data for the following columns:

* Origin time of the Earthquake
* Latitude and the longitude of the location.
* Depth – This means how much depth below the earth’s level the earthquake started.
* The magnitude of the earthquake
* Location

Now we shall print the structure of the dataset , that is its shape.

|  |
| --- |
| df.shape |

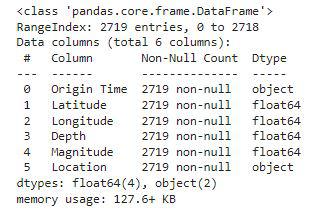
**Output:**

(2719, 6)

Now let’s see which data is present in which type of data format.

|  |
| --- |
| df.info() |

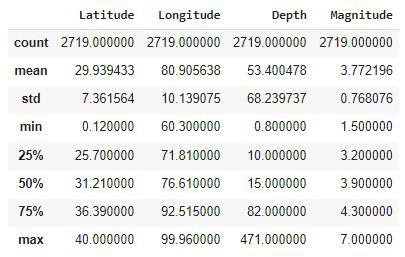
**Output:**



Looking at the descriptive statistical measures also gives us some idea regarding the distribution of the data.

|  |
| --- |
| df.describe() |

**Output:**



From the above description of the dataset, we can conclude that:

* The maximum magnitude of the Earthquake is 7.
* The maximum depth at which the earthquake started is 471 km below the ground.

**Feature Engineering**

Feature Engineeringhelps 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:**



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

|  |
| --- |
| splitted = df['Date'].str.split('-', expand=True)    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:**

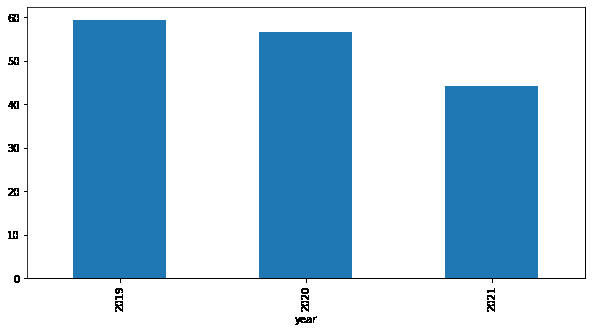


**Exploratory Data Analysis**

[EDA](https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/) 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.figure(figsize=(10, 5))  x = df.groupby('year').mean()['Depth']  x.plot.bar()  plt.show() |

