IBM NaanMudhalvan
ARTIFICIAL INTELLIGENCE

Project Title : Earthquake Prediction Using Python
Phase 4 : Development Part – 2

- Visualizing the data on the world map
- Splitting the dataset into Training and Testing sets.

Workbook Link : GOOGLE COLAB LINK

INTRODUCTION

The visualization process involves leveraging geospatial libraries like Basemap to represent earthquake occurrences worldwide, offering insights into distribution patterns and potential seismic hotspots. This spatial understanding is pivotal for informed decision-making in earthquake-prone regions. Additionally, the strategic split of the dataset into training and testing sets is essential for training machine learning models.

DATA VISUALIZATION

Data visualization plays a crucial role in unraveling the intricate tapestry of earthquake data, offering a lens through which patterns and insights emerge. Leveraging libraries such as Matplotlib, Seaborn, and Basemap, the seismic landscape can be visually represented, providing a comprehensive view of global seismic activity.

DATA SPLITTING

In the journey of constructing a reliable earthquake prediction model,one indispensable phase is the strategic splitting of the dataset into training and testing sets. This division is fundamental for evaluating the model's generalization performance, providing a robust assessment of its predictive capabilities on unseen data. Through libraries like scikit-learn, the dataset is partitioned, with a portion reserved for training the model and the rest set aside for testing its predictive accuracy.

PROGRAM :

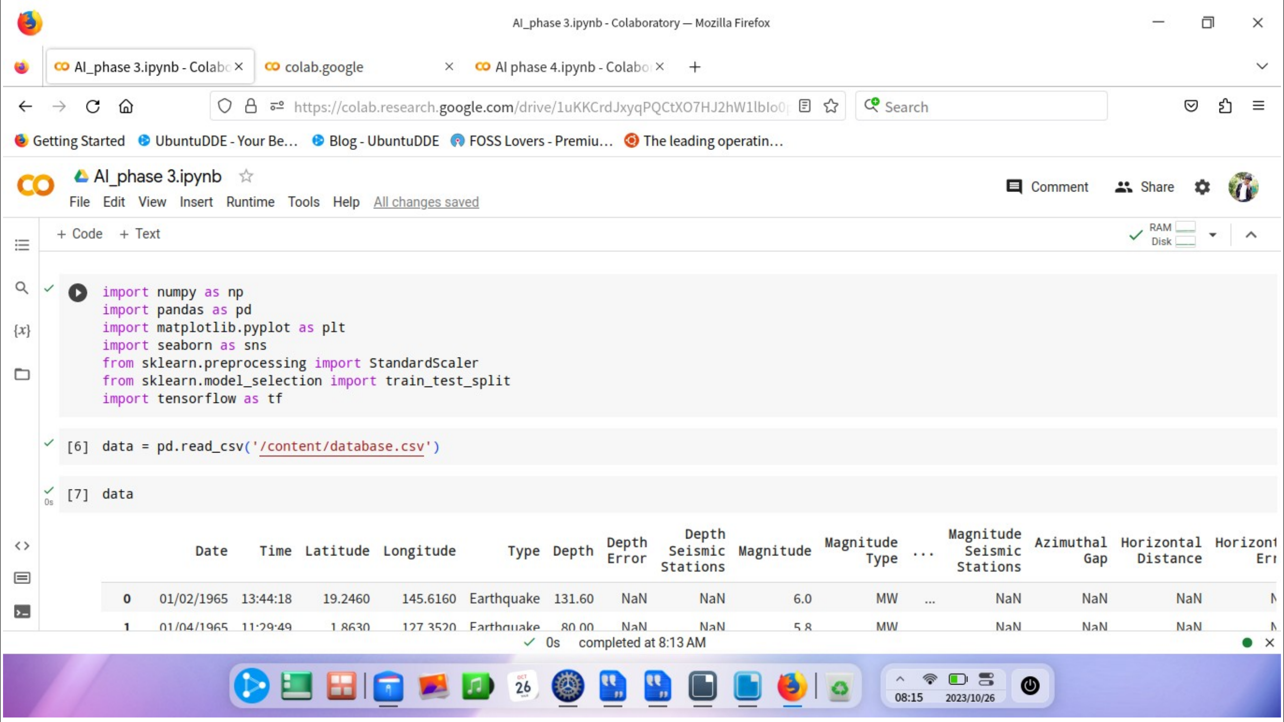
```
!pip3 install basemap

import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap
import seaborn as sns
sns.set(style="darkgrid")

print("Min Value: "+ str(data['Magnitude'].min()))
print("Max Value: " + str(data['Magnitude'].max()))

Greater_8 = data[data['Magnitude'] > 8]
Greater_8['Location Source'].value_counts()

Greater_7 = data[data['Magnitude'] > 7]
Greater_7['Location Source'].value_counts()
Greater_6 = data[data['Magnitude'] > 6]
Greater_6['Location Source'].value_counts()
```

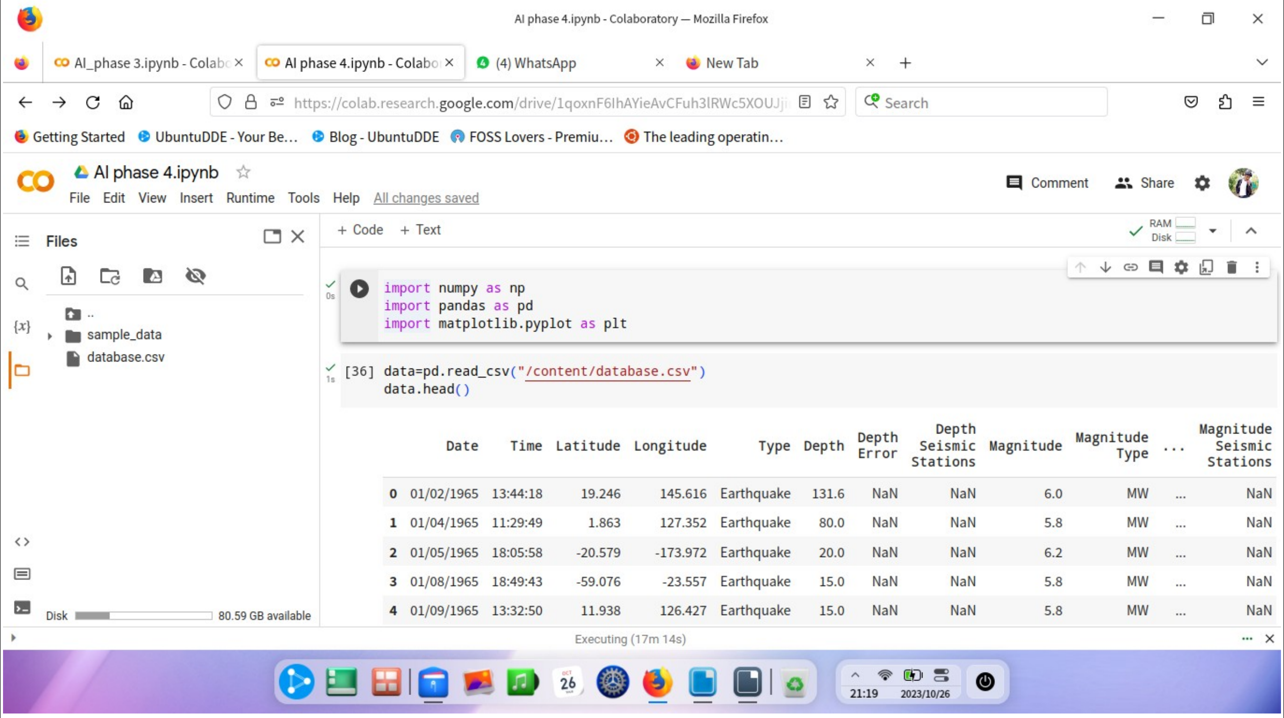
```
Greater_5 = data[data['Magnitude'] > 5]
Greater_5['Location Source'].value_counts()
Greater_4 = data[data['Magnitude'] > 4]
Greater_4['Location Source'].value_counts()

plt.hist(data['Magnitude'])
plt.xlabel('Magnitude Size')
plt.ylabel('Number of Occurrences')

sns.countplot(x="Magnitude Type", data=data)
plt.ylabel('Frequency')
plt.title('Magnitude Type VS Frequency')
print(" local magnitude (ML), surface-wave magnitude
(Ms), body-wave magnitude (Mb), moment magnitude
(Mm)")

def get_marker_color(magnitude):
if magnitude < 6.2:
return ('go')
elif magnitude < 7.5:
return ('yo')
else:
return ('ro')

plt.figure(figsize=(14,10))
eq_map = Basemap(projection='robin', resolution = 'l',
lat_0=0, lon_0=-130)
eq_map.drawcoastlines()
eq_map.drawcountries()
eq_map.fillcontinents(color='gray')
eq_map.drawmapboundary()
eq_map.drawmeridians(np.arange(0, 360, 30))
lons = data['Longitude'].values
lats = data['Latitude'].values
magnitudes = data['Magnitude'].values
timestrings = data['Date'].tolist()
min_marker_size = 0.5
for lon, lat, mag in zip(lons, lats, magnitudes):
x,y = eq_map(lon, lat)
msize = mag
marker_string = get_marker_color(mag)
eq_map.plot(x, y, marker_string, markersize=msize)
title_string = "Earthquakes of Magnitude 5.5 or Greater\n"
title_string += "%s - %s" % (timestrings[0][:10],
timestrings[-1][:10])
plt.title(title_string)
plt.show()

data['date'] = data['Date'].apply(lambda x:
pd.to_datetime(x))
data['year'] = data['date'].apply(lambda x: str(x).split('-')[0])
plt.figure(figsize=(15, 8))
```

