earthquake_explorer_project

October 5, 2023

Column Name	Description
Date	The date of the earthquake (e.g., "01/02/1965").
Time	The time of the earthquake (e.g., "13:44:18").
Latitude	The latitude coordinate of the earthquake (e.g., "19.246").
Longitude	The longitude coordinate of the earthquake (e.g., "145.616").
Type	The type of the earthquake (e.g., "Earthquake").
Depth	The depth of the earthquake (e.g., "131.6").
Depth Error	The error in depth (sometimes missing).
Depth Seismic Stations	The number of seismic stations at depth (sometimes missing).
Magnitude	The magnitude of the earthquake (e.g., "6.0").
Magnitude Type	The type of magnitude (e.g., "MW").
Magnitude Seismic Stations	The number of seismic stations for magnitude (sometimes missing).
Azimuthal Gap	The azimuthal gap (sometimes missing).
Horizontal Distance	The horizontal distance (sometimes missing).
Horizontal Error	The horizontal error (sometimes missing).
Root Mean Square	The root mean square (sometimes missing).
ID	The ID of the earthquake (e.g., "ISCGEM860706").
Source	The source of the data (e.g., "ISCGEM").
Location Source	The source of the location (e.g., "ISCGEM").
Magnitude Source	The source of magnitude information (e.g., "ISCGEM").
Status	The status of the earthquake (e.g., "Automatic").

Project: Earthquake Data Analysis

In this project, we will analyze a dataset of significant earthquakes worldwide from 1965 to 2016. We will perform data cleaning, exploratory data analysis (EDA), and data visualization.

Step 1: Load the Dataset

We start by importing the necessary libraries and loading the dataset from the "DATA" folder. We then display the first five observations to get an overview of the data.

[ ]:

```python
# Load the Dataset

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

We start by loading the dataset from the "data" folder. The dataset, named "database.csv," is loaded using the Pandas library.

[ ]:

```python
data_file = "data/database.csv"
```

You can do some basic data exploration to examine the contents of your dataset. For example, you can use the head() function to display the first five observations.

[ ]:

```python
# Data Discovery (EDA):
df = pd.read_csv(data_file)
df.head()
```

[ ]:

```
         Date      Time  Latitude  Longitude        Type  Depth  Depth Error  \
0  01/02/1965  13:44:18    19.246    145.616  Earthquake  131.6          NaN
1  01/04/1965  11:29:49     1.863    127.352  Earthquake   80.0          NaN
2  01/05/1965  18:05:58   -20.579   -173.972  Earthquake   20.0          NaN
3  01/08/1965  18:49:43   -59.076    -23.557  Earthquake   15.0          NaN
4  01/09/1965  13:32:50    11.938    126.427  Earthquake   15.0          NaN

   Depth Seismic Stations  Magnitude Magnitude Type  ...  \
0                     NaN        6.0             MW  ...
1                     NaN        5.8             MW  ...
2                     NaN        6.2             MW  ...
3                     NaN        5.8             MW  ...
4                     NaN        5.8             MW  ...

   Magnitude Seismic Stations  Azimuthal Gap  Horizontal Distance  \
0                         NaN            NaN                  NaN
1                         NaN            NaN                  NaN
2                         NaN            NaN                  NaN
3                         NaN            NaN                  NaN
4                         NaN            NaN                  NaN

   Horizontal Error  Root Mean Square            ID  Source Location Source  \
0               NaN               NaN  ISCGEM860706  ISCGEM          ISCGEM
1               NaN               NaN  ISCGEM860737  ISCGEM          ISCGEM
2               NaN               NaN  ISCGEM860762  ISCGEM          ISCGEM
3               NaN               NaN  ISCGEM860856  ISCGEM          ISCGEM
4               NaN               NaN  ISCGEM860890  ISCGEM          ISCGEM

  Magnitude Source     Status
0           ISCGEM  Automatic
1           ISCGEM  Automatic
2           ISCGEM  Automatic
3           ISCGEM  Automatic
4           ISCGEM  Automatic
```

```
[5 rows x 21 columns]
```

[ ]:
```
df.tail()
```

[ ]:
```
             Date      Time  Latitude  Longitude        Type  Depth  \
23407  12/28/2016  08:22:12   38.3917  -118.8941  Earthquake  12.30
23408  12/28/2016  09:13:47   38.3777  -118.8957  Earthquake   8.80
23409  12/28/2016  12:38:51   36.9179   140.4262  Earthquake  10.00
23410  12/29/2016  22:30:19   -9.0283   118.6639  Earthquake  79.00
23411  12/30/2016  20:08:28   37.3973   141.4103  Earthquake  11.94

       Depth Error  Depth Seismic Stations  Magnitude Magnitude Type  …  \
23407          1.2                    40.0        5.6             ML  …
23408          2.0                    33.0        5.5             ML  …
23409          1.8                     NaN        5.9            MWW  …
23410          1.8                     NaN        6.3            MWW  …
23411          2.2                     NaN        5.5             MB  …

       Magnitude Seismic Stations  Azimuthal Gap  Horizontal Distance  \
23407                        18.0          42.47                0.120
23408                        18.0          48.58                0.129
23409                         NaN          91.00                0.992
23410                         NaN          26.00                3.553
23411                       428.0          97.00                0.681

       Horizontal Error  Root Mean Square          ID Source Location Source  \
23407               NaN            0.1898  NN00570710     NN              NN
23408               NaN            0.2187  NN00570744     NN              NN
23409               4.8            1.5200  US10007NAF     US              US
23410               6.0            1.4300  US10007NL0     US              US
23411               4.5            0.9100  US10007NTD     US              US

      Magnitude Source    Status
23407               NN  Reviewed
23408               NN  Reviewed
23409               US  Reviewed
23410               US  Reviewed
23411               US  Reviewed

[5 rows x 21 columns]
```

[ ]:
```
df.columns
```

[ ]:
```
Index(['Date', 'Time', 'Latitude', 'Longitude', 'Type', 'Depth', 'Depth Error',
       'Depth Seismic Stations', 'Magnitude', 'Magnitude Type',
       'Magnitude Error', 'Magnitude Seismic Stations', 'Azimuthal Gap',
       'Horizontal Distance', 'Horizontal Error', 'Root Mean Square', 'ID',
       'Source', 'Location Source', 'Magnitude Source', 'Status'],
      dtype='object')
```

[ ]:
```
df.shape
```

[ ]:
```
(23412, 21)
```

[ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23412 entries, 0 to 23411
Data columns (total 21 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   Date                        23412 non-null  object 
 1   Time                        23412 non-null  object 
 2   Latitude                    23412 non-null  float64
 3   Longitude                   23412 non-null  float64
 4   Type                        23412 non-null  object 
 5   Depth                       23412 non-null  float64
 6   Depth Error                 4461 non-null   float64
 7   Depth Seismic Stations      7097 non-null   float64
 8   Magnitude                   23412 non-null  float64
 9   Magnitude Type              23409 non-null  object 
 10  Magnitude Error             327 non-null    float64
 11  Magnitude Seismic Stations  2564 non-null   float64
 12  Azimuthal Gap               7299 non-null   float64
 13  Horizontal Distance         1604 non-null   float64
 14  Horizontal Error            1156 non-null   float64
 15  Root Mean Square            17352 non-null  float64
 16  ID                          23412 non-null  object 
 17  Source                      23412 non-null  object 
 18  Location Source             23412 non-null  object 
 19  Magnitude Source            23412 non-null  object 
 20  Status                      23412 non-null  object 
dtypes: float64(12), object(9)
memory usage: 3.8+ MB
```