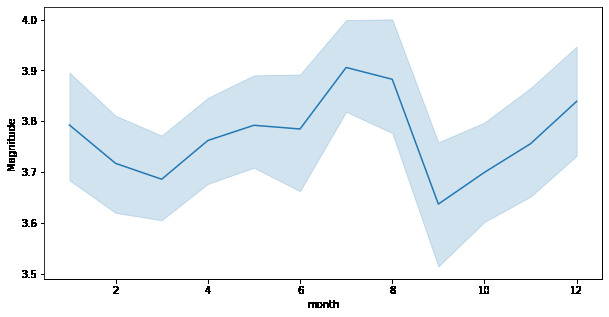
**Output:**



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() |

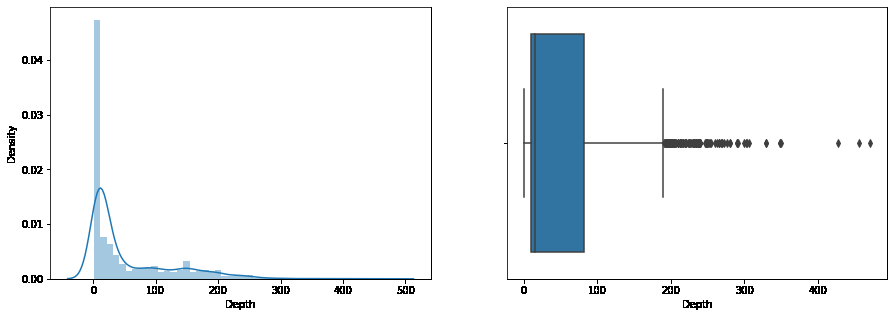
**Output:**



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() |

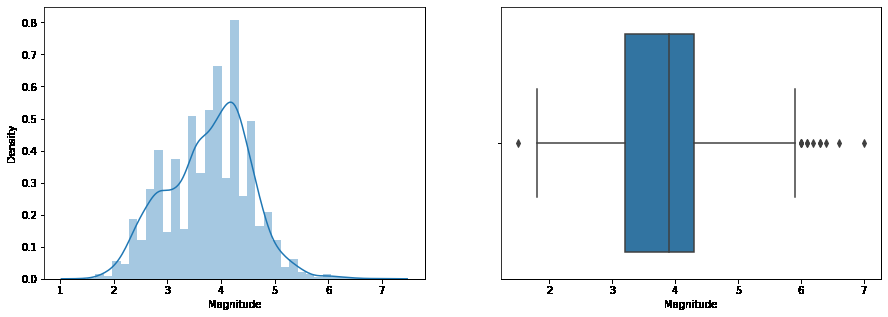
**Output:**



From the distribution graph, it is visible that there are some [outliers](https://www.geeksforgeeks.org/machine-learning-outlier/)that can be confirmed by using the [boxplot](https://www.geeksforgeeks.org/boxplot-using-seaborn-in-python/). 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() |

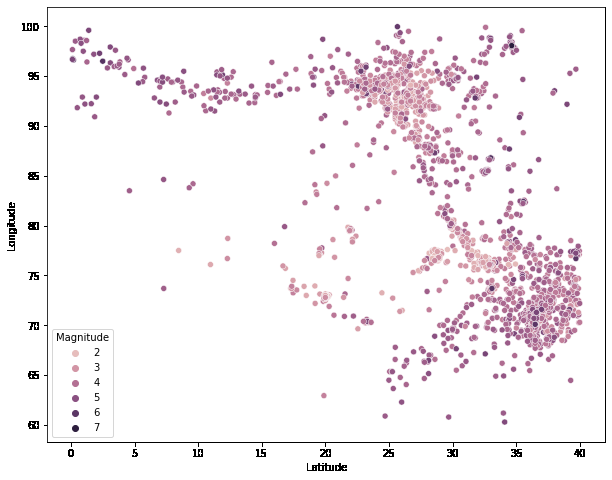
**Output:**



As we know that many natural phenomena follow a [normal distribution](https://www.geeksforgeeks.org/python-normal-distribution-in-statistics/)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() |

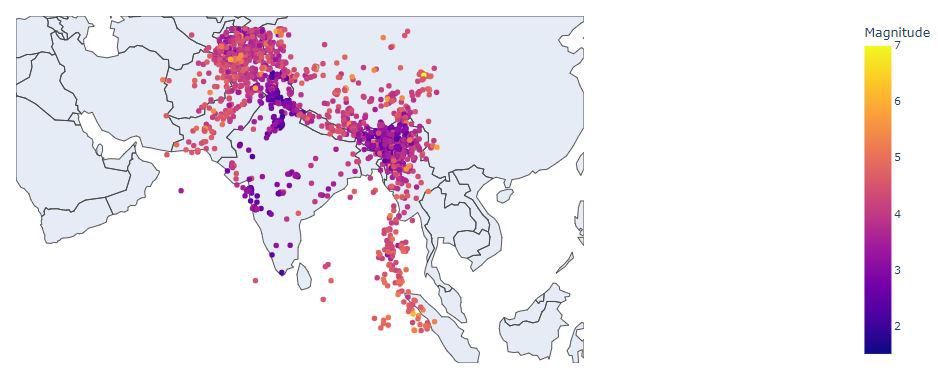
**Output:**



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:**

In this project we are using the python’s matplotlib library and pandas library and how to use it. We have also analysed and visualised the earthquake dataset using the matplotlib library. And also we predicted the zones that are prone to earthquake by using python.