Analysis of Natural Disaster Data

1. Introduction

Natural disasters cause significant loss of life, economic damage, and environmental destruction. This project analyzes a dataset of global natural disasters, focusing on their frequency, impact, and geographical distribution.

2. Objectives

- To visualize global disaster locations
- To analyze trends in disaster occurrences over time
- To assess the human and economic impact of disasters
- To explore correlations between different disaster-related factors

3. Methodology

3.1 Data Source

The dataset used is public_emdat_project.csv, which includes information on 15,784 disaster events, covering attributes like:

- Disaster type and subtype
- Location (country, region, coordinates)
- Start and end dates
- Number of deaths, injuries, and affected people
- Financial losses (total damage, insured damage)

3.2 Tools & Libraries

- **Python**: Data processing and visualization
- **Pandas**: Data manipulation
- **Geopandas & Matplotlib**: Geospatial visualization
- **Seaborn**: Statistical data visualization

4. Analysis and Results

4.1 Geospatial Analysis

- A scatter plot was generated to visualize disaster locations by type.
- The data showed high concentrations of disasters in Asia and North America.

4.2 Temporal Analysis

- The dataset was processed to extract yearly trends.
- A line graph depicted the increasing frequency of disasters over the years.

4.3 Impact Analysis

4.3.1 Human Impact

- Total deaths were aggregated by disaster type, with floods and earthquakes causing the highest fatalities.

4.3.2 Economic Impact

- The total damage costs by country were computed, revealing the top 10 most affected countries.
- Data processing encountered an error due to column formatting issues with Total Damage ('000 US$).

4.4 Correlation Analysis

- A heatmap displayed correlations between numerical attributes (e.g., deaths, affected individuals, damage).
- Strong correlations were found between the number of affected people and economic losses.

5. Challenges

- Missing or inconsistent data, especially in geolocation and financial attributes.
- Encoding issues required multiple attempts to load the dataset.
- Column name mismatches led to errors in financial impact analysis.

6. Conclusion

The analysis highlights the increasing frequency and impact of natural disasters, emphasizing the need for better preparedness and response strategies. Future work can focus on data cleaning and predictive modeling to forecast disaster trends.

```
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
try:
    df = pd.read_csv('public_emdat_project.csv', encoding='utf-8')
except UnicodeDecodeError:
    try:
        df = pd.read_csv('public_emdat_project.csv',
encoding='latin1')
    except UnicodeDecodeError:
        df = pd.read_csv('public_emdat_project.csv', encoding='ISO-
8859-1')

# Display basic info
print(df.info())
print(df.head())

# Geospatial Analysis: Plot disaster locations
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Longitude', y='Latitude', data=df, hue='Disaster
Type', palette='viridis', alpha=0.6)
plt.title('Global Disaster Locations by Type')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()

# Temporal Analysis: Disasters over time
df['Start Date'] = pd.to_datetime(df[['Start Year', 'Start Month',
'Start Day']].dropna().apply(
    lambda row: f"{int(row['Start Year'])}-{int(row['Start Month'])}-
{int(row['Start Day'])}", axis=1
))
disasters_over_time = df.groupby(df['Start Date'].dt.year)
['DisNo.'].count()
plt.figure(figsize=(10, 6))
disasters_over_time.plot(kind='line', marker='o')
plt.title('Number of Disasters Over Time')
plt.xlabel('Year')
plt.ylabel('Number of Disasters')
plt.grid()
plt.show()

# Impact Analysis: Total deaths by disaster type
total_deaths_by_type = df.groupby('Disaster Type')['Total
Deaths'].sum().sort_values(ascending=False)
plt.figure(figsize=(10, 6))
```