[ ]:
```
df.describe()
```

[ ]:
```
           Latitude     Longitude         Depth  Depth Error  \
count  23412.000000  23412.000000  23412.000000  4461.000000   
mean       1.679033     39.639961     70.767911     4.993115   
std       30.113183    125.511959    122.651898     4.875184   
min      -77.080000   -179.997000     -1.100000     0.000000   
25%      -18.653000    -76.349750     14.522500     1.800000   
50%       -3.568500    103.982000     33.000000     3.500000   
```

```
75%       26.190750    145.026250     54.000000     6.300000
max       86.005000    179.998000    700.000000    91.295000

       Depth Seismic Stations     Magnitude  Magnitude Error  \
count             7097.000000  23412.000000       327.000000
mean               275.364098      5.882531         0.071820
std                162.141631      0.423066         0.051466
min                  0.000000      5.500000         0.000000
25%                146.000000      5.600000         0.046000
50%                255.000000      5.700000         0.059000
75%                384.000000      6.000000         0.075500
max                934.000000      9.100000         0.410000

       Magnitude Seismic Stations  Azimuthal Gap  Horizontal Distance  \
count                 2564.000000    7299.000000          1604.000000
mean                    48.944618      44.163532             3.992660
std                     62.943106      32.141486             5.377262
min                      0.000000       0.000000             0.004505
25%                     10.000000      24.100000             0.968750
50%                     28.000000      36.000000             2.319500
75%                     66.000000      54.000000             4.724500
max                    821.000000     360.000000            37.874000

       Horizontal Error  Root Mean Square
count       1156.000000      17352.000000
mean           7.662759          1.022784
std           10.430396          0.188545
min            0.085000          0.000000
25%            5.300000          0.900000
50%            6.700000          1.000000
75%            8.100000          1.130000
max           99.000000          3.440000
```

[ ]:
```python
# Check for missing data
df.isnull().sum()
```

[ ]:
```
Date                              0
Time                              0
Latitude                          0
Longitude                         0
Type                              0
Depth                             0
Depth Error                   18951
Depth Seismic Stations        16315
Magnitude                         0
Magnitude Type                    3
Magnitude Error               23085
Magnitude Seismic Stations    20848
Azimuthal Gap                 16113
Horizontal Distance           21808
Horizontal Error              22256
Root Mean Square               6060
ID                                0
Source                            0
Location Source                   0
Magnitude Source                  0
Status                            0
dtype: int64
```

[ ]:

```
df.dtypes
```

[ ]:

```
Date                           object
Time                           object
Latitude                      float64
Longitude                     float64
Type                           object
Depth                         float64
Depth Error                   float64
Depth Seismic Stations        float64
Magnitude                     float64
Magnitude Type                 object
Magnitude Error               float64
Magnitude Seismic Stations    float64
Azimuthal Gap                 float64
Horizontal Distance           float64
Horizontal Error              float64
Root Mean Square              float64
ID                             object
Source                         object
Location Source                object
Magnitude Source               object
Status                         object
dtype: object
```

[ ]:

```
# Eksik değerleri doldurma
df['Depth Error'].fillna(df['Depth Error'].mean(), inplace=True)
df['Magnitude Error'].fillna(df['Magnitude Error'].mean(), inplace=True)
```

Distribution by Date of Earthquakes:

In this section, we will visualize the number of earthquakes by year, as well as analyze the earthquakes by the beginning of each year.

[ ]:

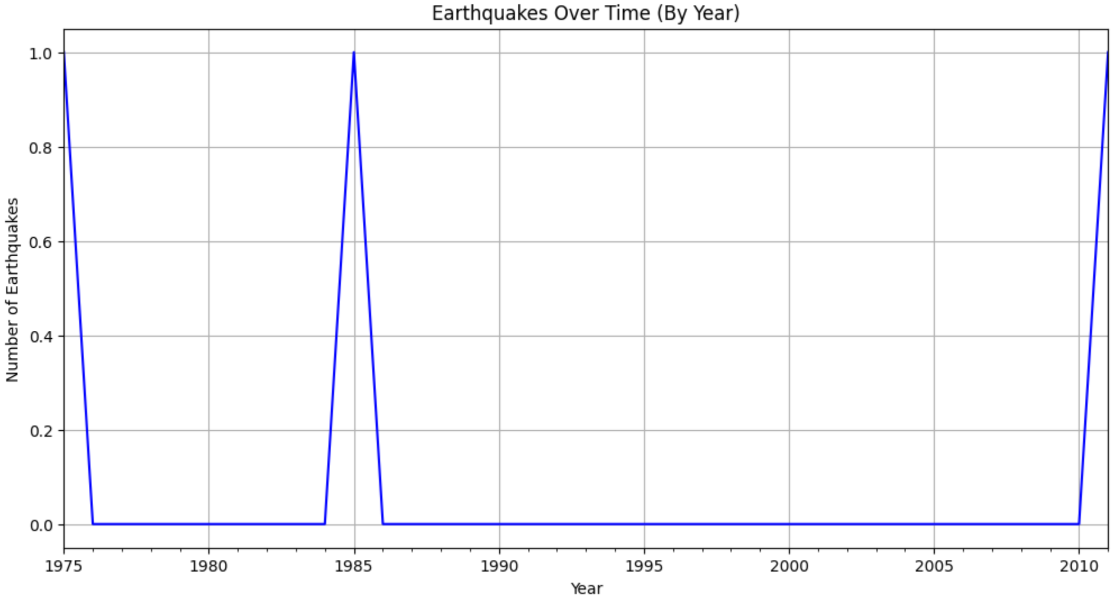
```
df['Date'] = pd.to_datetime(df['Date'], format='%Y-%m-%dT%H:%M:%S.%fZ',␣
 ↪errors='coerce')
```

[ ]:

```
earthquake_count_by_year = df.resample('Y', on='Date').size()
```

[ ]:

```
plt.figure(figsize=(12, 6))
earthquake_count_by_year.plot(kind='line', color='blue')
plt.title('Earthquakes Over Time (By Year)')
plt.xlabel('Year')
plt.ylabel('Number of Earthquakes')
plt.grid(True)
plt.show()
```



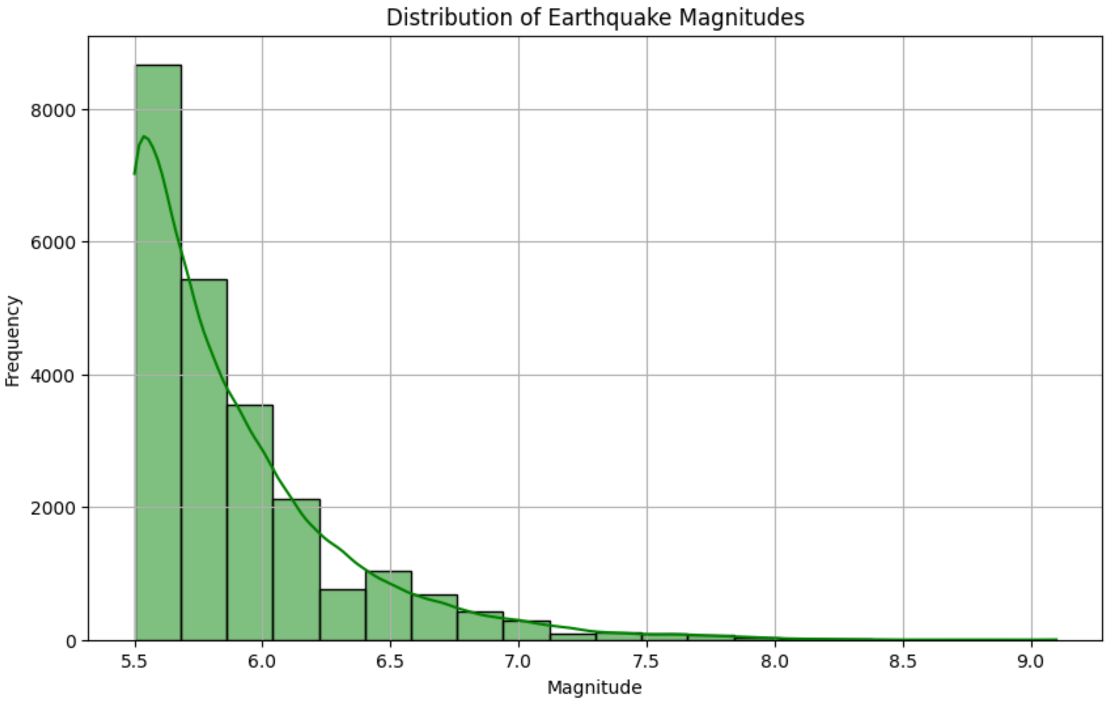
This code will create a line chart showing how earthquakes are distributed over the years.