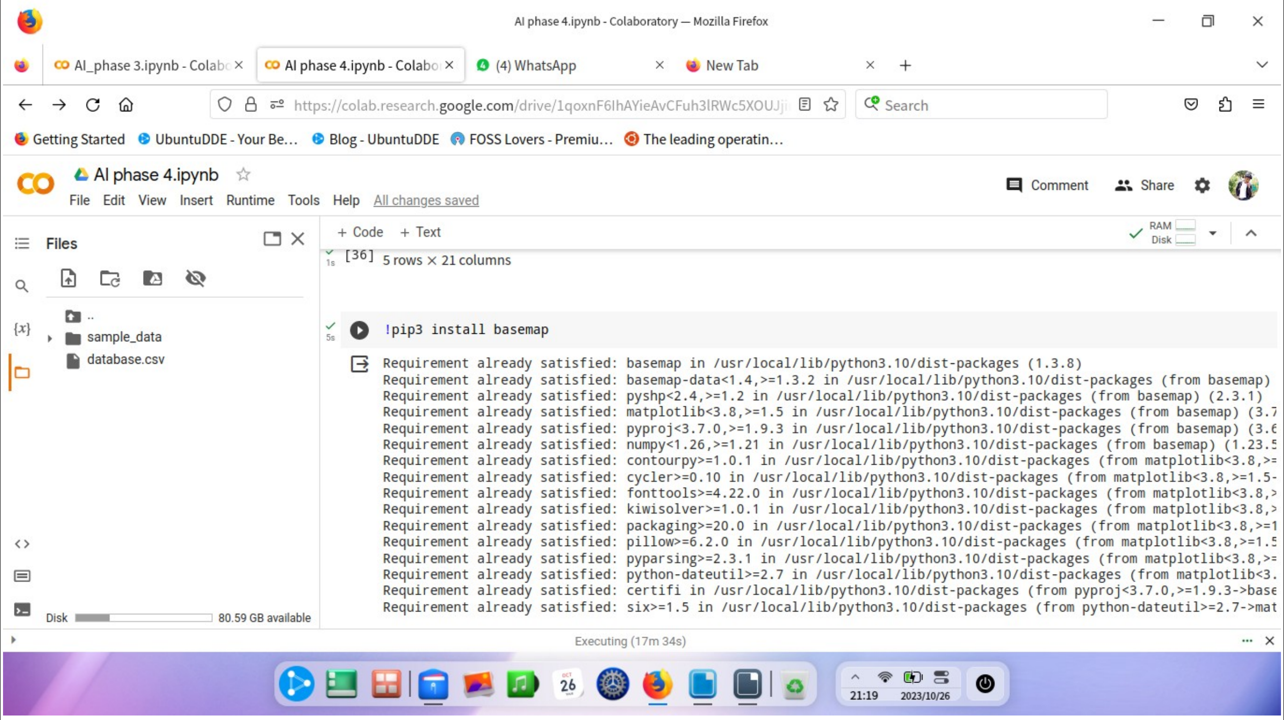
```
sns.set(font_scale=1.0)
ax = sns.countplot(x="year", data=data, color="blue")
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)
plt.ylabel('Number Of Earthquakes')
plt.title('Number of Earthquakes In Each Year')

data['year'].value_counts()[:5]

import datetime
data['date'] = data['Date'].apply(lambda x:
pd.to_datetime(x))
data['mon'] = data['date'].apply(lambda x: str(x).split('-')[1])
plt.figure(figsize=(10, 6))
sns.set(font_scale=1)
ax = sns.countplot(x="mon", data=data, color="green")
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)
plt.ylabel('Number Of Earthquakes')
plt.title('Number of Earthquakes In Each month')

data['mon'].value_counts()[:5]

import datetime
data['date'] = data['Date'].apply(lambda x:
pd.to_datetime(x))
data['days'] = data['date'].apply(lambda x: str(x).split('-')[-
1])
plt.figure(figsize=(16, 8))
sns.set(font_scale=1.0)
ax = sns.countplot(x="days", data=data, color="orange")
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)
plt.ylabel('Number Of Earthquakes')
plt.title('Number of Earthquakes In Each days')
data['days'].value_counts()[:5]
x = data['year'].unique()
y = data['year'].value_counts()
count = []
for i in range(len(x)):
key = x[i]
count.append(y[key])
plt.figure(figsize=(15,12))
plt.scatter(x, count)
plt.title("Earthquake per year from 1995 to 2016")
plt.xlabel("Year")
plt.xticks(rotation=90)
plt.ylabel("Number of Earthquakes")
plt.yticks(rotation=30)
plt.show()

data.loc[data['Magnitude'] >= 8, 'Class'] = 'Disastrous'
data.loc[(data['Magnitude'] >= 7) & (data['Magnitude'] <
7.9), 'Class'] = 'Major'
data.loc[(data['Magnitude'] >= 6) & (data['Magnitude'] <
```

