**Earthquake Prediction model using python**

It is well known that if a disaster has happened in a region, it is likely to happen there again. Some regions really have frequent earthquakes, but this is just a comparative quantity compared to other regions. So, predicting the earthquake with Date and Time, Latitude and Longitude from previous data is not a trend which follows like other things, it is natural occuring.

**Example**

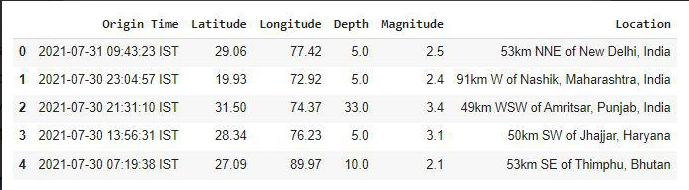
|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sb  import warnings  warnings.filterwarnings('ignore') |
|  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Now, let’s load the dataset into the panda’s 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

**example**

df.shape

**output**

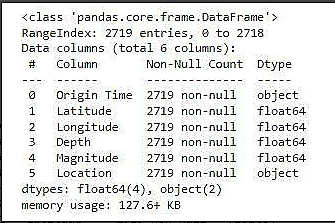
(2719, 6)

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

**Example**

df.info()

**Output**

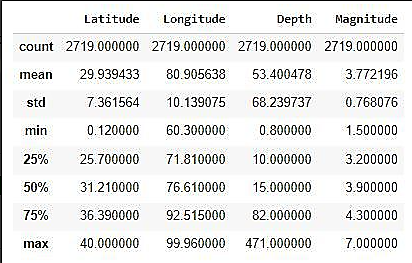


Looking at the descriptive statistical measures also gives us some idea regarding the distribution of the data.

**Example**

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.

**Example**

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.

**Example**

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 representation

**Example**

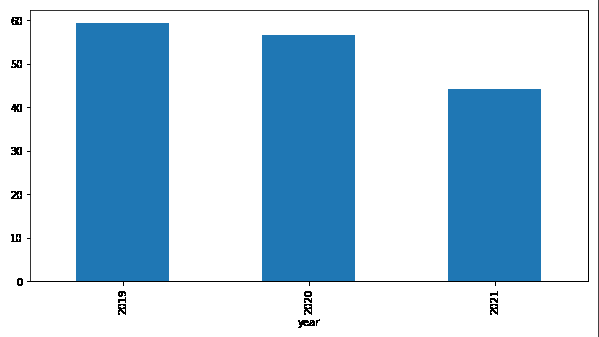
plt.figure(figsize=(10, 5))

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.

**Example**

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

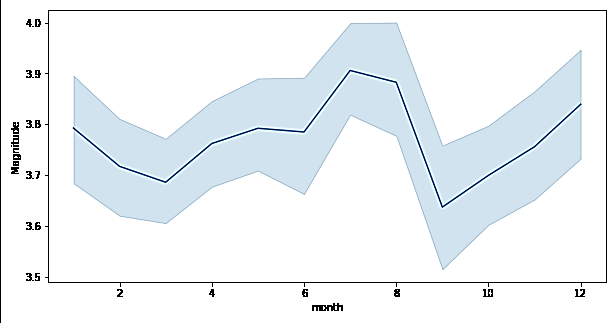
sb.lineplot(data=df,

            x='month',

            y='Magnitude')

plt.show()

**output**



**Example**

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

plt.subplot(1, 2, 1)

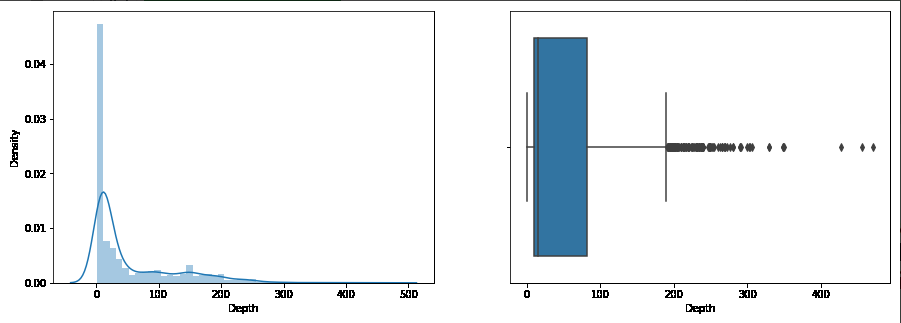
sb.distplot(df['Depth'])

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.

**Report:**

All the above instruction are installed and execution successfully

**Project:**

**Name : S.Priya Dharshini**

**Dept : III year CSE**

**Reg no : 621421104039**

**College code: 6214**

**Group : IBM-GROUP 5**