Magnitude Distribution of Earthquakes:

Now, let's examine the magnitude distribution of earthquakes and visualize this distribution as a histogram.

[ ]:

```
# Depremlerin büyüklüklerini histogram olarak çizdirin
plt.figure(figsize=(10, 6))
sns.histplot(df['Magnitude'], bins=20, color='green', kde=True)
plt.title('Distribution of Earthquake Magnitudes')
plt.xlabel('Magnitude')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



[ ]:
```
type_counts = df['Type'].value_counts()
print(type_counts)
```

```
Type
Earthquake           23232
Nuclear Explosion      175
Explosion                4
Rock Burst               1
Name: count, dtype: int64
```

Correlation Analysis:

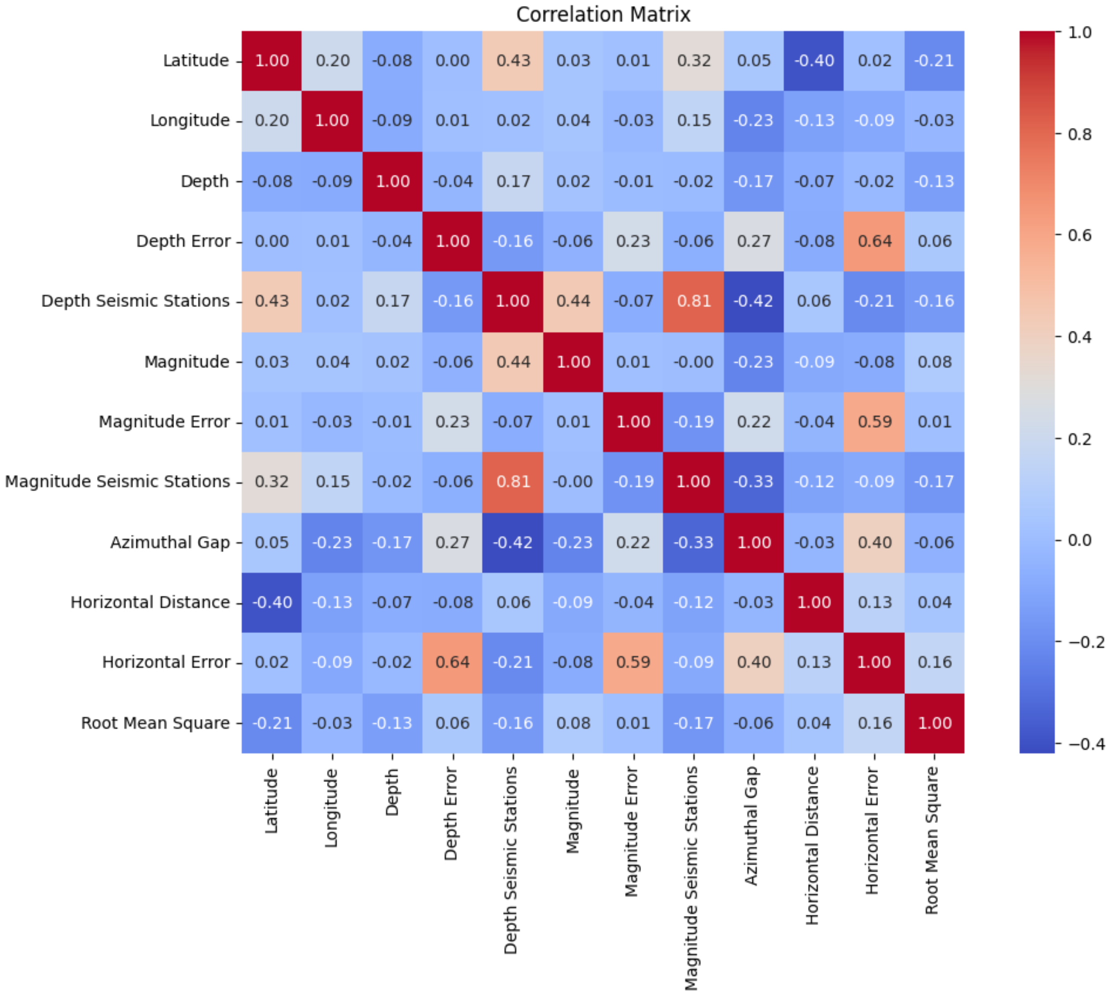
You can calculate the correlation matrix to examine relationships between numerical columns in your data set. The correlation matrix shows how related the columns are to each other.

[ ]:
```
# Sayısal olmayan sütunları çıkar
numerical_columns = df.select_dtypes(include=['float64']).columns
numerical_df = df[numerical_columns]
```

[ ]:
```
# Korelasyon matrisini hesapla
correlation_matrix = numerical_df.corr()
```

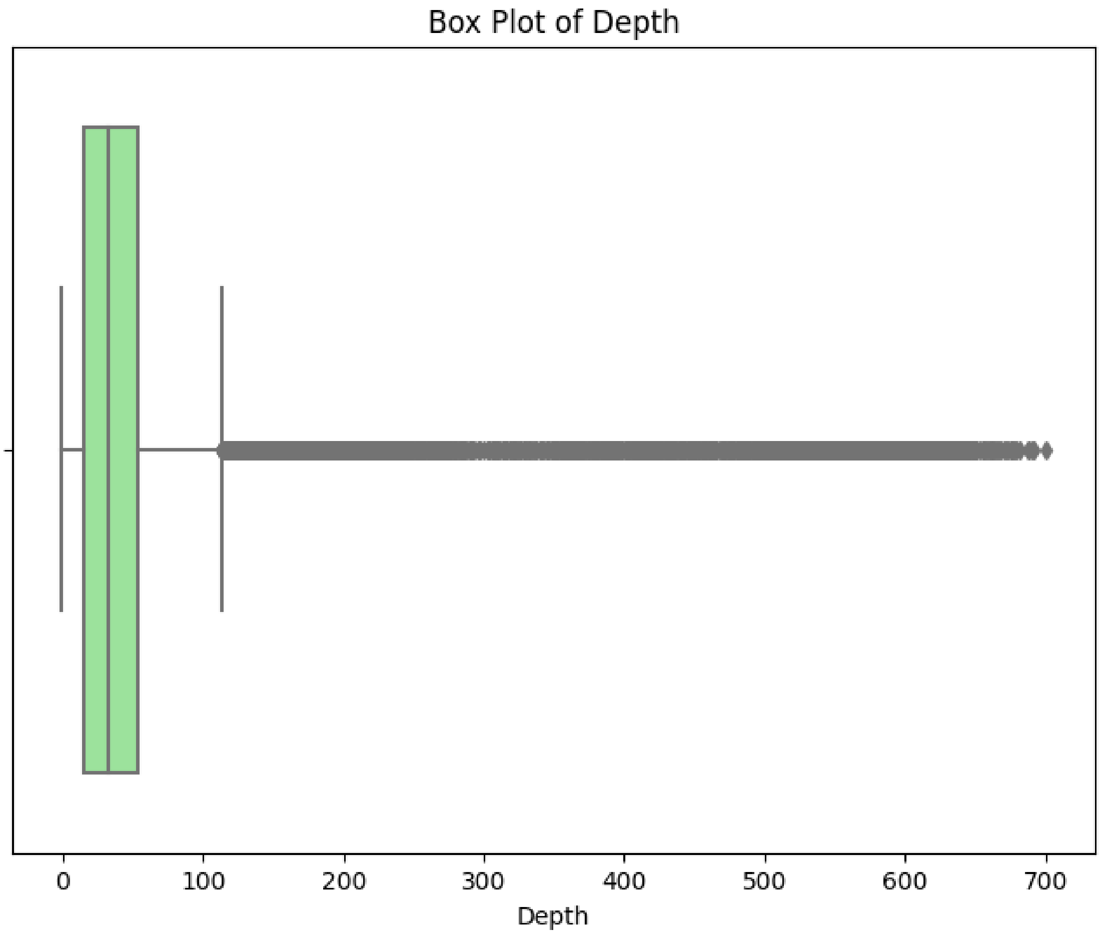
[ ]:
```
# Korelasyon matrisini görselleştir
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",␣
  ↪square=True)
```