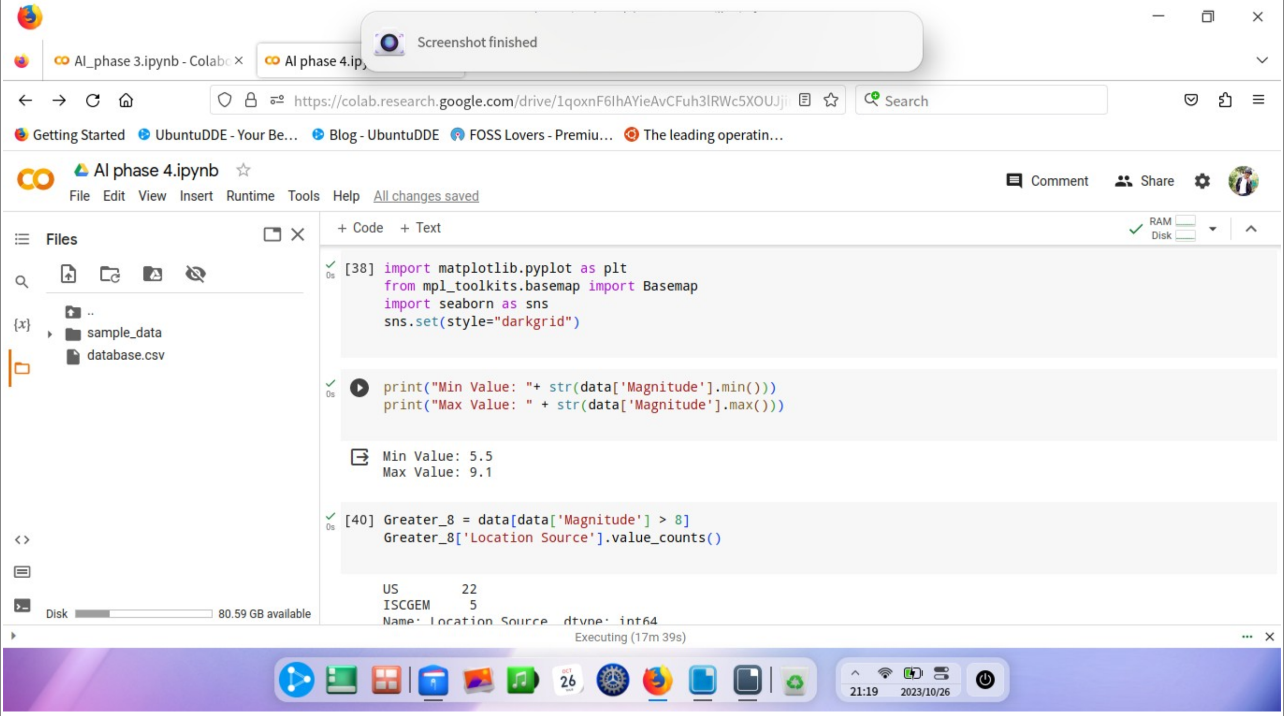
```
6.9), 'Class'] = 'Strong'
data.loc[(data['Magnitude'] >= 5.5) & (data['Magnitude'] <
5.9), 'Class'] = 'Moderate'
sns.countplot(x='Class', data=data)
plt.ylabel('Frequency')
plt.title('Magnitude Class vs Frequency')


X = final_data[['Timestamp', 'Latitude', 'Longitude']]
y = final_data[['Magnitude', 'Depth']]
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
print(X_train.shape, X_test.shape, y_train.shape,
X_test.shape)
```

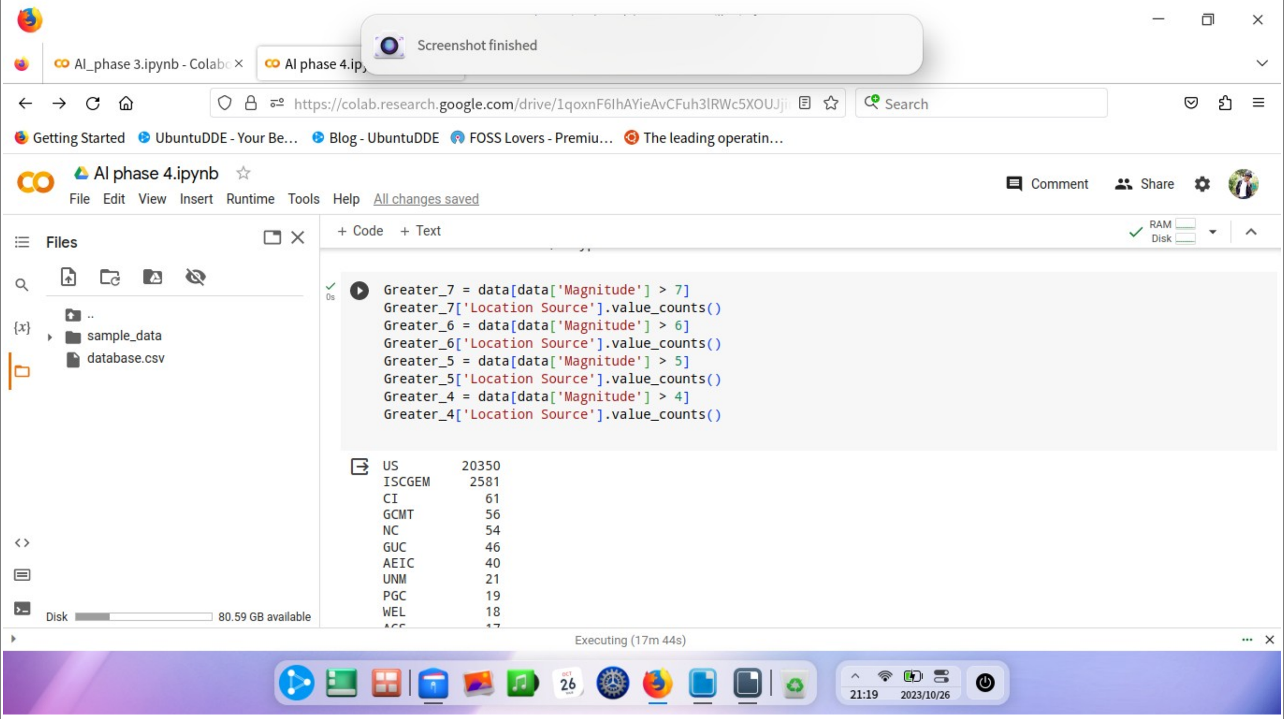
OUTPUT:

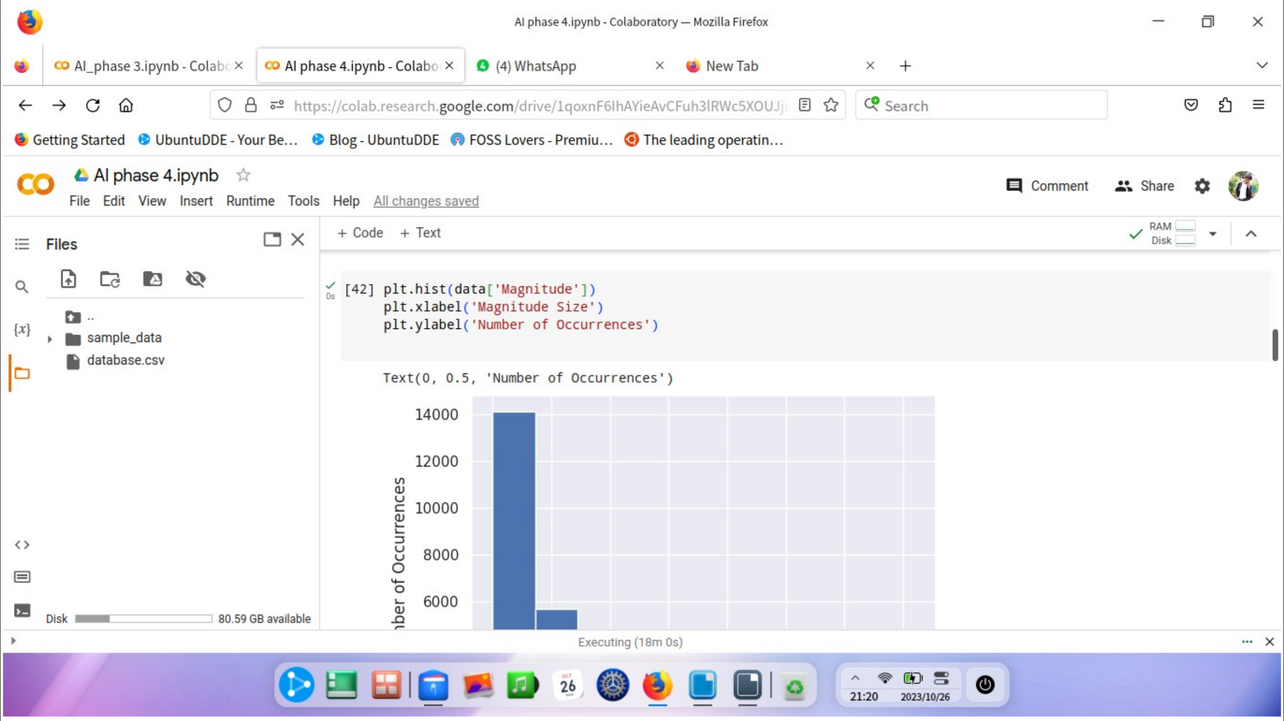


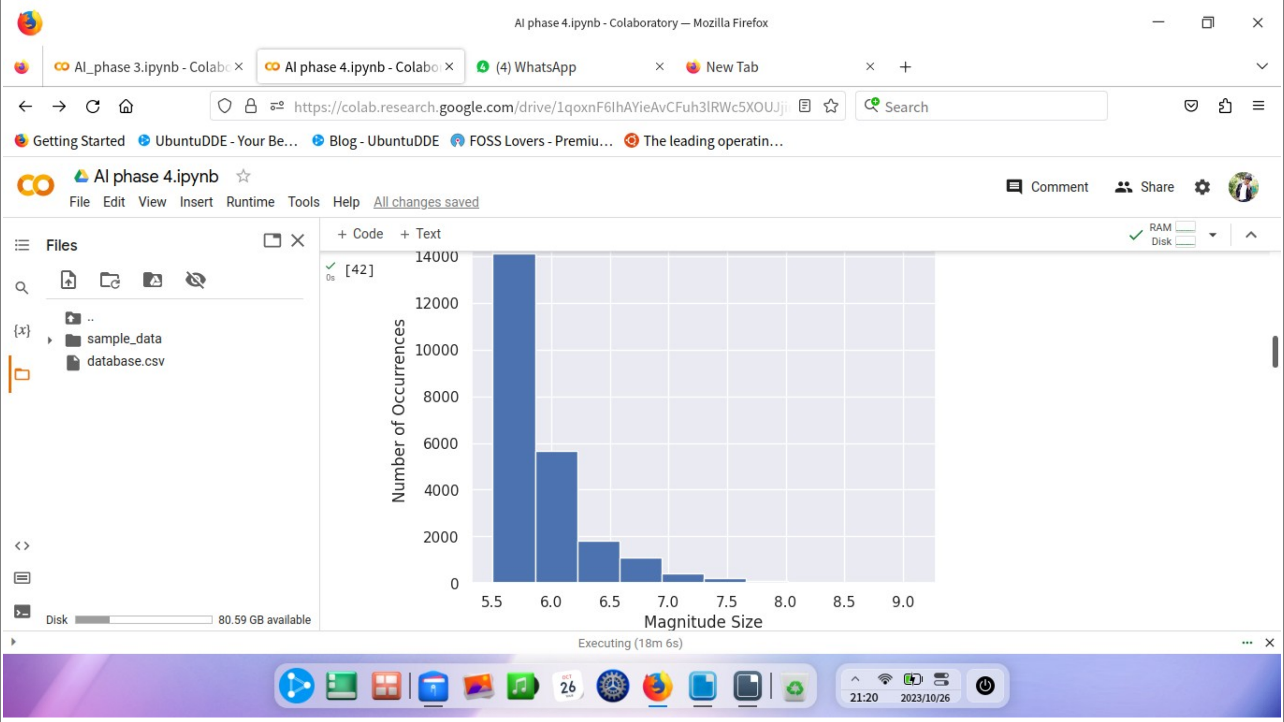

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI_phase 3.ipynb - Colab
AI phase 4.ipynb - Colab
(4) WhatsApp
New Tab
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Search
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
database.csv
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
[36] data=pd.read_csv("/content/database.csv")
data.head()
Date Time Latitude Longitude Type Depth Depth Error Depth Seismic Stations Magnitude Magnitude Type ... Magnitude Seismic Stations
0 01/02/1965 13:44:18 19.246 145.616 Earthquake 131.6 NaN NaN 6.0 MW ... NaN
1 01/04/1965 11:29:49 1.863 127.352 Earthquake 80.0 NaN NaN 5.8 MW ... NaN
2 01/05/1965 18:05:58 -20.579 -173.972 Earthquake 20.0 NaN NaN 6.2 MW ... NaN
3 01/08/1965 18:49:43 -59.076 -23.557 Earthquake 15.0 NaN NaN 5.8 MW ... NaN
4 01/09/1965 13:32:50 11.938 126.427 Earthquake 15.0 NaN NaN 5.8 MW ... NaN
Disk 80.59 GB available
Executing (17m 14s)
21:19
2023/10/26