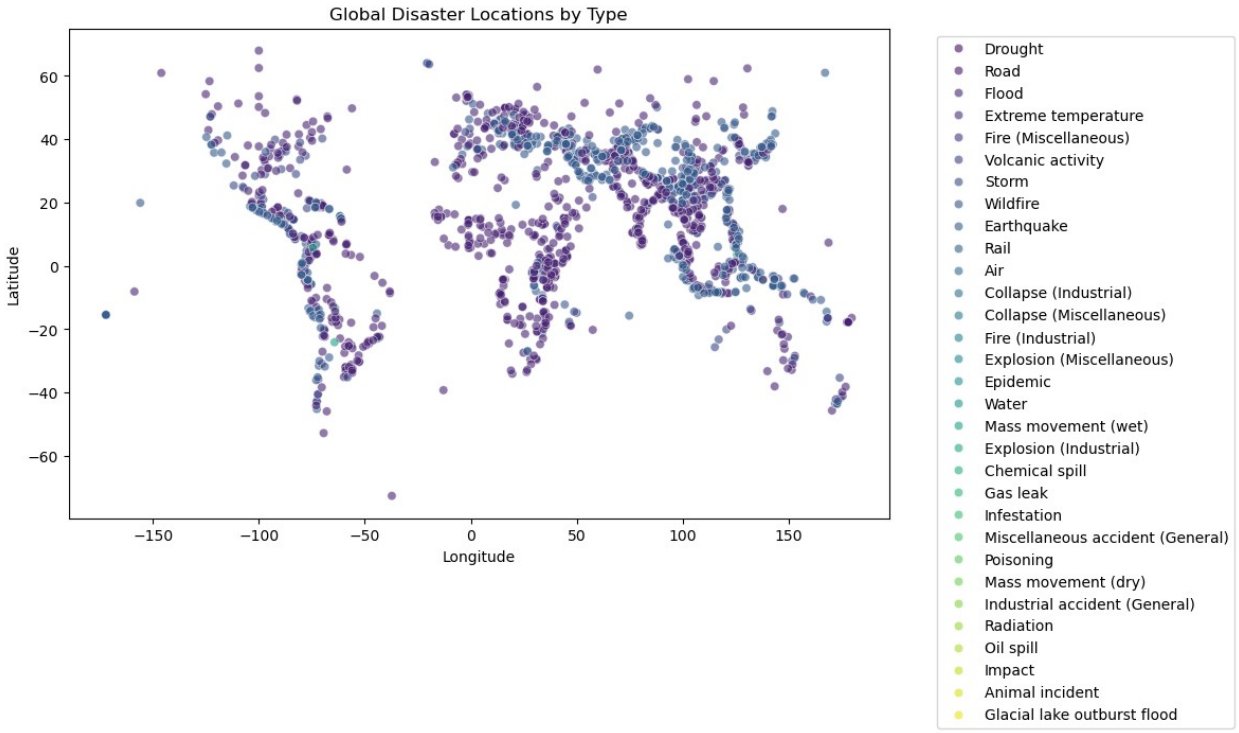
```
total_deaths_by_type.plot(kind='bar', color='skyblue')
plt.title('Total Deaths by Disaster Type')
plt.xlabel('Disaster Type')
plt.ylabel('Total Deaths')
plt.xticks(rotation=45, ha='right')
plt.show()

# Impact Analysis: Total damage by country
total_damage_by_country = df.groupby('Country')['Total Damage (`000
US$)'].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(10, 6))
total_damage_by_country.plot(kind='bar', color='orange')
plt.title('Top 10 Countries by Total Damage (in thousands of US$)')
plt.xlabel('Country')
plt.ylabel('Total Damage (in thousands of US$)')
plt.xticks(rotation=45, ha='right')
plt.show()

# Correlation Analysis: Heatmap of numerical variables
numerical_columns = ['Total Deaths', 'No. Injured', 'No. Affected',
'Total Damage (`000 US$)', 'Latitude', 'Longitude']
correlation_matrix = df[numerical_columns].corr()
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt='.2f')
plt.title('Correlation Heatmap of Numerical Variables')
plt.show()
```