```
plt.title('Correlation Matrix')
plt.show()
```



1. More Visualization

[ ]:
```
# Box plot of Depth
plt.figure(figsize=(8, 6))
sns.boxplot(x='Depth', data=df, color='lightgreen')
plt.title('Box Plot of Depth')
plt.xlabel('Depth')
plt.show()
```



2. Special Analysis Annual Distribution of Major Earthquakes

[ ]:

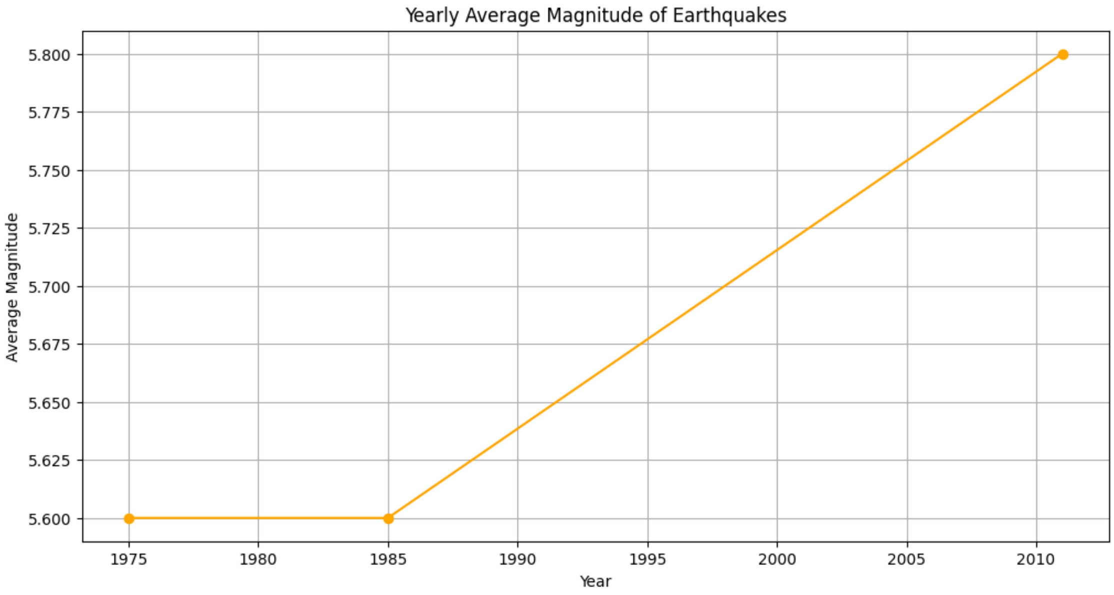
```
# Extract year from Date column
df['Year'] = pd.to_datetime(df['Date']).dt.year
```

[ ]:

```
# Group by year and calculate the average magnitude
yearly_avg_magnitude = df.groupby('Year')['Magnitude'].mean()
```

[ ]:

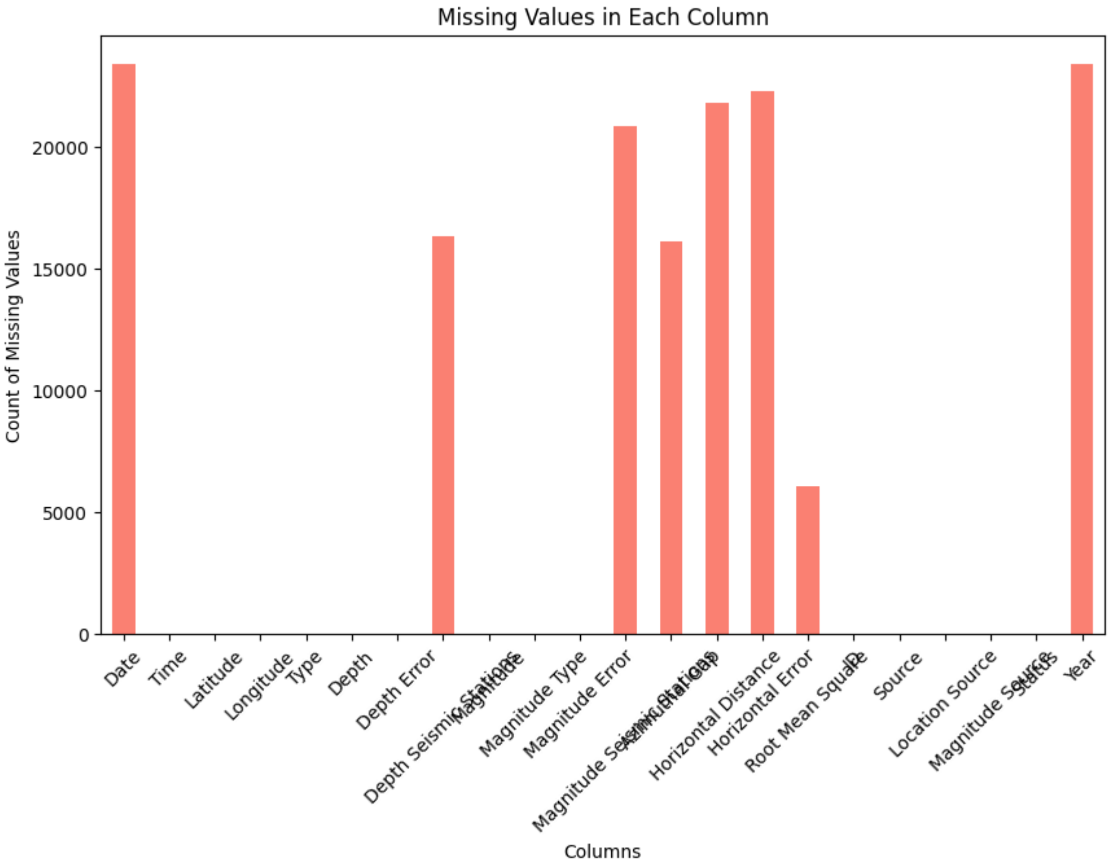
```
# Plot the yearly average magnitude
plt.figure(figsize=(12, 6))
yearly_avg_magnitude.plot(marker='o', color='orange')
plt.title('Yearly Average Magnitude of Earthquakes')
plt.xlabel('Year')
plt.ylabel('Average Magnitude')
plt.grid(True)
plt.show()
```



3. Incomplete Data Processing

[ ]:
```python
# Count missing values in each column
missing_values = df.isnull().sum()
```

[ ]:
```python
# Plot missing values
plt.figure(figsize=(10, 6))
missing_values.plot(kind='bar', color='salmon')
plt.title('Missing Values in Each Column')
plt.xlabel('Columns')
plt.ylabel('Count of Missing Values')
plt.xticks(rotation=45)
plt.show()
```



4. Statistical Analysis Basic Statistics