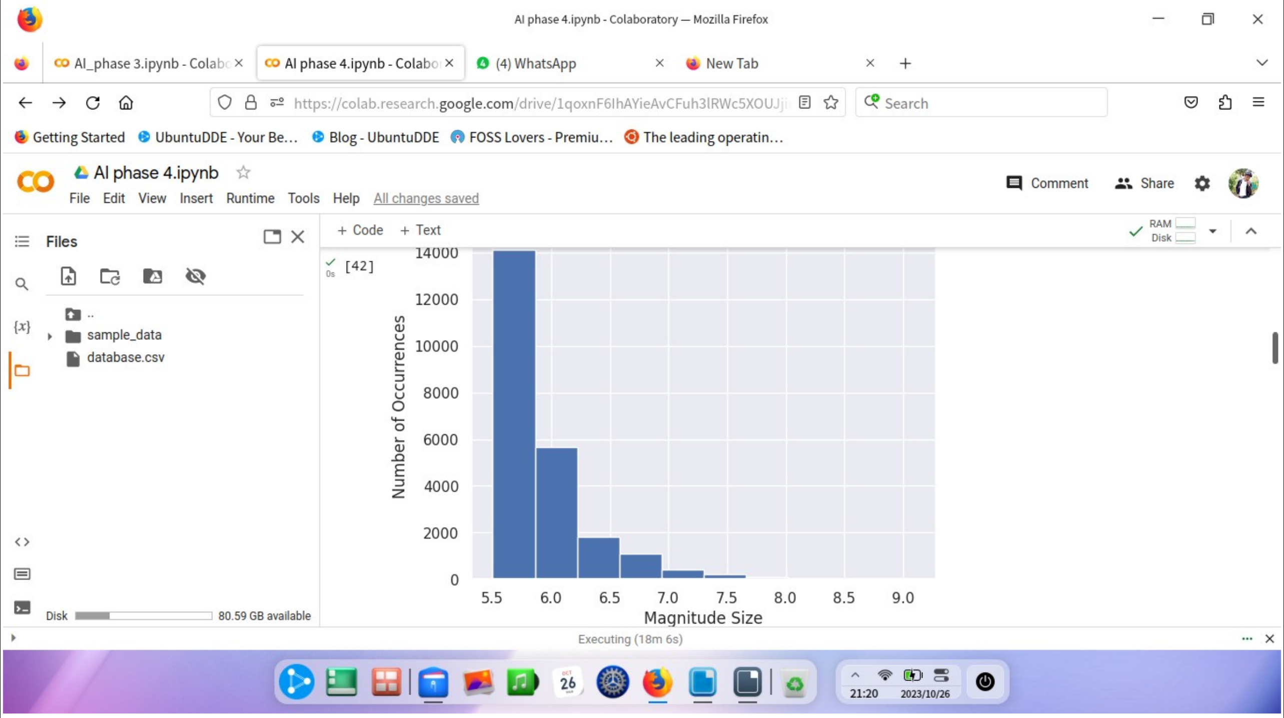

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI_phase 3.ipynb - Colab
AI phase 4.ipynb - Colab
(4) WhatsApp
New Tab
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Search
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
database.csv
[36] 5 rows × 21 columns
!pip3 install basemap
Requirement already satisfied: basemap in /usr/local/lib/python3.10/dist-packages (1.3.8)
Requirement already satisfied: basemap-data<1.4,>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from basemap)
Requirement already satisfied: pyshp<2.4,>=1.2 in /usr/local/lib/python3.10/dist-packages (from basemap) (2.3.1)
Requirement already satisfied: matplotlib<3.8,>=1.5 in /usr/local/lib/python3.10/dist-packages (from basemap) (3.7
Requirement already satisfied: pyproj<3.7.0,>=1.9.3 in /usr/local/lib/python3.10/dist-packages (from basemap) (3.6
Requirement already satisfied: numpy<1.26,>=1.21 in /usr/local/lib/python3.10/dist-packages (from basemap) (1.23.5
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>=
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>=1.5-
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>=1
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>=1.5
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.8,>=
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from pyproj<3.7.0,>=1.9.3->base
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->mat
Disk 80.59 GB available
Executing (17m 34s)
21:19
2023/10/26

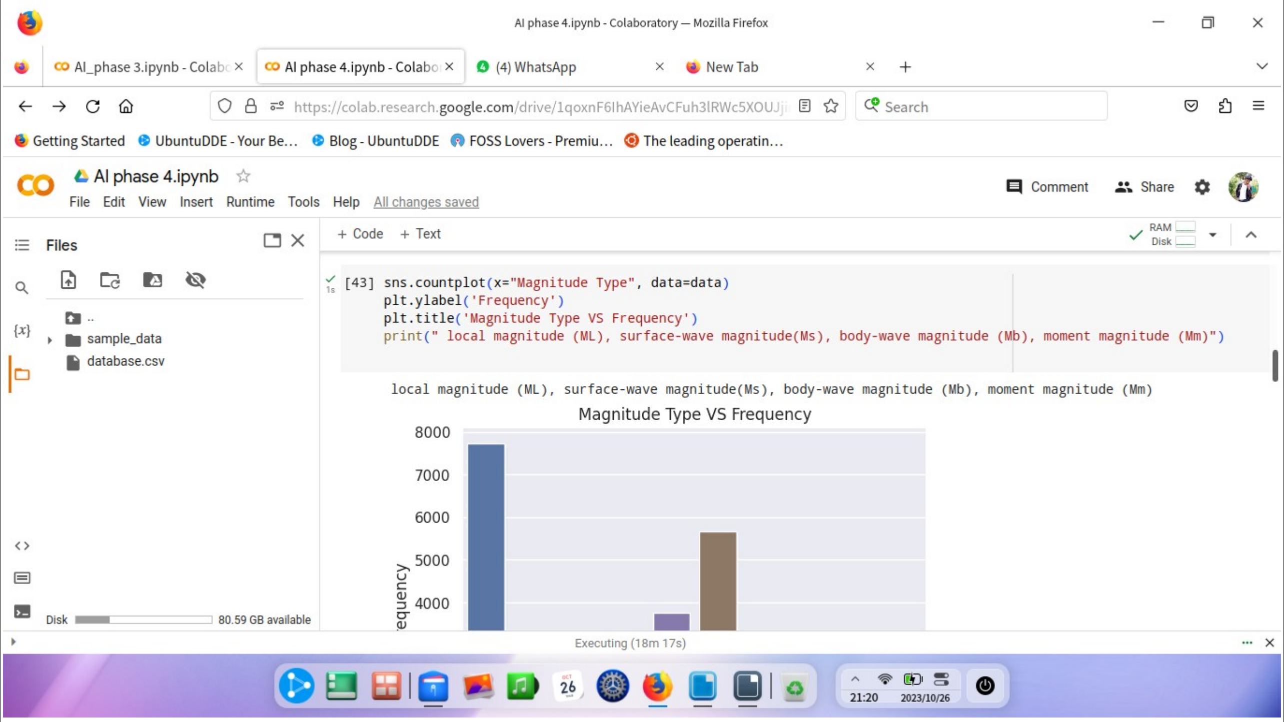

Screenshot finished
AI_phase 3.ipynb - Colab
AI phase 4.ipynb
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
sample_data
database.csv
[38] import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap
import seaborn as sns
sns.set(style="darkgrid")
print("Min Value: "+ str(data['Magnitude'].min()))
print("Max Value: " + str(data['Magnitude'].max()))
Min Value: 5.5
Max Value: 9.1
[40] Greater_8 = data[data['Magnitude'] > 8]
Greater_8['Location Source'].value_counts()
US 22
ISCGEM 5
Name: Location Source, dtype: int64
Disk 80.59 GB available
Executing (17m 39s)
21:19
2023/10/26