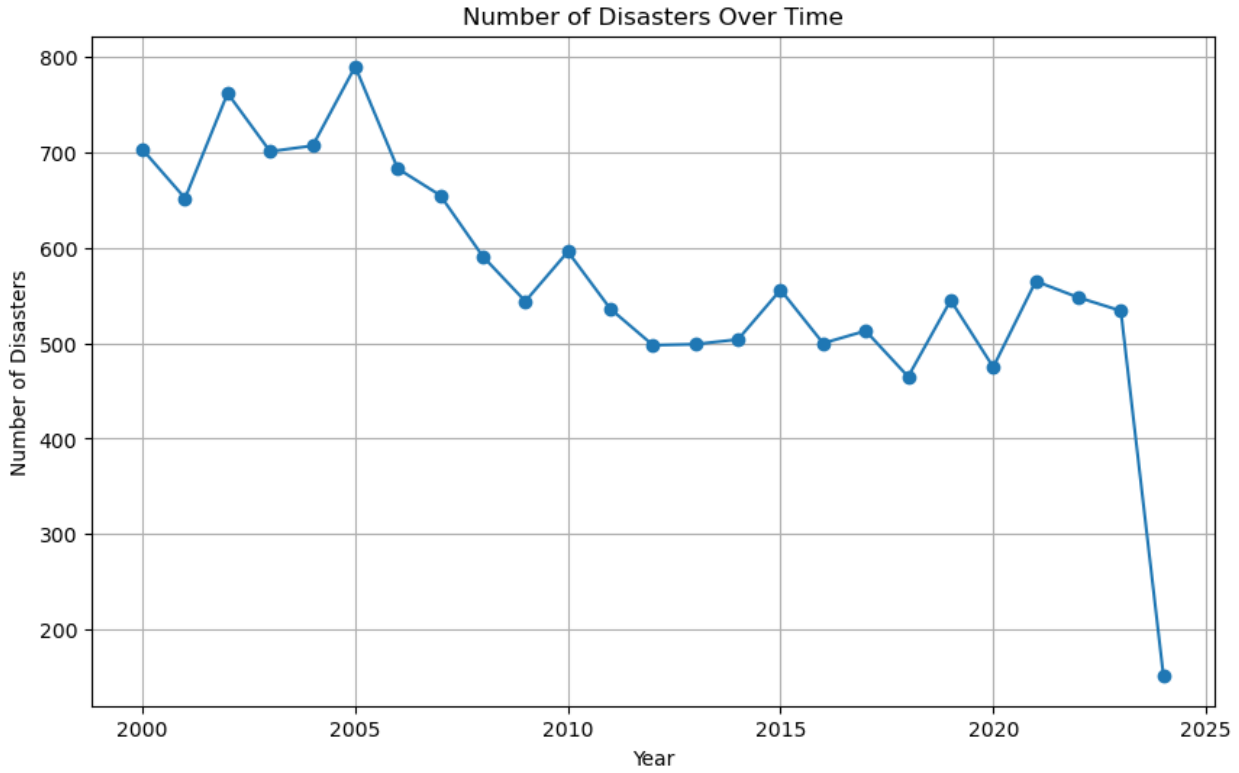
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15784 entries, 0 to 15783
Data columns (total 46 columns):
 #   Column                                  Non-Null Count  Dtype