```
[ ]: # Calculate basic statistics
     basic_stats = df[['Latitude', 'Longitude', 'Depth', 'Magnitude']].describe()
```

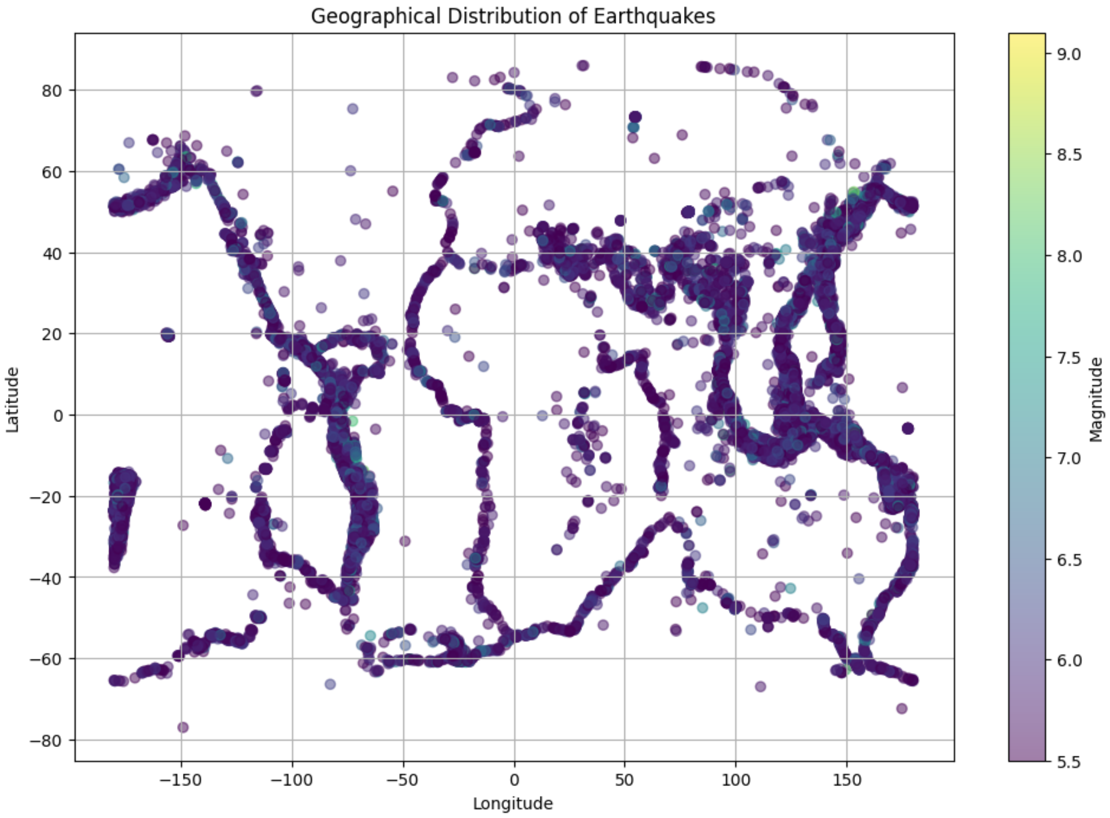
```
[ ]: # Display basic statistics
     basic_stats
```

```
[ ]:            Latitude     Longitude         Depth     Magnitude
     count  23412.000000  23412.000000  23412.000000  23412.000000
     mean       1.679033     39.639961     70.767911      5.882531
     std       30.113183    125.511959    122.651898      0.423066
     min      -77.080000   -179.997000     -1.100000      5.500000
     25%      -18.653000    -76.349750     14.522500      5.600000
     50%       -3.568500    103.982000     33.000000      5.700000
     75%       26.190750    145.026250     54.000000      6.000000
     max       86.005000    179.998000    700.000000      9.100000
```

5. Geographic Analyzes Geographic Distribution Map

[ ]:

```
# Create a scatter plot of earthquake locations
plt.figure(figsize=(12, 8))
plt.scatter(df['Longitude'], df['Latitude'], c=df['Magnitude'], cmap='viridis',␣
  ↪alpha=0.5)
plt.title('Geographical Distribution of Earthquakes')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.colorbar(label='Magnitude')
plt.grid(True)
plt.show()
```



6. Time Series Analysis Distribution of Earthquakes over Time

[ ]:

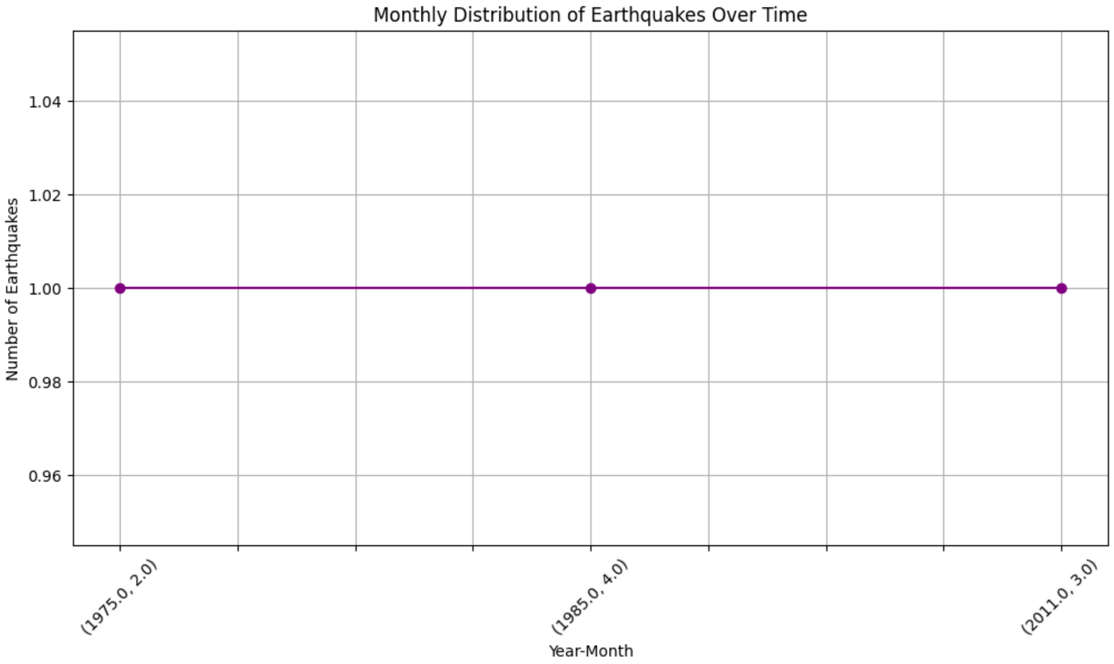
```
# Convert Date column to datetime
df['Date'] = pd.to_datetime(df['Date'], format='%m/%d/%Y')
```

[ ]:

```
# Group by year and month, and count the number of earthquakes
monthly_earthquakes = df.groupby([df['Date'].dt.year, df['Date'].dt.
  ↪month])['Magnitude'].count()
```

[ ]:

```
# Plot the number of earthquakes over time
plt.figure(figsize=(12, 6))
monthly_earthquakes.plot(kind='line', marker='o', color='purple')
plt.title('Monthly Distribution of Earthquakes Over Time')
plt.xlabel('Year-Month')
plt.ylabel('Number of Earthquakes')
plt.grid(True)
plt.xticks(rotation=45)
plt.show()
```



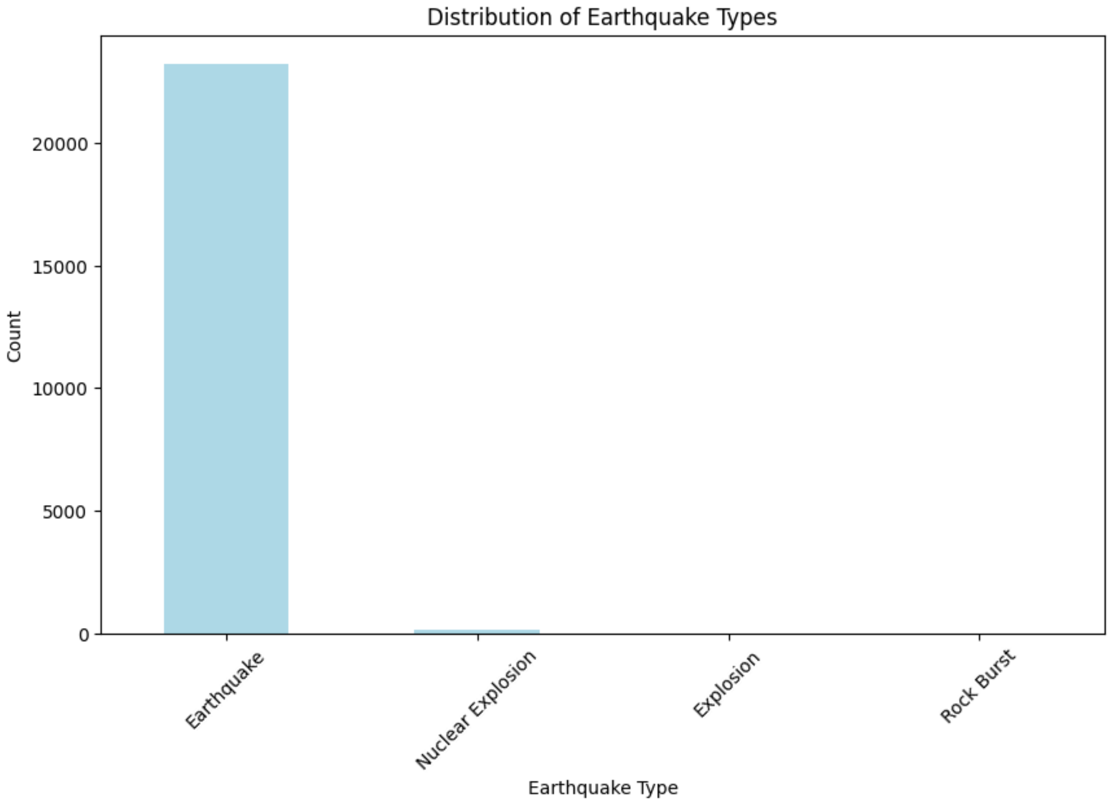
7. Categorical Data Analysis Type Distribution of Earthquakes

[ ]:

```
# Count the number of earthquakes for each type
earthquake_types = df['Type'].value_counts()
```

[ ]:

```
# Plot the distribution of earthquake types
plt.figure(figsize=(10, 6))
earthquake_types.plot(kind='bar', color='lightblue')
plt.title('Distribution of Earthquake Types')
plt.xlabel('Earthquake Type')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



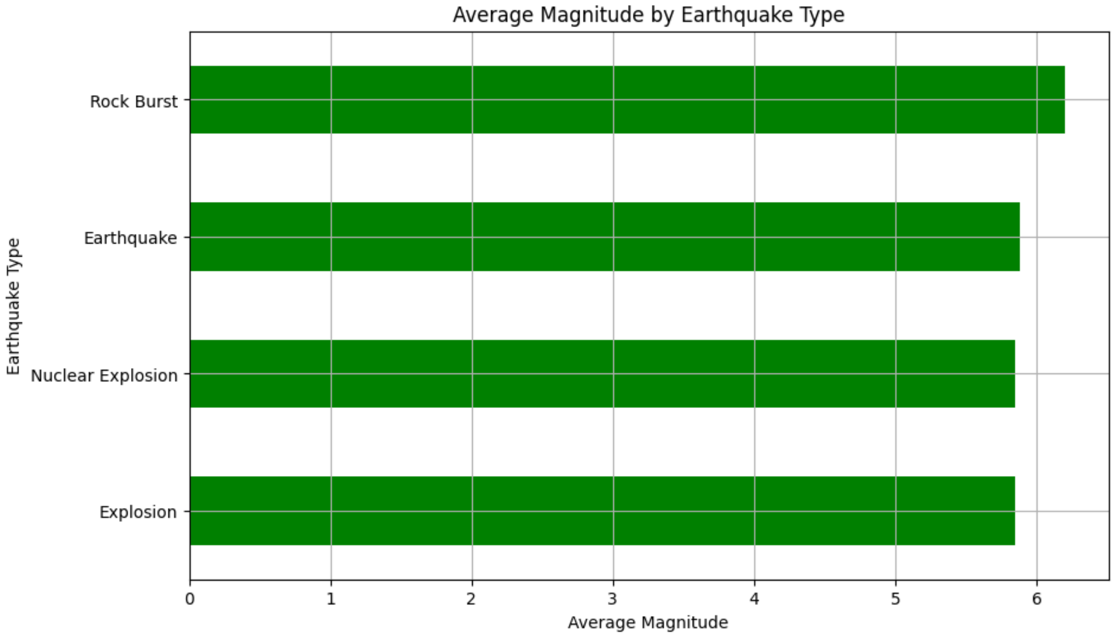
8. Grouping and Statistical Analysis Average Size by Type

[ ]:

```python
# Calculate the average magnitude for each earthquake type
avg_magnitude_by_type = df.groupby('Type')['Magnitude'].mean()
```

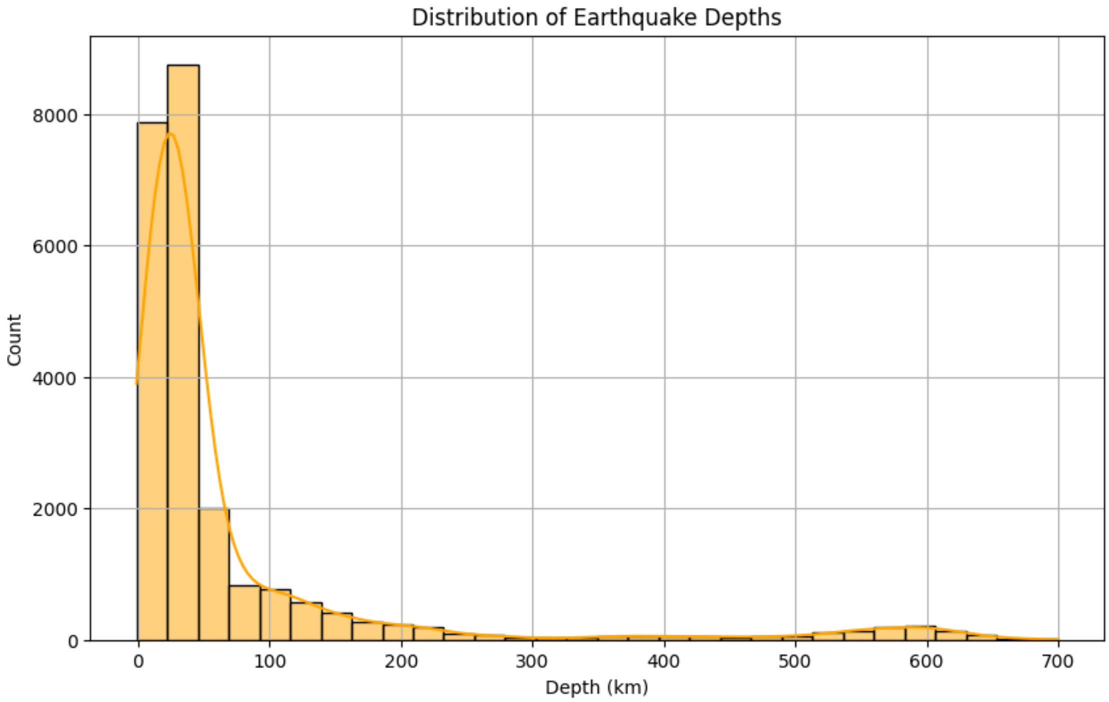
[ ]:

```python
# Plot the average magnitude by type
plt.figure(figsize=(10, 6))
avg_magnitude_by_type.sort_values().plot(kind='barh', color='green')
plt.title('Average Magnitude by Earthquake Type')
plt.xlabel('Average Magnitude')
plt.ylabel('Earthquake Type')
plt.grid(True)
plt.show()
```



9. Depth Analysis Depth Distribution

```
[ ]: # Plot the distribution of earthquake depths
     plt.figure(figsize=(10, 6))
     sns.histplot(df['Depth'], bins=30, kde=True, color='orange')
     plt.title('Distribution of Earthquake Depths')
     plt.xlabel('Depth (km)')
     plt.ylabel('Count')
     plt.grid(True)
     plt.show()
```