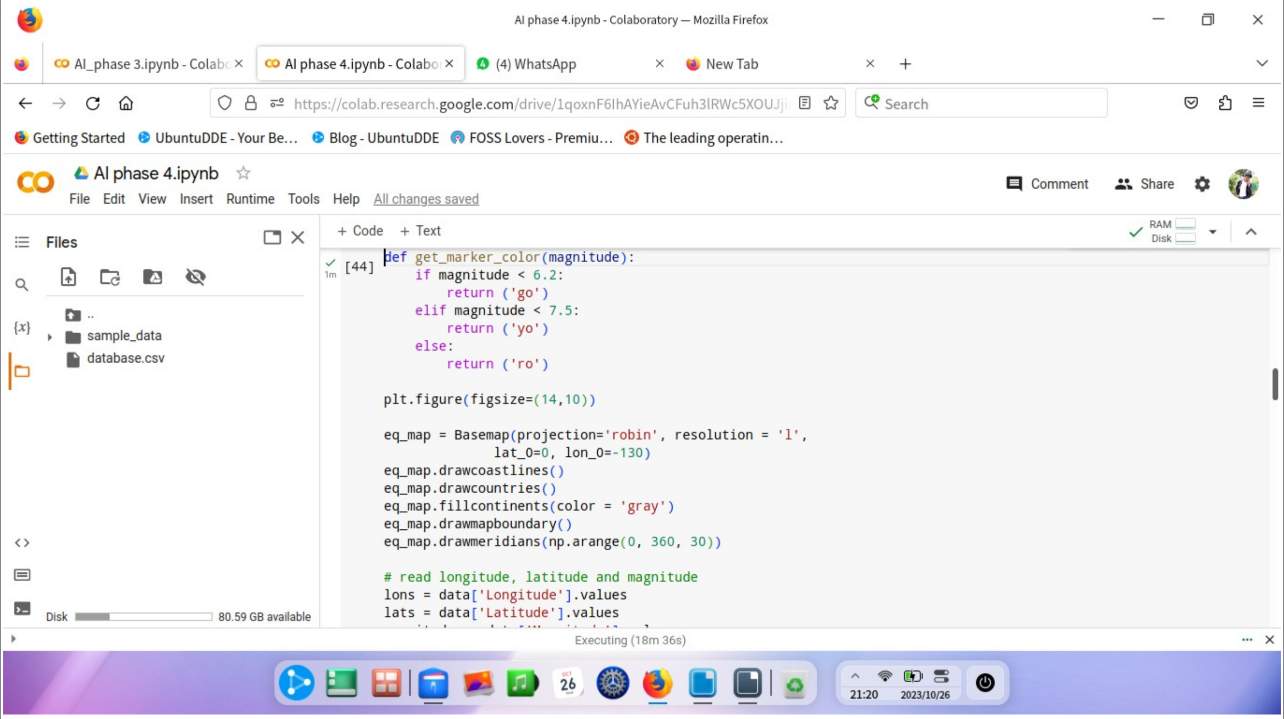

Screenshot finished
AI_phase 3.ipynb - Colab
AI phase 4.ipynb
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
sample_data
database.csv
Greater_7 = data[data['Magnitude'] > 7]
Greater_7['Location Source'].value_counts()
Greater_6 = data[data['Magnitude'] > 6]
Greater_6['Location Source'].value_counts()
Greater_5 = data[data['Magnitude'] > 5]
Greater_5['Location Source'].value_counts()
Greater_4 = data[data['Magnitude'] > 4]
Greater_4['Location Source'].value_counts()
US 20350
ISCGEM 2581
CI 61
GCMT 56
NC 54
GUC 46
AEIC 40
UNM 21
PGC 19
WEL 18
Disk 80.59 GB available
Executing (17m 44s)
21:19
2023/10/26

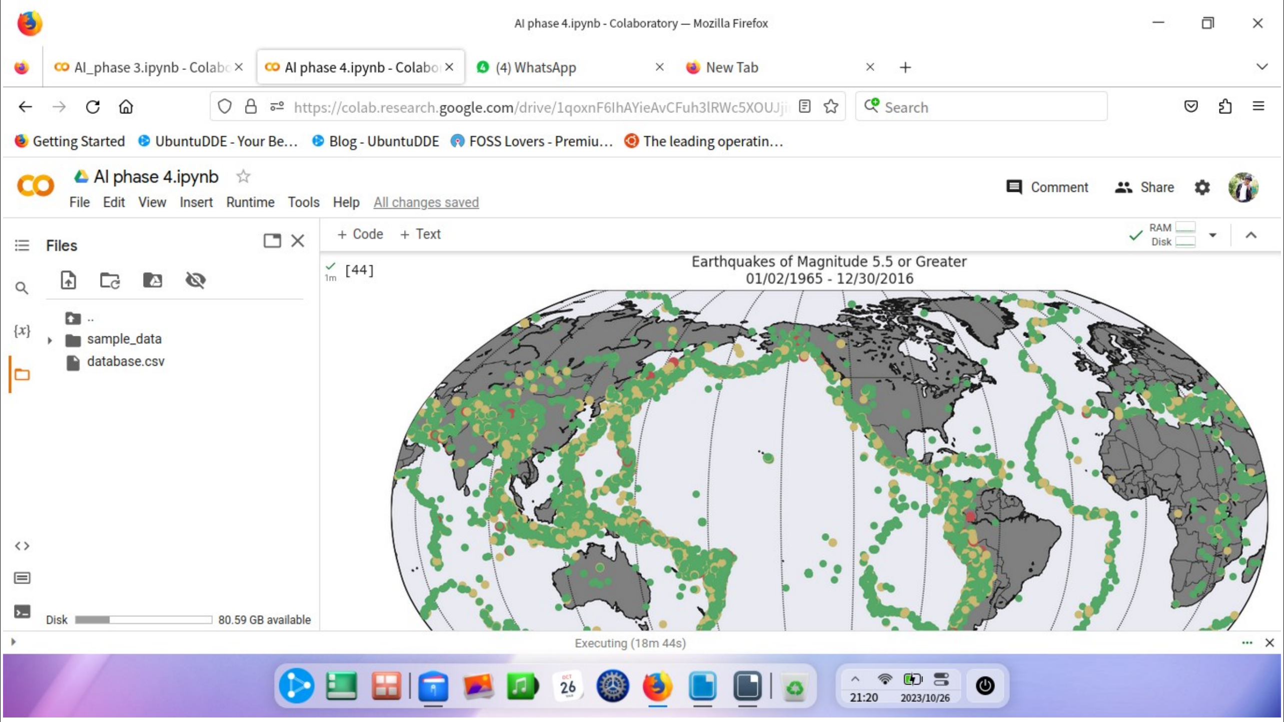

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
database.csv
[42] plt.hist(data['Magnitude'])
plt.xlabel('Magnitude Size')
plt.ylabel('Number of Occurrences')
Text(0, 0.5, 'Number of Occurrences')
14000
12000
10000
8000
6000
Executing (18m 0s)
21:20
2023/10/26