---  ------                                  --------------  -----

 0   DisNo.                                  15784 non-null  object

 1   Historic                                15784 non-null  object

 2   Classification Key                      15784 non-null  object

 3   Disaster Group                          15784 non-null  object

 4   Disaster Subgroup                       15784 non-null  object

 5   Disaster Type                           15784 non-null  object

 6   Disaster Subtype                        15784 non-null  object

 7   External IDs                            2405 non-null   object
```

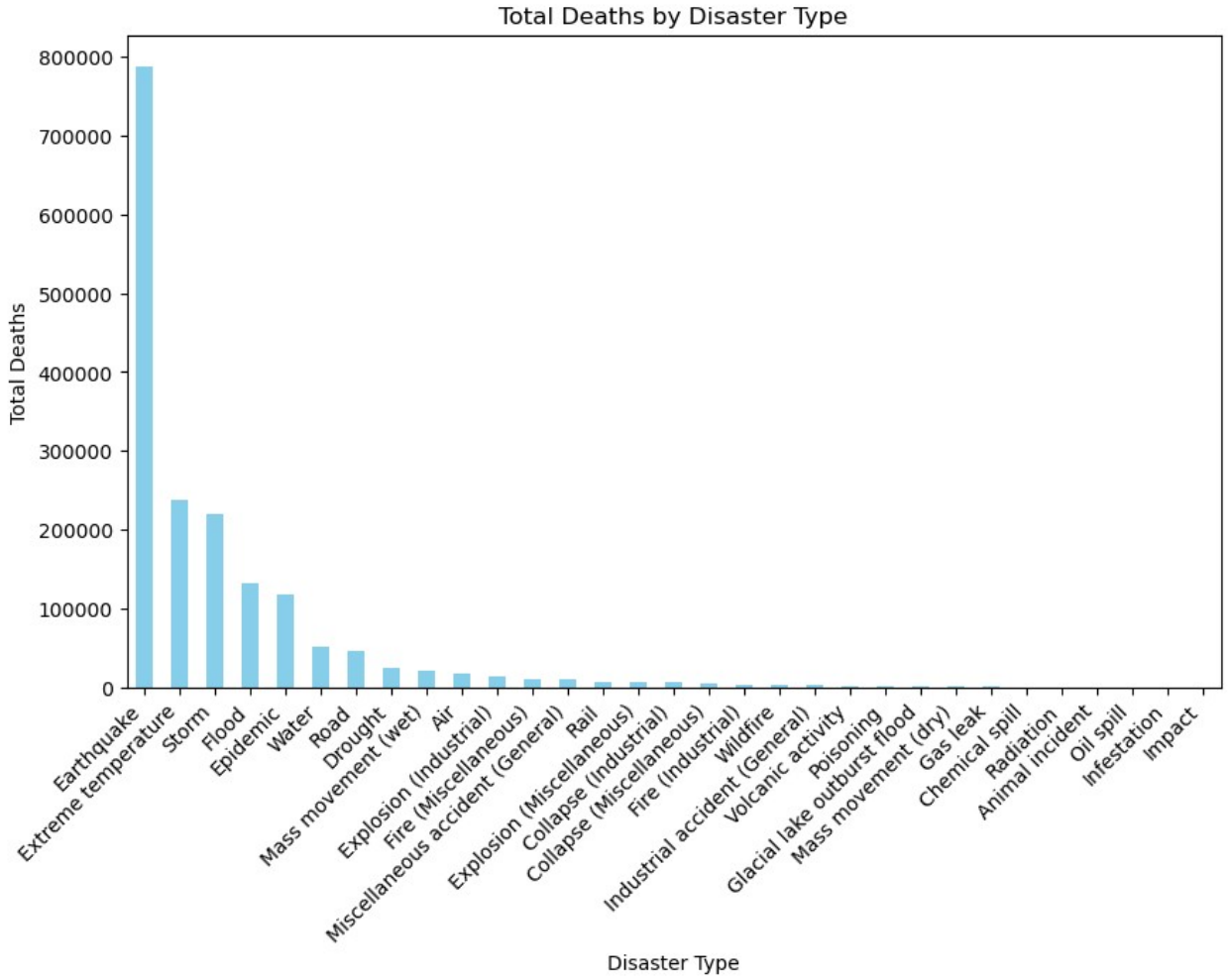
```
 8   Event Name                                  4955 non-null   object
 9   ISO                                         15784 non-null  object
 10  Country                                     15784 non-null  object
 11  Subregion                                   15784 non-null  object
 12  Region                                      15784 non-null  object
 13  Location                                    15136 non-null  object
 14  Origin                                      3955 non-null   object
 15  Associated Types                            3296 non-null   object
 16  OFDA/BHA Response                           15784 non-null  object
 17  Appeal                                      15784 non-null  object
 18  Declaration                                 15784 non-null  object
 19  AID Contribution ('000 US$)                 490 non-null
float64
 20  Magnitude                                   3378 non-null
float64
 21  Magnitude Scale                             9892 non-null   object
 22  Latitude                                    1815 non-null
float64
 23  Longitude                                   1815 non-null
float64
 24  River Basin                                 1212 non-null   object
 25  Start Year                                  15784 non-null  int64
 26  Start Month                                 15715 non-null
float64
 27  Start Day                                   14275 non-null
float64
 28  End Year                                    15784 non-null  int64
 29  End Month                                   15622 non-null
float64
 30  End Day                                     14342 non-null
float64
 31  Total Deaths                                12655 non-null
float64
 32  No. Injured                                 5790 non-null
float64
 33  No. Affected                                7172 non-null
```

```
float64
 34  No. Homeless                                1324 non-null
float64
 35  Total Affected                              11682 non-null
float64
 36  Reconstruction Costs ('000 US$)             33 non-null
float64
 37  Reconstruction Costs, Adjusted ('000 US$)  33 non-null
float64
 38  Insured Damage ('000 US$)                   695 non-null
float64
 39  Insured Damage, Adjusted ('000 US$)         694 non-null
float64
 40  Total Damage ('000 US$)                     3126 non-null
float64
 41  Total Damage, Adjusted ('000 US$)           3111 non-null
float64
 42  CPI                                         15621 non-null
float64
 43  Admin Units                                 8498 non-null   object

 44  Entry Date                                  15784 non-null  object

 45  Last Update                                 15784 non-null  object

dtypes: float64(20), int64(2), object(24)
memory usage: 5.5+ MB
None
          DisNo. Historic Classification Key Disaster Group Disaster
Subgroup  \
0  1999-9388-DJI       No    nat-cli-dro-dro        Natural
Climatological
1  1999-9388-SDN       No    nat-cli-dro-dro        Natural
Climatological
2  1999-9388-SOM       No    nat-cli-dro-dro        Natural
Climatological
3  2000-0001-AGO       No    tec-tra-roa-roa  Technological
Transport
4  2000-0002-AGO       No    nat-hyd-flo-riv        Natural
Hydrological

  Disaster Type Disaster Subtype External IDs Event Name  ISO  ...  \
0       Drought          Drought          NaN        NaN  DJI  ...
1       Drought          Drought          NaN        NaN  SDN  ...
2       Drought          Drought          NaN        NaN  SOM  ...
3          Road             Road          NaN        NaN  AGO  ...
4         Flood   Riverine flood          NaN        NaN  AGO  ...

  Reconstruction Costs ('000 US$) Reconstruction Costs, Adjusted ('000
US$)  \
```

```
0                                NaN
NaN
1                                NaN
NaN
2                                NaN
NaN
3                                NaN
NaN
4                                NaN
NaN

  Insured Damage ('000 US$) Insured Damage, Adjusted ('000 US$)  \
0                       NaN                                   NaN
1                       NaN                                   NaN
2                       NaN                                   NaN
3                       NaN                                   NaN
4                       NaN                                   NaN

  Total Damage ('000 US$) Total Damage, Adjusted ('000 US$)        CPI
\
0                     NaN                                  NaN  58.111474

1                     NaN                                  NaN  56.514291

2                     NaN                                  NaN  56.514291

3                     NaN                                  NaN  56.514291

4                 10000.0                              17695.0  56.514291


                                          Admin Units  Entry Date  Last
Update
0  [{"adm1_code":1093,"adm1_name":"Ali Sabieh"},{...  2006-03-01
2023-09-25
1  [{"adm1_code":2757,"adm1_name":"Northern Darfu...  2006-03-08
2023-09-25
2  [{"adm1_code":2691,"adm1_name":"Bay"},{"adm1_c...  2006-03-08
2023-09-25
3                                                NaN  2004-10-27
2023-09-25
4  [{"adm2_code":4214,"adm2_name":"Baia Farta"},{...  2005-02-03
2023-09-25

[5 rows x 46 columns]
```


Global Disaster Locations by Type
Latitude
Longitude
60
40
20
0
−20
−40
−60
−150
−100
−50
0
50
100
150
Drought
Road
Flood
Extreme temperature
Fire (Miscellaneous)
Volcanic activity
Storm
Wildfire
Earthquake
Rail
Air
Collapse (Industrial)
Collapse (Miscellaneous)
Fire (Industrial)
Explosion (Miscellaneous)
Epidemic
Water
Mass movement (wet)
Explosion (Industrial)
Chemical spill
Gas leak
Infestation
Miscellaneous accident (General)
Poisoning
Mass movement (dry)
Industrial accident (General)
Radiation
Oil spill
Impact
Animal incident
Glacial lake outburst flood


Number of Disasters Over Time
Number of Disasters
Year
800
700
600
500
400
300
200
2000
2005
2010
2015
2020
2025




```
---------------------------------------------------------------------------
-----
KeyError                                  Traceback (most recent call
last)
Cell In[15], line 52
     49 plt.show()
     51 # Impact Analysis: Total damage by country
---> 52 total_damage_by_country = df.groupby('Country')['Total Damage
(`000 US$)'].sum().sort_values(ascending=False).head(10)
     53 plt.figure(figsize=(10, 6))
     54 total_damage_by_country.plot(kind='bar', color='orange')

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
generic.py:1951, in DataFrameGroupBy.__getitem__(self, key)
   1944 if isinstance(key, tuple) and len(key) > 1:
   1945     # if len == 1, then it becomes a SeriesGroupBy and this is
actually
   1946     # valid syntax, so don't raise
   1947     raise ValueError(
```

```
   1948         "Cannot subset columns with a tuple with more than one
element. "
   1949         "Use a list instead."
   1950     )
-> 1951 return super().__getitem__(key)

File ~\anaconda3\Lib\site-packages\pandas\core\base.py:244, in
SelectionMixin.__getitem__(self, key)
    242 else:
    243     if key not in self.obj:
--> 244         raise KeyError(f"Column not found: {key}")
    245     ndim = self.obj[key].ndim
    246     return self._gotitem(key, ndim=ndim)

KeyError: 'Column not found: Total Damage (`000 US$)'
```