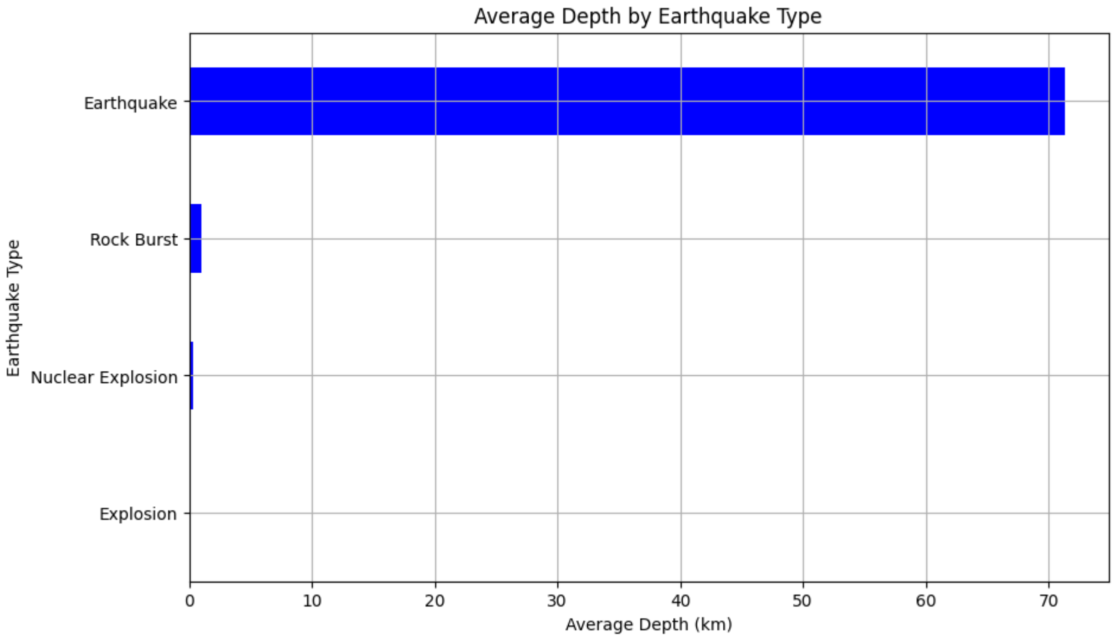


[ ]:

```
# Calculate the average depth for each earthquake type
avg_depth_by_type = df.groupby('Type')['Depth'].mean()
```

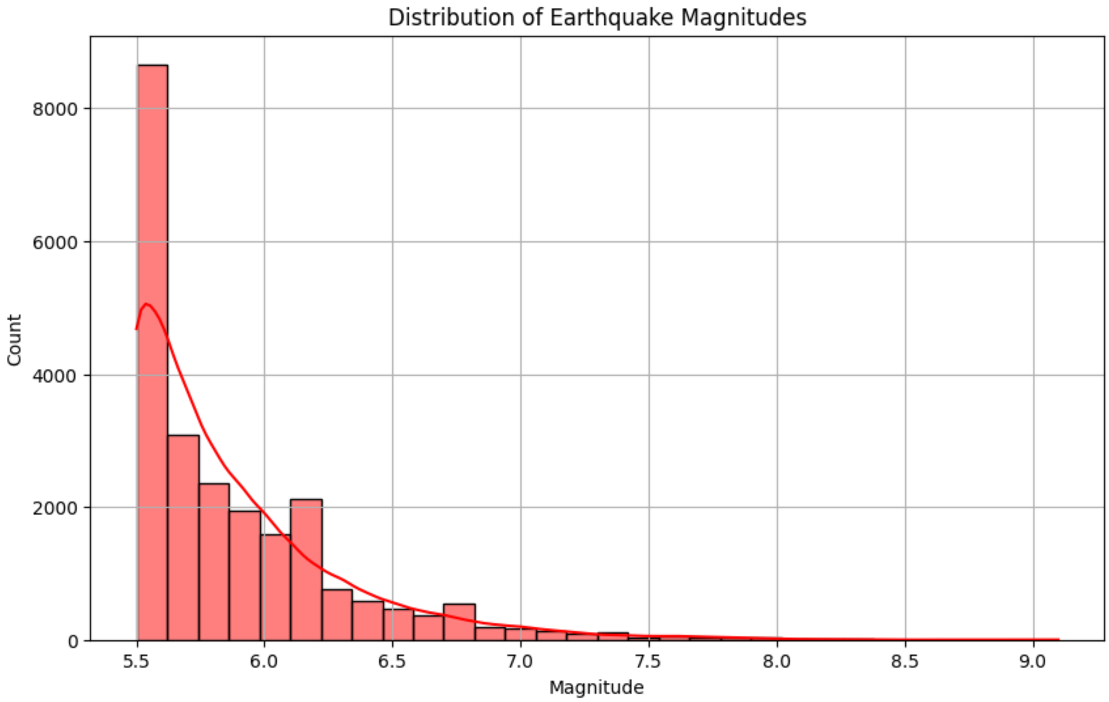
[ ]:

```
# Plot the average depth by type
plt.figure(figsize=(10, 6))
avg_depth_by_type.sort_values().plot(kind='barh', color='blue')
plt.title('Average Depth by Earthquake Type')
plt.xlabel('Average Depth (km)')
plt.ylabel('Earthquake Type')
plt.grid(True)
plt.show()
```



10. Size Analysis Size Distribution

```
[ ]: # Plot the distribution of earthquake magnitudes
     plt.figure(figsize=(10, 6))
     sns.histplot(df['Magnitude'], bins=30, kde=True, color='red')
     plt.title('Distribution of Earthquake Magnitudes')
     plt.xlabel('Magnitude')
     plt.ylabel('Count')
     plt.grid(True)
     plt.show()
```



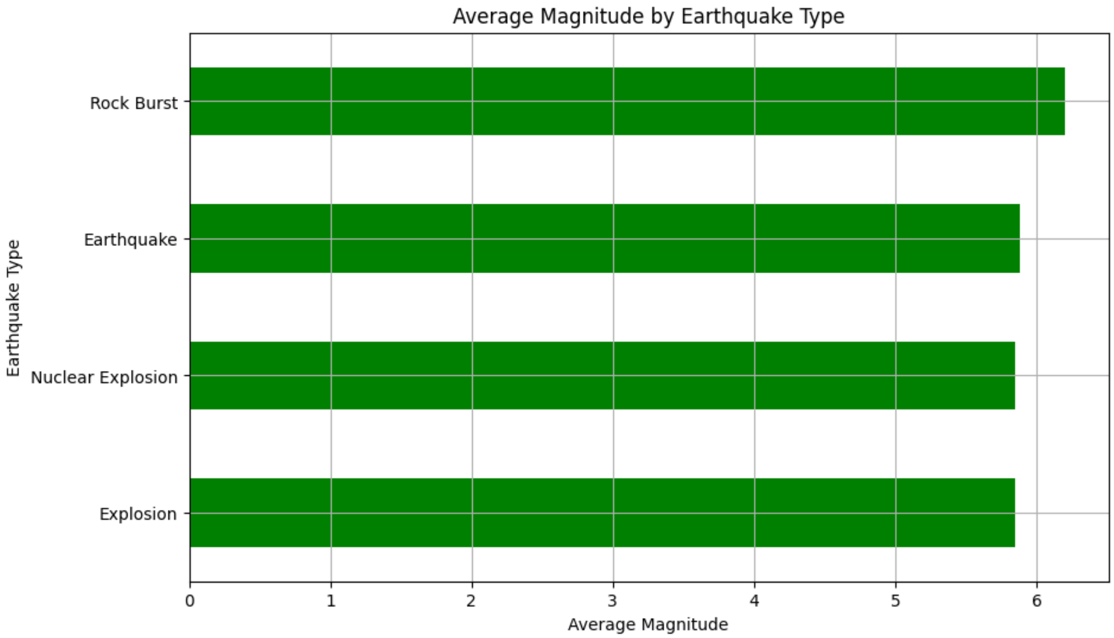
Average Size by Types

[ ]:

```
# Calculate the average magnitude for each earthquake type
avg_magnitude_by_type = df.groupby('Type')['Magnitude'].mean()
```

[ ]:

```
# Plot the average magnitude by type
plt.figure(figsize=(10, 6))
avg_magnitude_by_type.sort_values().plot(kind='barh', color='green')
plt.title('Average Magnitude by Earthquake Type')
plt.xlabel('Average Magnitude')
plt.ylabel('Earthquake Type')
plt.grid(True)
plt.show()
```

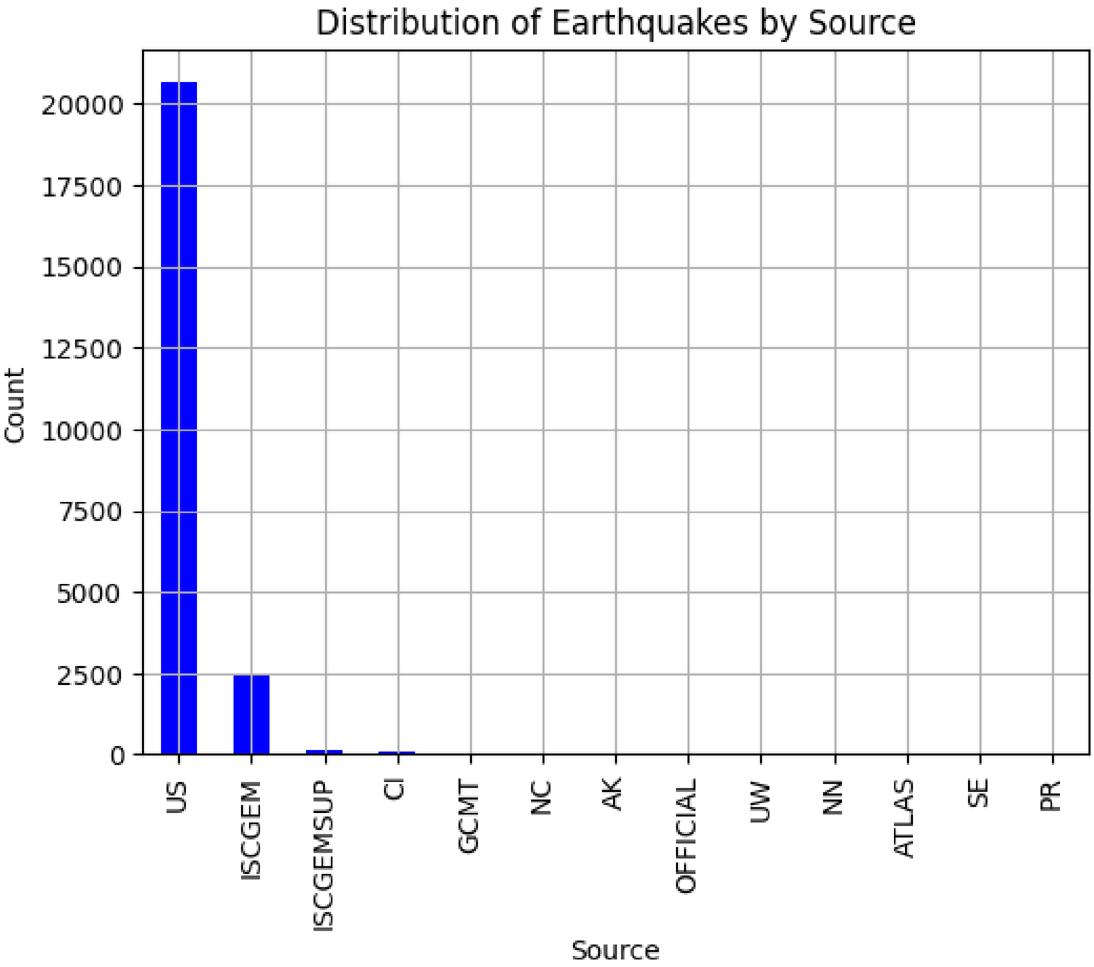


11. Source Analysis

By examining the source from which the data comes, you can determine reliable and unreliable sources.

[ ]:

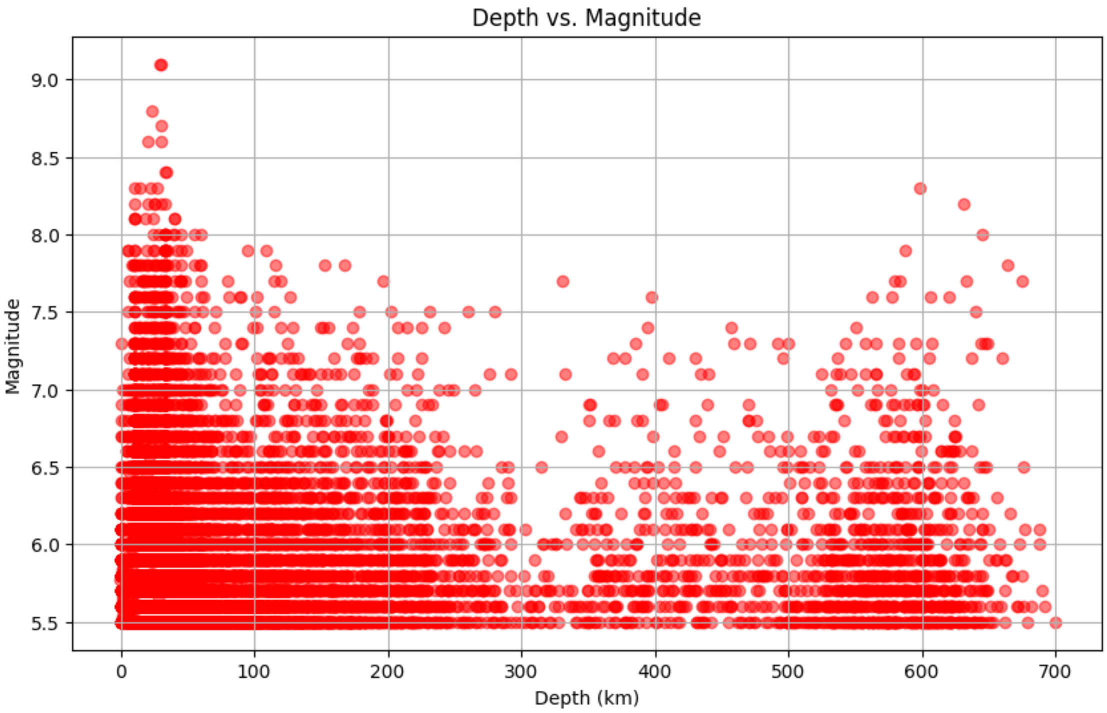
```
# Distribution of earthquakes according to sourcesplt.figure(figsize=(10, 6))
df['Source'].value_counts().plot(kind='bar', color='blue')
plt.title('Distribution of Earthquakes by Source')
plt.xlabel('Source')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```



12. Relationship between Depth and Size

You can draw a scatter plot to see the relationship between the earthquake's depth and magnitude.

```
[ ]: # Derinlik ve büyüklük arasındaki ilişkinin scatter plot'u
plt.figure(figsize=(10, 6))
plt.scatter(df['Depth'], df['Magnitude'], c='red', alpha=0.5)
plt.title('Depth vs. Magnitude')
plt.xlabel('Depth (km)')
plt.ylabel('Magnitude')
plt.grid(True)
plt.show()
```



Download and Import the Folium Library

First, download and import the folium library. If you haven't installed it before, you can download it with pip:

[ ]:
```
import folium
```

Create Map First, create a map. First, determine the map center and initial zoom level:

[ ]:
```
# Harita oluşturun
m = folium.Map(location=[0, 0], zoom_start=2)
```

Add Earthquakes to the Map

Add earthquake points to the map using your data. You can create a marker for each earthquake:

[ ]:
```
# Verileri döngü ile işleyin ve her bir depremi haritaya ekleyin
for index, row in df.iterrows():
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=f"Date: {row['Date']}<br>Depth: {row['Depth']}<br>Magnitude:␣
  ↪{row['Magnitude']}",
        icon=folium.Icon(color='blue')
    ).add_to(m)
```

This code will create a marker for each earthquake and add an information window (popup) with the date, depth and magnitude of the earthquake.

Show Map

Finally, you can use the m variable to represent the map you created:

[ ]:

```
m.save('earthquake_map.html')  # Haritayı bir HTML dosyasına kaydetmek␣
  ↪isterseniz
m
```

[ ]:

```
<folium.folium.Map at 0x22fb157a8d0>
```