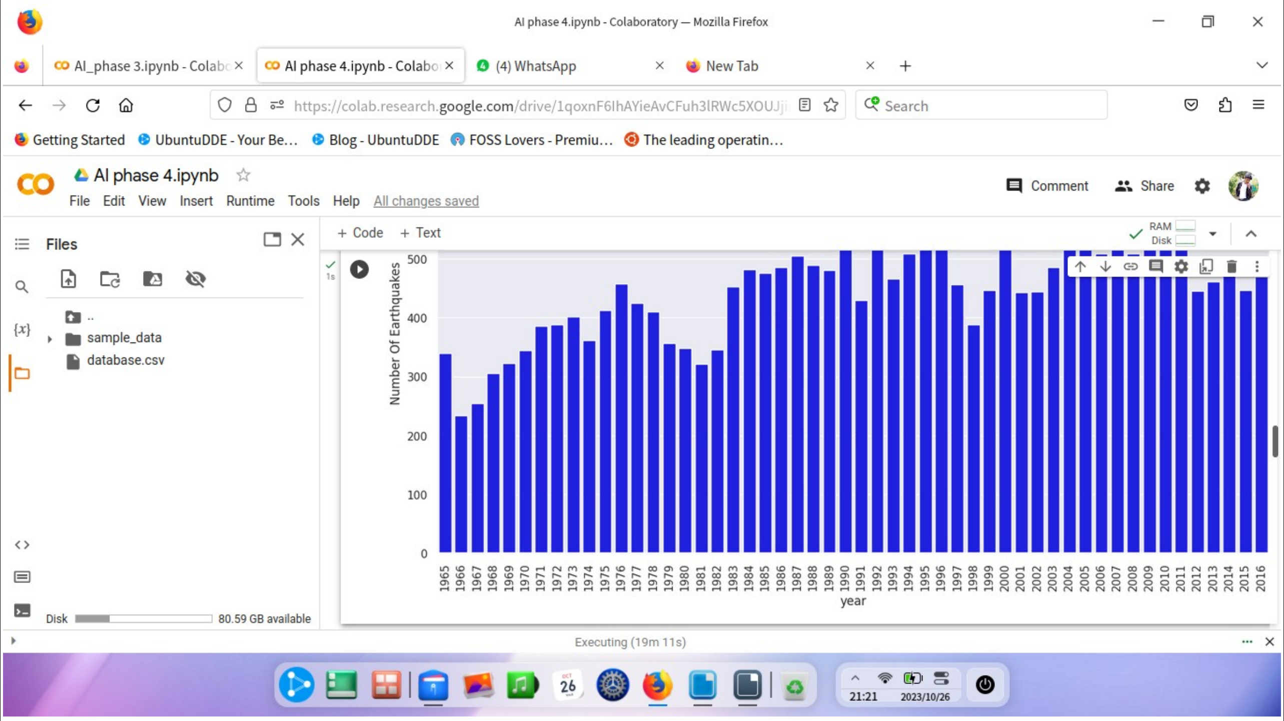

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
database.csv
[42]
Number of Occurrences
14000
12000
10000
8000
6000
4000
2000
0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
Magnitude Size
Executing (18m 6s)
21:20
2023/10/26

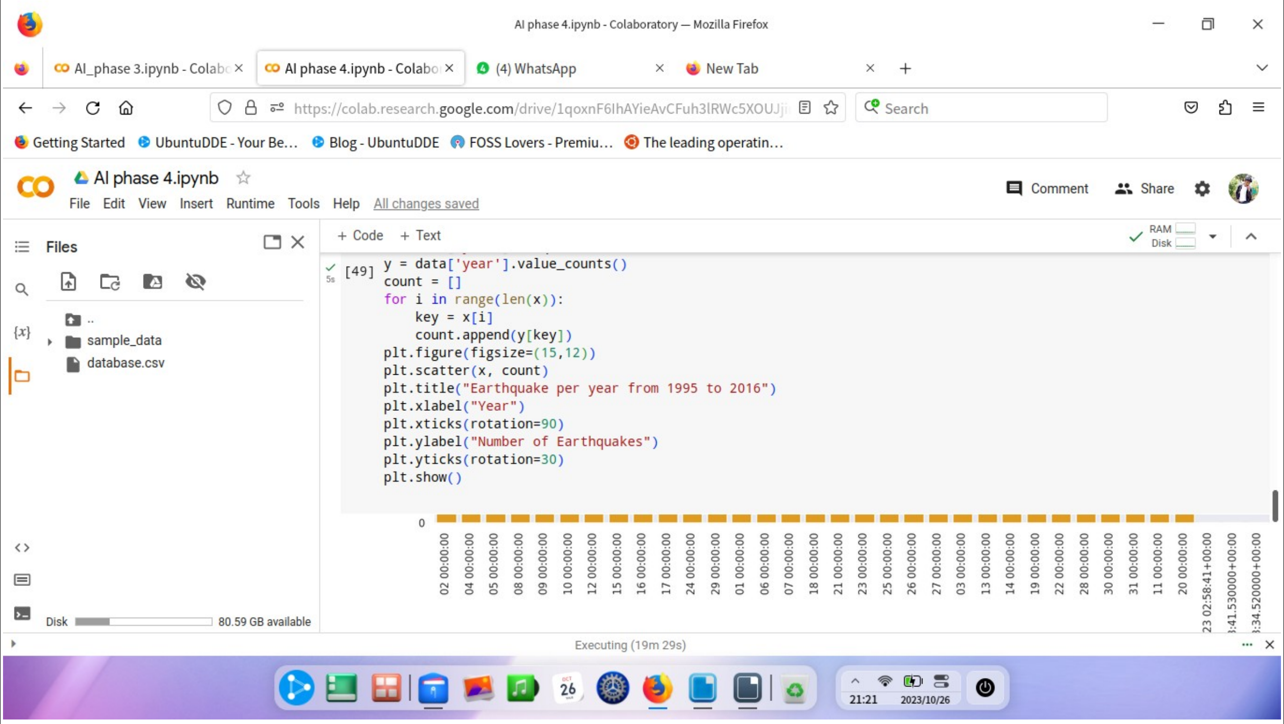

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
[42]
Number of Occurrences
14000
12000
10000
8000
6000
4000
2000
0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
Magnitude Size
Executing (18m 6s)

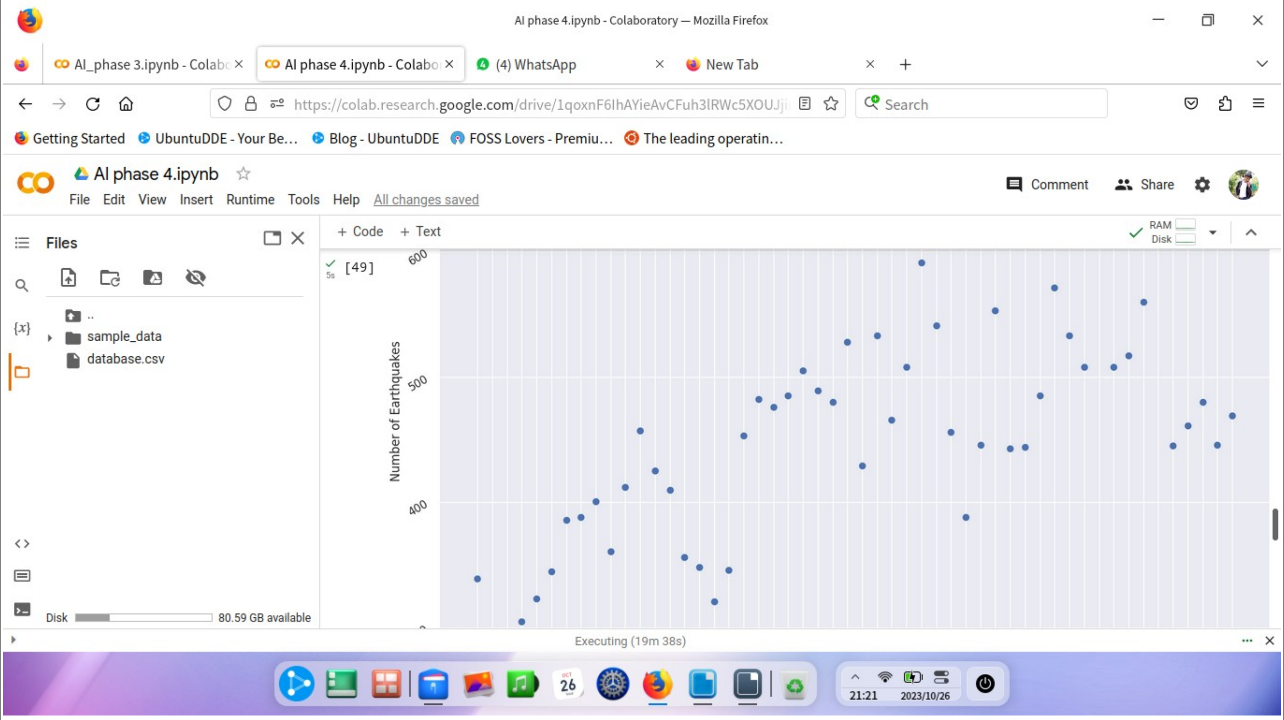

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
[43] sns.countplot(x="Magnitude Type", data=data)
plt.ylabel('Frequency')
plt.title('Magnitude Type VS Frequency')
print(" local magnitude (ML), surface-wave magnitude(Ms), body-wave magnitude (Mb), moment magnitude (Mm)")
local magnitude (ML), surface-wave magnitude(Ms), body-wave magnitude (Mb), moment magnitude (Mm)
Magnitude Type VS Frequency
8000
7000
6000
5000
4000
Executing (18m 17s)

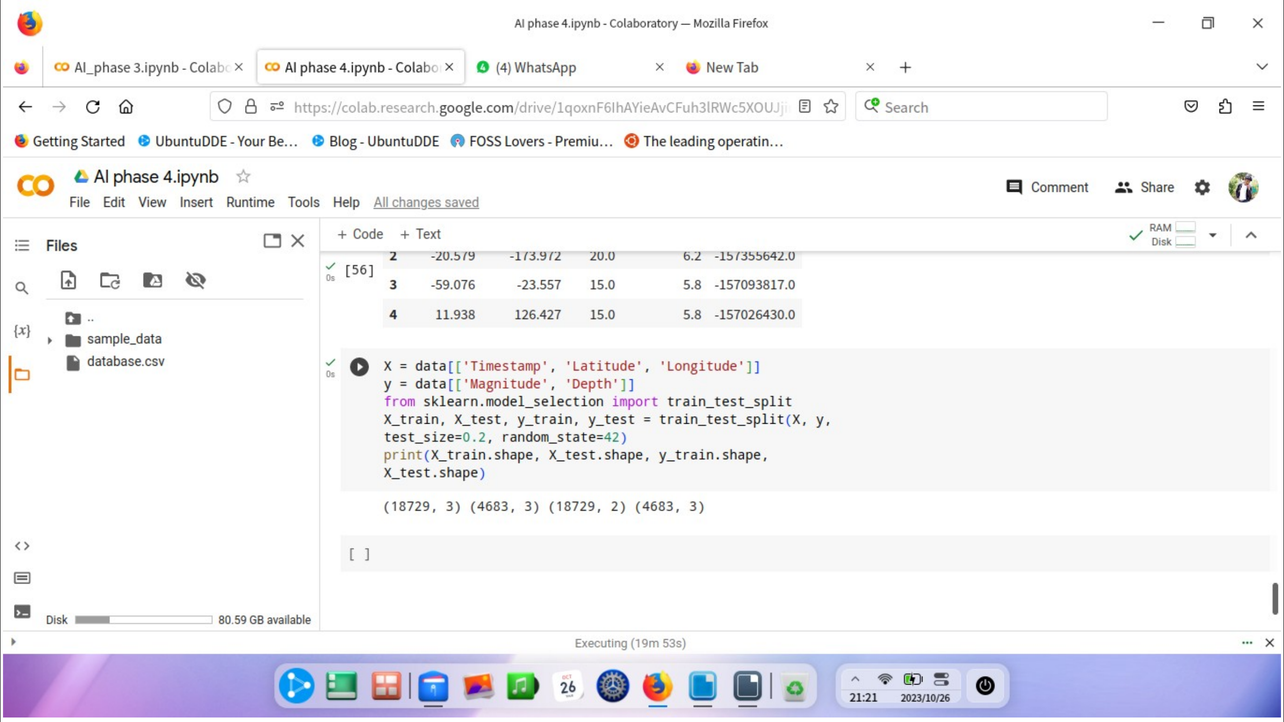

AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI_phase 3.ipynb - Colab
AI phase 4.ipynb - Colab
(4) WhatsApp
New Tab
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Search
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
..
sample_data
database.csv
[44]
def get_marker_color(magnitude):
    if magnitude < 6.2:
        return ('go')
    elif magnitude < 7.5:
        return ('yo')
    else:
        return ('ro')
plt.figure(figsize=(14,10))
eq_map = Basemap(projection='robin', resolution = 'l',
              lat_0=0, lon_0=-130)
eq_map.drawcoastlines()
eq_map.drawcountries()
eq_map.fillcontinents(color = 'gray')
eq_map.drawmapboundary()
eq_map.drawmeridians(np.arange(0, 360, 30))
# read longitude, latitude and magnitude
lons = data['Longitude'].values
lats = data['Latitude'].values
Disk
80.59 GB available
Executing (18m 36s)
21:20
2023/10/26


AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI_phase 3.ipynb - Colab
AI phase 4.ipynb - Colab
(4) WhatsApp
New Tab
https://colab.research.google.com/drive/1qoxnF6lhAYieAvCFuh3lRWc5XOUJ
Search
Getting Started
UbuntuDDE - Your Be...
Blog - UbuntuDDE
FOSS Lovers - Premiu...
The leading operatin...
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
..
sample_data
database.csv
[44]
Earthquakes of Magnitude 5.5 or Greater
01/02/1965 - 12/30/2016
Disk
80.59 GB available
Executing (18m 44s)
21:20
2023/10/26


AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
database.csv
Number Of Earthquakes
year
Disk 80.59 GB available
Executing (19m 11s)


AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
[49]
y = data['year'].value_counts()
count = []
for i in range(len(x)):
    key = x[i]
    count.append(y[key])
plt.figure(figsize=(15,12))
plt.scatter(x, count)
plt.title("Earthquake per year from 1995 to 2016")
plt.xlabel("Year")
plt.xticks(rotation=90)
plt.ylabel("Number of Earthquakes")
plt.yticks(rotation=30)
plt.show()
Disk 80.59 GB available
Executing (19m 29s)


AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
database.csv
[49]
Number of Earthquakes
600
500
400
Disk 80.59 GB available
Executing (19m 38s)


AI phase 4.ipynb - Colaboratory — Mozilla Firefox
AI phase 4.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
database.csv
[56]
2 -20.579 -173.972 20.0 6.2 -157355642.0
3 -59.076 -23.557 15.0 5.8 -157093817.0
4 11.938 126.427 15.0 5.8 -157026430.0
X = data[['Timestamp', 'Latitude', 'Longitude']]
y = data[['Magnitude', 'Depth']]
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
print(X_train.shape, X_test.shape, y_train.shape,
X_test.shape)
(18729, 3) (4683, 3) (18729, 2) (4683, 3)
Disk 80.59 GB available
Executing (19m 53s)

CONCLUSION:

In conclusion, the data visualization efforts employing tools suchas Basemap have provided crucial insights into the geographical distribution of earthquakes, offering a comprehensive view of seismic activities worldwide. This spatial understanding is pivotalfor identifying regions prone to seismic events and informs subsequent modeling endeavors. Simultaneously, the strategic process of data splitting into training and testing sets marks a crucial preparatory phase in developing a robust earthquake prediction model.