Earthquake prediction using Python

Phase -3

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.

- Pandas – 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/Seaborn – This library is used to draw visualizations.

Code:

All the libraries and the file required for the analysis is imported in the code.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import warnings

warnings.filterwarnings('ignore')
df = pd.read_csv('dataset.csv')
```

df.head()

Dataset Used:

dataset.csv

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

Analysis of the dataset:

Describing the CSV file

Code:

df.describe()

Output:

	Latitude	Longitude	Depth	Magnitude
count	2719.000000	2719.000000	2719.000000	2719.000000
mean	29.939433	80.905638	53.400478	3.772196
std	7.361564	10.139075	68.239737	0.768076
min	0.120000	60.300000	0.800000	1.500000
25%	25.700000	71.810000	10.000000	3.200000
50%	31.210000	76.610000	15.000000	3.900000
75%	36.390000	92.515000	82.000000	4.300000
max	40.000000	99.960000	471.000000	7.000000

Feature Engineering:

Code:

```
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

Info Gathering:

Code:

```
df.shape
df.info()
```

Output:

(2719, 6)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2719 entries, 0 to 2718
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Origin Time   2719 non-null   object 
 1   Latitude      2719 non-null   float64
 2   Longitude     2719 non-null   float64
 3   Depth         2719 non-null   float64
 4   Magnitude     2719 non-null   float64
 5   Location      2719 non-null   object 
dtypes: float64(4), object(2)
memory usage: 127.6+ KB
```

Data Visualization:

Code:

```
plt.figure(figsize=(10, 5))
x = df.groupby('year').mean()['Depth']
x.plot.bar()
plt.show()
```

Output: