🌍 Disaster Impact and Response Analysis in Africa 2005 - 2019

Prepared by: IHEJIETO, ANITA CHISOM

Email: anitaihejieto@gmail.com

Date: 2025-05-31

📑 Executive Summary

This report analyzes epidemic and disease outbreak events across African countries using data from the EM-DAT international disaster database. The aim is to uncover trends and patterns in disease occurrence, duration, severity, Impact, geographic spread, and international responses.

✅ **Objectives:**

- Clean and explore EM-DAT disaster data.
- Integrate ISO country codes and UN sub-region classifications.
- Generate actionable visualizations and insights by;
  - Identify the most frequent epidemic subtypes, affected sub-regions and countries.
  - Measure the duration and human impact of disease outbreaks.
  - Measure Impact through metrics like total deaths, affected population, and disaster duration.
  - Evaluate the distribution of international support such as Appeals, Declarations, and OFDA/BHA responses.
  - Detect Outliers in durations and number of deaths tolls.
- Generate a reproducible report for stakeholders, policy makers, or NGOs.

Key findings:

- Epidemic outbreaks, especially bacterial and viral diseases, are frequent across many African countries.
- Western, Eastern and Midle Africa are among the most affected UN Sub-regions.
- Cholera and Meningococcal diseases are the most occurred disease across the African Countries
- 2005-2006 had the highest spike in Epidemic disasters and it depreciated over time
- Africa is amongst the less developed countries, so there was little or no Declarations, Appeals and International responses acrosss the UN Sub-Regions.

Data source:

Dataset: Emergency Event database (EMDAT)

Region of Focus: Africa

Timeframe: 2005 - 2019

📘 Introduction

Disasters continue to have a profound impact on African development and human well-being. In the past years, the world has experienced a series of epidemic and disease outbreaks, significantly impacting global health systems, economies, and communities. Understanding the nature and scope of these health emergencies is essential for improving disaster preparedness and response strategies. This project leverages a dataset containing detailed information on epidemic events globally, sourced from disaster databases like EM-DAT. The dataset encompasses records of disease outbreaks across multiple African countries and the United Nations Sub-regions, It also provides insight into various metrics like: The data spans multiple years and UN Sub-Regions, offering a robust base for comprehensive analysis of health-related disasters.

📂 Dataset Description

Key Columns:

- **Event Name:** The various diseases affecting the countries (e.g., Cholera, Ebola, COVID-19)
- **Start and End Dates:** The begining and the End dates for all events that occurred.
- **Disaster Type and Subtype:** The Disaster type is Epidemic while the Subtype are the disease range, that is; Viral, bacterial or Infectious diseases.
- **UN Sub-region and Development Region:** The Sub regions are Western, Eastern, Midlde, Southern and Northern and they are all less and least developed.
- **Total Deaths:** Specific number of people that died
- **Appeal/Declaration Status:** Was the disaster declared serious, were official Appeals made seeking for international aids?
- **OFDA/BHA Humanitarian Response:** Did the US respond and provide humanitarian aid/funding or not?
- **Duration of Event (in days):** How long did one event last for?

Importing necessary libraries and loading the excel file

In [1]:
```
import pandas as pd
import numpy as np

Anita_df = pd.read_excel ('public_emdat_custom_request_2025-05-10_cea27815-4dbd-4999-b051-f551b580e7bb.xl
```

Data cleaning and Transformation steps

Let's view the file and see what it looks like.

In [2]:
```
Anita_df
# We can also see how many rows and columns are in this dataset
```

Out[2]:

	DisNo.	Historic	Classification Key	Disaster Group	Disaster Subgroup	Disaster Type	Disaster Subtype	External IDs	Event Name	ISO	...	Reconstru Costs (
0	2005-0050-CMR	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	NaN	Cholera	CMR	...	
1	2005-0058-COD	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	NaN	Cholera	COD	...	
2	2005-0082-SDN	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	NaN	Meningococcal disease	SDN	...	
3	2005-0105-NGA	No	nat-bio-epi-dis	Natural	Biological	Epidemic	Infectious disease (General)	NaN	Acute watery diarrhoeal syndrome	NGA	...	
4	2005-0134-NGA	No	nat-bio-epi-vir	Natural	Biological	Epidemic	Viral disease	NaN	NaN	NGA	...	
...	...	...	...	...	...	...	...	...	...	...	...	
339	2019-0667-SDN	No	nat-bio-epi-dis	Natural	Biological	Epidemic	Infectious disease (General)	NaN	Visceral Leishmaniasis (Kala-Azar)	SDN	...	
340	2020-0039-NGA	No	nat-bio-epi-vir	Natural	Biological	Epidemic	Viral disease	GLIDE:EP-2020-000028	Lassa fever	NGA	...	
341	2020-0497-COD	No	nat-bio-epi-vir	Natural	Biological	Epidemic	Viral disease	NaN	Ebola	COD	...	
342	2020-0528-NGA	No	nat-bio-epi-vir	Natural	Biological	Epidemic	Viral disease	GLIDE:EP-2020-000230	Yellow fever	NGA	...	
343	2020-0607-SDN	No	nat-bio-epi-par	Natural	Biological	Epidemic	Parasitic disease	NaN	Visceral Leishmaniasis (Kala-Azar)	SDN	...	

344 rows × 46 columns

from the dataset above, some columns and rows contains null values and invalid data entries

In [3]:

```
#Now lets view and describe how many columns have Non-null datas (Non- null means not empty)
print(Anita_df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 46 columns):
 #   Column                                     Non-Null Count  Dtype  
---  ------                                     --------------  -----  
 0   DisNo.                                     344 non-null    object 
 1   Historic                                   344 non-null    object 
 2   Classification Key                         344 non-null    object 
 3   Disaster Group                             344 non-null    object 
 4   Disaster Subgroup                          344 non-null    object 
 5   Disaster Type                              344 non-null    object 
 6   Disaster Subtype                           344 non-null    object 
 7   External IDs                               128 non-null    object 
 8   Event Name                                 323 non-null    object 
 9   ISO                                        344 non-null    object 
 10  Country                                    344 non-null    object 
 11  Subregion                                  344 non-null    object 
 12  Region                                     344 non-null    object 
 13  Location                                   313 non-null    object 
 14  Origin                                     19 non-null     object 
 15  Associated Types                           0 non-null      float64
 16  OFDA/BHA Response                          344 non-null    object 
 17  Appeal                                     344 non-null    object 
 18  Declaration                                344 non-null    object 
 19  AID Contribution ('000 US$)                2 non-null      float64
 20  Magnitude                                  6 non-null      float64
 21  Magnitude Scale                            139 non-null    object 
 22  Latitude                                   0 non-null      float64
 23  Longitude                                  0 non-null      float64
 24  River Basin                                0 non-null      float64
 25  Start Year                                 344 non-null    int64  
 26  Start Month                                335 non-null    float64
 27  Start Day                                  199 non-null    float64
 28  End Year                                   344 non-null    int64  
 29  End Month                                  341 non-null    float64
 30  End Day                                    236 non-null    float64
 31  Total Deaths                               314 non-null    float64
 32  No. Injured                                89 non-null     float64
 33  No. Affected                               244 non-null    float64
 34  No. Homeless                               0 non-null      float64
 35  Total Affected                             329 non-null    float64
 36  Reconstruction Costs ('000 US$)            0 non-null      float64
 37  Reconstruction Costs, Adjusted ('000 US$)  0 non-null      float64
 38  Insured Damage ('000 US$)                  0 non-null      float64
 39  Insured Damage, Adjusted ('000 US$)        0 non-null      float64
 40  Total Damage ('000 US$)                    0 non-null      float64
 41  Total Damage, Adjusted ('000 US$)          0 non-null      float64
 42  CPI                                        344 non-null    float64
 43  Admin Units                                7 non-null      object 
 44  Entry Date                                 344 non-null    object 
 45  Last Update                                344 non-null    object 
dtypes: float64(22), int64(2), object(22)
memory usage: 123.8+ KB
None
```

This means we have some columns that are empty, 0 non-null means there's no column that is not empty, 344 non-null means 344 are not empty. So they have to be dropped to avoid biased data.

Steps to be taken:

- Removed duplicates
- Removing extra spaces between column strings
- Drop columns with >50% missing data.
- Droping multiple columns that couldn't be dropped because their headers included special characters within like , '' 0987,$
- Creat disaster_duration = End Date - Start Date.
- Extract year from Start Date

- Merge ISO dataset with this current data

In [4]:
```
#Removing duplicates
Anita_df = Anita_df.drop_duplicates()

#Removing extra spaces between strings in columns
Anita_df.columns = Anita_df.columns.str.strip()

#dropping columns that are more than 50% empty
Anita_df= Anita_df.drop(columns = ['External IDs','Origin','Associated Types','Magnitude','Magnitude Sca
```

In [5]:
```
print(Anita_df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 344 entries, 0 to 343
Data columns (total 35 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   DisNo.                                    344 non-null    object 
 1   Historic                                  344 non-null    object 
 2   Classification Key                        344 non-null    object 
 3   Disaster Group                            344 non-null    object 
 4   Disaster Subgroup                         344 non-null    object 
 5   Disaster Type                             344 non-null    object 
 6   Disaster Subtype                          344 non-null    object 
 7   Event Name                                323 non-null    object 
 8   ISO                                       344 non-null    object 
 9   Country                                   344 non-null    object 
 10  Subregion                                 344 non-null    object 
 11  Region                                    344 non-null    object 
 12  Location                                  313 non-null    object 
 13  OFDA/BHA Response                         344 non-null    object 
 14  Appeal                                    344 non-null    object 
 15  Declaration                               344 non-null    object 
 16  AID Contribution ('000 US$)               2 non-null      float64
 17  Start Year                                344 non-null    int64  
 18  Start Month                               335 non-null    float64
 19  Start Day                                 199 non-null    float64
 20  End Year                                  344 non-null    int64  
 21  End Month                                 341 non-null    float64
 22  End Day                                   236 non-null    float64
 23  Total Deaths                              314 non-null    float64
 24  No. Affected                              244 non-null    float64
 25  Total Affected                            329 non-null    float64
 26  Reconstruction Costs ('000 US$)           0 non-null      float64
 27  Reconstruction Costs, Adjusted ('000 US$) 0 non-null      float64
 28  Insured Damage ('000 US$)                 0 non-null      float64
 29  Insured Damage, Adjusted ('000 US$)       0 non-null      float64
 30  Total Damage ('000 US$)                   0 non-null      float64
 31  Total Damage, Adjusted ('000 US$)         0 non-null      float64
 32  CPI                                       344 non-null    float64
 33  Entry Date                                344 non-null    object 
 34  Last Update                               344 non-null    object 
dtypes: float64(15), int64(2), object(18)
memory usage: 96.8+ KB
None
```

35 columns now left out of 44 that was Initially loaded.

In [6]:
```
#Droping multiple columns that couldn't be dropped because their headers included special characters with
Anita_df = Anita_df.drop(Anita_df.columns[[16,26,27,28,29,30,31]], axis = 1)
```

In [7]:
```
print(Anita_df)
```

```
          DisNo. Historic Classification Key Disaster Group  \
0    2005-0050-CMR       No    nat-bio-epi-bac        Natural   
1    2005-0058-COD       No    nat-bio-epi-bac        Natural   
2    2005-0082-SDN       No    nat-bio-epi-bac        Natural   
3    2005-0105-NGA       No    nat-bio-epi-dis        Natural   
4    2005-0134-NGA       No    nat-bio-epi-vir        Natural   
..             ...      ...                ...            ...   
339  2019-0667-SDN       No    nat-bio-epi-dis        Natural   
340  2020-0039-NGA       No    nat-bio-epi-vir        Natural   
341  2020-0497-COD       No    nat-bio-epi-vir        Natural   
342  2020-0528-NGA       No    nat-bio-epi-vir        Natural   
343  2020-0607-SDN       No    nat-bio-epi-par        Natural   

    Disaster Subgroup Disaster Type             Disaster Subtype  \
0          Biological      Epidemic            Bacterial disease   
1          Biological      Epidemic            Bacterial disease   
2          Biological      Epidemic            Bacterial disease   
3          Biological      Epidemic  Infectious disease (General)   
4          Biological      Epidemic                Viral disease   
..                ...           ...                          ...   
339        Biological      Epidemic  Infectious disease (General)   
340        Biological      Epidemic                Viral disease   
341        Biological      Epidemic                Viral disease   
342        Biological      Epidemic                Viral disease   
343        Biological      Epidemic            Parasitic disease   

                             Event Name  ISO  \
0                               Cholera  CMR   
1                               Cholera  COD   
2                 Meningococcal disease  SDN   
3      Acute watery diarrhoeal syndrome  NGA   
4                                   NaN  NGA   
..                                  ...  ...   
339  Visceral Leishmaniasis (Kala-Azar)  SDN   
340                         Lassa fever  NGA   
341                               Ebola  COD   
342                        Yellow fever  NGA   
343  Visceral Leishmaniasis (Kala-Azar)  SDN   

                              Country  ... Start Day End Year End Month  \
0                            Cameroon  ...       3.0     2005       5.0   
1    Democratic Republic of the Congo  ...      17.0     2005       1.0   
2                               Sudan  ...      22.0     2005       2.0   
3                             Nigeria  ...       3.0     2005       3.0   
4                             Nigeria  ...      28.0     2005       3.0   
..                                ...  ...       ...      ...       ...   
339                             Sudan  ...       NaN     2019      12.0   
340                           Nigeria  ...       1.0     2020       2.0   
341  Democratic Republic of the Congo  ...       1.0     2020      11.0   
342                           Nigeria  ...       1.0     2020      12.0   
343                             Sudan  ...       NaN     2020      12.0   

    End Day Total Deaths No. Affected  Total Affected        CPI  Entry Date  \
0       NaN         42.0       1400.0          1400.0  64.092740  2005-05-11   
1      30.0         34.0       2152.0          2152.0  64.092740  2006-01-19   
2       2.0        124.0       3579.0          3579.0  64.092740  2005-05-11   
3      16.0         46.0        200.0           200.0  64.092740  2005-03-23   
4      23.0        561.0      23575.0         23575.0  64.092740  2005-04-21   
..      ...          ...          ...             ...        ...         ...   
339     NaN         29.0          NaN          2098.0  83.904170  2022-05-19   
340    17.0         47.0          NaN           365.0  84.939198  2020-01-29   
341    11.0         55.0          NaN             NaN  84.939198  2020-11-24   
342     4.0         76.0          NaN           222.0  84.939198  2020-12-04   
343     NaN         38.0          NaN          2137.0  84.939198  2022-05-19   

     Last Update  
0     2023-09-25  
1     2023-09-25  
2     2023-09-25  
3     2023-09-25  
4     2023-09-25  
```

```
..          ...
339   2023-09-25
340   2023-09-25
341   2023-09-25
342   2023-09-25
343   2023-09-25

[344 rows x 28 columns]
```

The columns have been removed and we're left with 28, but we stil have null rows

In [8]:
```
#Droping all empty rows
Anita_df = Anita_df.dropna()
```

In [9]:
```
#Let's view the data info
Anita_df.isnull().sum()
```

Out[9]:
```
DisNo.                0
Historic              0
Classification Key    0
Disaster Group        0
Disaster Subgroup     0
Disaster Type         0
Disaster Subtype      0
Event Name            0
ISO                   0
Country               0
Subregion             0
Region                0
Location              0
OFDA/BHA Response     0
Appeal                0
Declaration           0
Start Year            0
Start Month           0
Start Day             0
End Year              0
End Month             0
End Day               0
Total Deaths          0
No. Affected          0
Total Affected        0
CPI                   0
Entry Date            0
Last Update           0
dtype: int64
```

This means we have 0 null values accross rows and columns. Our dataset is now clean and non biased

In [10]:
```
#Changing the data type columns with dates into int type and renaming the column haeaders to a format und
#then changing the format into a datetime
Anita_df['Start Day'] = Anita_df['Start Day'].astype(int)

Anita_df['Start Month'] = Anita_df['Start Month'].astype(int)

Anita_df['Start Year'] = Anita_df['Start Year'].astype(int)

Anita_df['Start Date'] = pd.to_datetime(Anita_df.rename(columns={'Start Day' :'day',
                                                                  'Start Month':'month',
                                                                  'Start Year':'year'
                                                                  })[['year','month','day']])



Anita_df['End Day'] = Anita_df['End Day'].astype(int)

Anita_df['End Month'] = Anita_df['End Month'].astype(int)

Anita_df['End Year'] = Anita_df['End Year'].astype(int)

Anita_df['End Date'] = pd.to_datetime(Anita_df.rename(columns={'End Day' :'day',
```

```
                                             'End Month':'month',
                                             'End Year':'year'
                                             })[['year','month','day']])
```

In [11]:
```
#Calculate the duration of Epidemic disaster
Anita_df['Disaster duration'] = (Anita_df['End Date'] - Anita_df['Start Date']).dt.days
```

In [12]:
```
Anita_df
```

Out[12]:

	DisNo.	Historic	Classification Key	Disaster Group	Disaster Subgroup	Disaster Type	Disaster Subtype	Event Name	ISO	Country	...	End Day	T De
1	2005-0058-COD	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	COD	Democratic Republic of the Congo	...	30	
2	2005-0082-SDN	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Meningococcal disease	SDN	Sudan	...	2	1
3	2005-0105-NGA	No	nat-bio-epi-dis	Natural	Biological	Epidemic	Infectious disease (General)	Acute watery diarrhoeal syndrome	NGA	Nigeria	...	16	
7	2005-0176-SEN	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	SEN	Senegal	...	23	3
8	2005-0186-COD	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	COD	Democratic Republic of the Congo	...	8	
...	...	...	...	...	...	...	...	...	...	...	...	...	
303	2017-0549-MWI	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	MWI	Malawi	...	11	
309	2018-0076-UGA	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	UGA	Uganda	...	28	
317	2018-0436-AGO	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	AGO	Angola	...	12	
330	2019-0352-ETH	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	ETH	Ethiopia	...	23	
333	2019-0525-SDN	No	nat-bio-epi-bac	Natural	Biological	Epidemic	Bacterial disease	Cholera	SDN	Sudan	...	5	

119 rows × 31 columns

We have 3 new columns added at the end of the table and a total of 119 rows now.

Next is merging the ISO UN data and extracting the 3 columns needed for analysis which are ISO Code, UN Sub regions and development regions

In [13]:
```
#Merge the ISO excel data with current dataset, but first we have to load the dataset
ISO_df = pd.read_excel("JME_Regional-Classifications.xlsx")
```

In [14]:
```
ISO_df
```

Out[14]:

	ISO Code	Country	UN Region	UN Sub Region	SDG Region	Development Regions	UNICEF Region	UNICEF Sub-Regions	WHO Region	World Bank Income Groups	World Bank Income Groups Combined	R
0	AFG	Afghanistan	Asia	Southern Asia	Central Asia and Southern Asia	Least Developed	ROSA	SA	EMRO	Low Income	Low Income	
1	ALB	Albania	Europe	Southern Europe	Northern America and Europe	More Developed	ECA	EECA	EURO	Upper Middle Income	Middle Income	I (
2	DZA	Algeria	Africa	Northern Africa	Western Asia and Northern Africa	Less Developed	MENA	MENA	AFRO	Upper Middle Income	Middle Income	I
3	AND	Andorra	Europe	Southern Europe	Northern America and Europe	More Developed	ECA	WE	EURO	High Income	High Income	I (
4	AGO	Angola	Africa	Middle Africa	Sub-Saharan Africa	Least Developed	SSA	ESA	AFRO	Lower Middle Income	Middle Income	S
...	...	...	...	...	...	...	...	...	...	...	...	
201	ZWE	Zimbabwe	Africa	Eastern Africa	Sub-Saharan Africa	Less Developed	SSA	ESA	AFRO	Low Income	Low Income	S
202	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
203	1 UNICEF regional abbreviations and full names...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
204	2 WHO regional abbreviations and full names: A...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
205	3 Based on FY18 World Bank income classification	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

206 rows × 12 columns

In [15]:

```
#Changing our 'ISO Code' column in the new ISO Dataset to 'ISO' so it matches with the ISO column name ir
ISO_df = ISO_df.rename(columns = {'ISO Code':'ISO'})
```

In [16]:

```
#Selecting only the needed columns from the ISO dataset
ISO_subset= ISO_df[['UN Sub Region', 'Development Regions', 'ISO']]

#Merge with Anita_df using ISO as the key
Anita_df = pd.merge (Anita_df, ISO_subset, how='left', on='ISO')
```

In [17]:

```
print(Anita_df)
```

```
            DisNo. Historic Classification Key Disaster Group  \
0    2005-0058-COD       No    nat-bio-epi-bac        Natural
1    2005-0082-SDN       No    nat-bio-epi-bac        Natural
2    2005-0105-NGA       No    nat-bio-epi-dis        Natural
3    2005-0176-SEN       No    nat-bio-epi-bac        Natural
4    2005-0186-COD       No    nat-bio-epi-bac        Natural
..             ...      ...                ...            ...
114  2017-0549-MWI       No    nat-bio-epi-bac        Natural
115  2018-0076-UGA       No    nat-bio-epi-bac        Natural
116  2018-0436-AGO       No    nat-bio-epi-bac        Natural
117  2019-0352-ETH       No    nat-bio-epi-bac        Natural
118  2019-0525-SDN       No    nat-bio-epi-bac        Natural

    Disaster Subgroup Disaster Type              Disaster Subtype  \
0          Biological      Epidemic             Bacterial disease
1          Biological      Epidemic             Bacterial disease
2          Biological      Epidemic  Infectious disease (General)
3          Biological      Epidemic             Bacterial disease
4          Biological      Epidemic             Bacterial disease
..                ...           ...                           ...
114        Biological      Epidemic             Bacterial disease
115        Biological      Epidemic             Bacterial disease
116        Biological      Epidemic             Bacterial disease
117        Biological      Epidemic             Bacterial disease
118        Biological      Epidemic             Bacterial disease

                         Event Name  ISO                           Country  \
0                           Cholera  COD  Democratic Republic of the Congo
1              Meningococcal disease  SDN                             Sudan
2    Acute watery diarrhoeal syndrome  NGA                           Nigeria
3                           Cholera  SEN                           Senegal
4                           Cholera  COD  Democratic Republic of the Congo
..                              ...  ...                               ...
114                         Cholera  MWI                            Malawi
115                         Cholera  UGA                            Uganda
116                         Cholera  AGO                            Angola
117                         Cholera  ETH                          Ethiopia
118                         Cholera  SDN                             Sudan

     ... No. Affected Total Affected        CPI  Entry Date Last Update  \
0    ...       2152.0         2152.0  64.092740  2006-01-19  2023-09-25
1    ...       3579.0         3579.0  64.092740  2005-05-11  2023-09-25
2    ...        200.0          200.0  64.092740  2005-03-23  2023-09-25
3    ...      23022.0        23022.0  64.092740  2005-10-24  2023-09-25
4    ...       1420.0         1420.0  64.092740  2005-06-02  2023-09-25
..   ...          ...            ...        ...         ...         ...
114  ...        450.0          450.0  80.445779  2018-03-02  2023-09-25
115  ...       1000.0         1000.0  82.410668  2018-03-01  2023-09-25
116  ...        139.0          139.0  82.410668  2018-11-28  2023-09-25
117  ...        871.0          871.0  83.904170  2019-07-22  2023-09-25
118  ...        510.0          510.0  83.904170  2019-10-29  2023-09-25

    Start Date   End Date  Disaster duration    UN Sub Region  \
0   2005-01-17 2005-01-30                 13    Middle Africa
1   2005-01-22 2005-02-02                 11  Northern Africa
2   2005-02-03 2005-03-16                 41   Western Africa
3   2005-01-01 2005-09-23                265   Western Africa
4   2005-01-01 2005-05-08                127    Middle Africa
..         ...        ...                ...              ...
114 2017-11-24 2018-02-11                 79   Eastern Africa
115 2018-02-15 2018-02-28                 13   Eastern Africa
116 2018-10-09 2018-11-12                 34    Middle Africa
117 2019-04-25 2019-06-23                 59   Eastern Africa
118 2019-08-29 2019-12-05                 98  Northern Africa

     Development Regions
0        Least Developed
1        Least Developed
2         Less Developed
3        Least Developed
4        Least Developed
```

```
..                  ...
114       Least Developed
115       Least Developed
116       Least Developed
117       Least Developed
118       Least Developed

[119 rows x 33 columns]
```

The development regions and UN Sub-Regions column has been added

📊 Exploratory Data Analysis

We explore distributions, frequencies and impacts by sub-region, death and Disaster subtypes, severity over time.

In [18]:

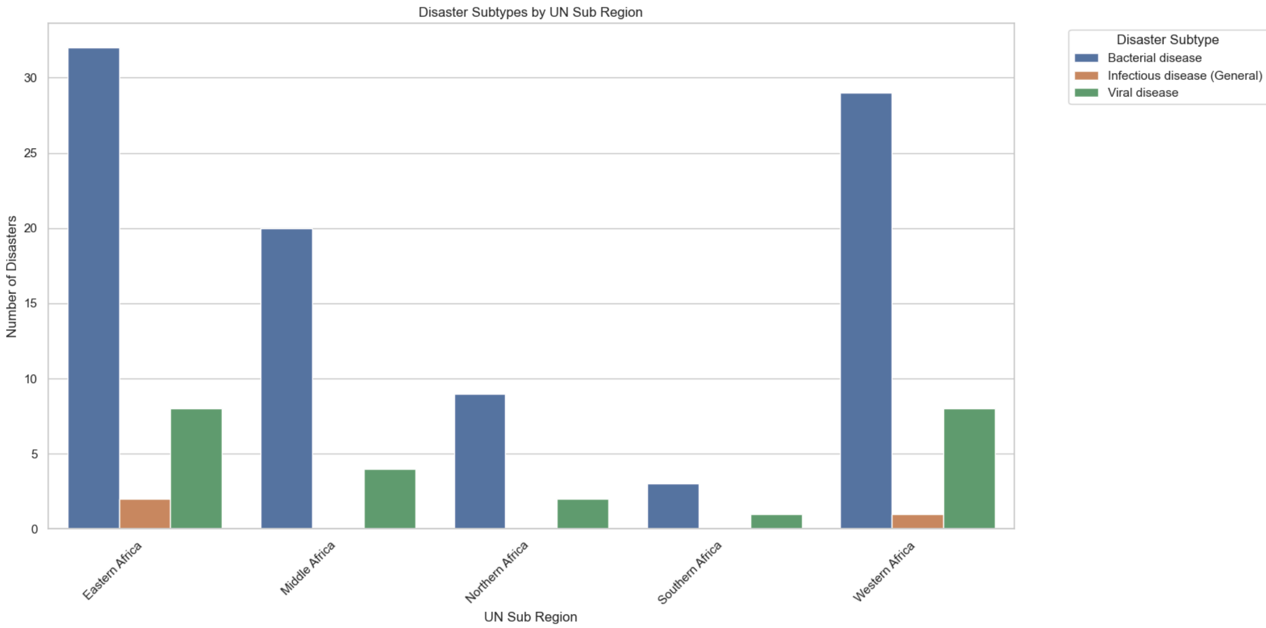
```
# Importing the necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

1. Disaster Subtypes by UN Sub-region

Using **Bar plot:** A bar plot is an effective and intuitive visualization tool for comparing categorical data across multiple groups. In the context of Disaster Subtypes by UN Sub-region, the objective is to understand the frequency or distribution of different disaster subtypes within each regional grouping. So using barplot makes the visualization Scalable, Readable and Efective for easy understanding.

In [19]:

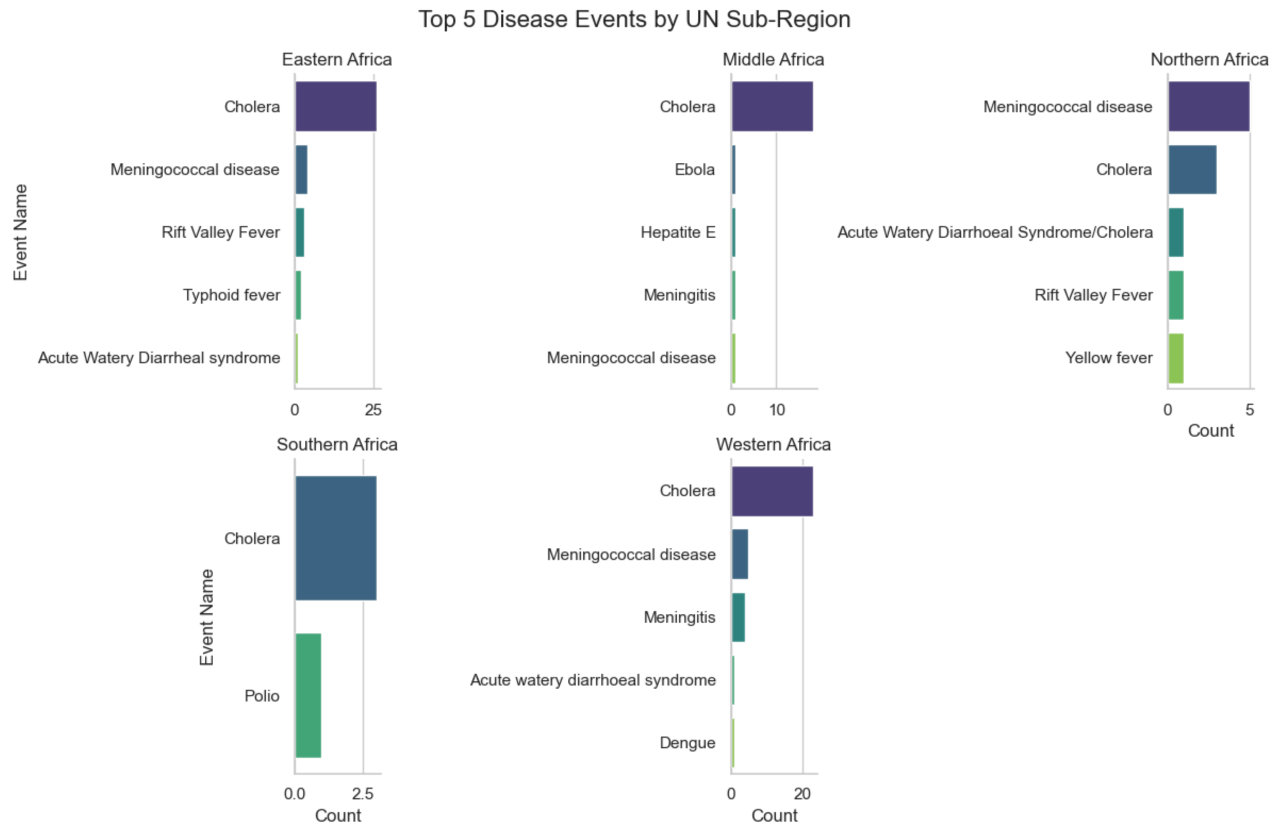
```
# Group the data
Plot_1 = Anita_df.groupby(['UN Sub Region', 'Disaster Subtype']).size().reset_index(name='Count')

# Set up the plot
plt.figure(figsize=(16, 8))
sns.set(style="whitegrid")

# Create the barplot
sns.barplot(data=Plot_1, x='UN Sub Region', y='Count', hue='Disaster Subtype')

# Rotate x-axis labels for better readability
plt.xticks(rotation=45, ha='right')
plt.title('Disaster Subtypes by UN Sub Region')
plt.xlabel('UN Sub Region')
plt.ylabel('Number of Disasters')
plt.legend(title='Disaster Subtype', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```



2. Top 5 Disease Events by UN Sub Region

The Top 5 Disease Events by UN Sub-region represent a combination of categorical variables (disease names and regions) and their frequencies or counts. A bar plot is the most effective tool for visualizing this type of summary data due to its'ability of Easy Comparison of Top Disease Events, Highlighting Rankings Effectively and Supports Categorical Interpretation.

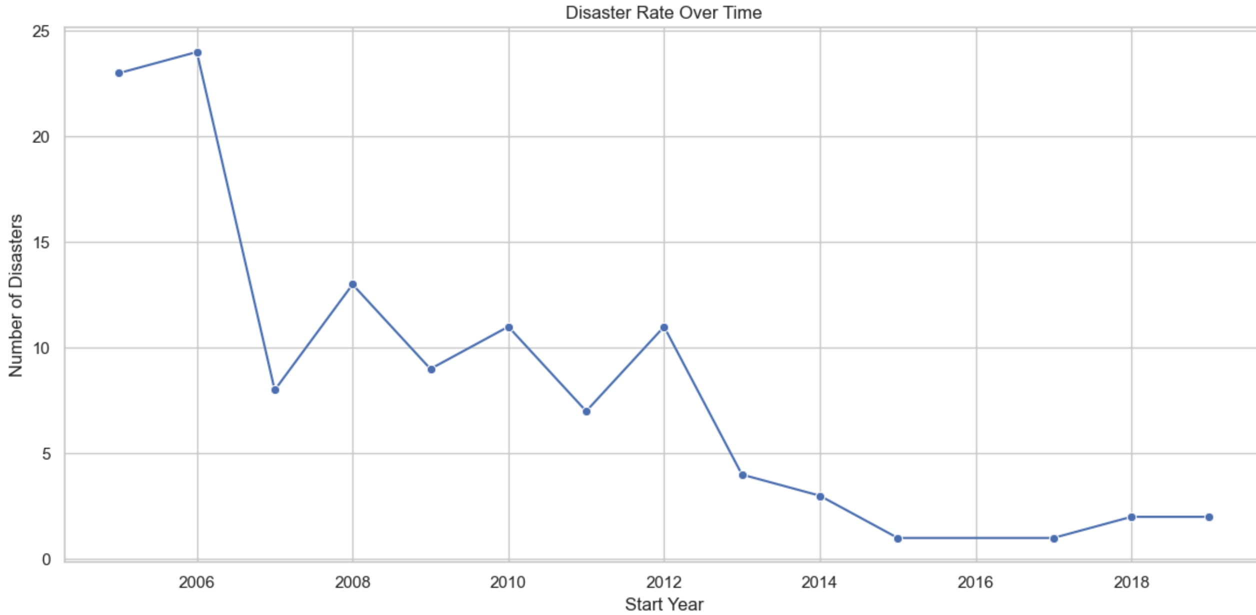
In [20]:
```
# Step 1: Get top 5 Disaster event names per UN Sub-region
Top5 = (
    Anita_df.groupby('UN Sub Region')['Event Name']
    .value_counts()
    .groupby(level=0)
    .head(5)
    .reset_index(name='Count')
)

# Step 2: Plot with seaborn using FacetGrid
g = sns.FacetGrid(Top5, col='UN Sub Region', col_wrap=3, sharex=False, sharey=False, height=4)
g.map_dataframe(sns.barplot, y='Event Name', x='Count', palette='viridis')

# Step 3: Adjust layout and titles
g.set_titles('{col_name}')
g.set_axis_labels('Count', 'Event Name')
g.fig.subplots_adjust(top=0.9)
g.fig.suptitle('Top 5 Disease Events by UN Sub-Region', fontsize=16)
plt.tight_layout()
plt.show()
```



3. Disaster Rate Over Time

When analyzing how disaster events change across years or months, we're working with time series data. A line plot is the most appropriate and effective way to visualize these trends and fluctuations over time due to the ability to track Trends Across Time, Highlight Patterns, Spikes, dips, and consistent growth or decline in disaster frequency and Supports Multi-Series Comparison

In [21]:
```
# Group the dataset by year and count the number of Disasters
disaster_trends = Anita_df['Start Year'].value_counts().sort_index()

# Plot the result
plt.figure(figsize=(12, 6))
sns.lineplot(x=disaster_trends.index, y=disaster_trends.values, marker='o')
plt.title('Disaster Rate Over Time')
plt.xlabel('Start Year')
plt.ylabel('Number of Disasters')
plt.grid(True)
plt.tight_layout()
plt.show()
```



4. Disaster_Event_count_table

Tables allow users to compare event frequency across multiple years for each UN Sub-region while also displaying the total events and deaths in a structured and easily readable way. It gives the Precise and Exact Values Unlike charts (e.g. bar or line plots), tables provide exact numeric values. The use of a table in this context ensures clarity, accuracy, and comparative insight providing both an overview of disaster frequency and a sense of severity (through deaths), all in one view.

In [22]:
```
# Extract year from Start Date
Anita_df['Year'] = Anita_df['Start Date'].dt.year

# Create pivot table for Event occurance per year: UN Sub-region as rows, Year as columns
# Count of disasters
Disaster_Event_count_table = pd.pivot_table(
    Anita_df,
    values='Disaster duration',
    index='UN Sub Region',
    columns='Year',
    aggfunc='count',
    fill_value=0
)

# Add a Total column
Disaster_Event_count_table['Total'] = Disaster_Event_count_table.sum(axis=1)

# Pivot table for total deaths
death_table = Anita_df.groupby('UN Sub Region')['Total Deaths'].sum().rename('Total Deaths')

# Combine both tables
final_table = Disaster_Event_count_table.join(death_table)

# Reset index for neat display
final_table.reset_index(inplace=True)

# Show final result
(final_table)
```

Out[22]:

	UN Sub Region	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2017	2018	2019	Total	Total Deaths
0	Eastern Africa	4	10	2	3	5	6	2	3	2	1	1	1	1	1	42	6225.0
1	Middle Africa	5	4	2	2	2	1	4	1	0	2	0	0	1	0	24	5449.0
2	Northern Africa	2	4	2	1	0	0	0	1	0	0	0	0	0	1	11	1951.0
3	Southern Africa	0	1	0	3	0	0	0	0	0	0	0	0	0	0	4	86.0
4	Western Africa	12	5	2	4	2	4	1	6	2	0	0	0	0	0	38	7513.0

In [23]:

```
# Ensure 'Start Date' is datetime
Anita_df['Start Date'] = pd.to_datetime(Anita_df['Start Date'])

# Extract year
Anita_df['Year'] = Anita_df['Start Date'].dt.year

# Group by year and sum total deaths, then transpose
total_deaths_horizontal = Anita_df.groupby('Year')['Total Deaths'].sum().to_frame().T

# Set row index name
total_deaths_horizontal.index = ['Total Deaths']

# Display
total_deaths_horizontal
```

Out[23]:

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2017	2018	2019
Total Deaths	1775.0	4862.0	2672.0	4782.0	2058.0	1379.0	1943.0	1356.0	34.0	238.0	47.0	6.0	33.0	39.0

5. Average Affected and Deaths by Disaster Subtype in UN Sub Regions

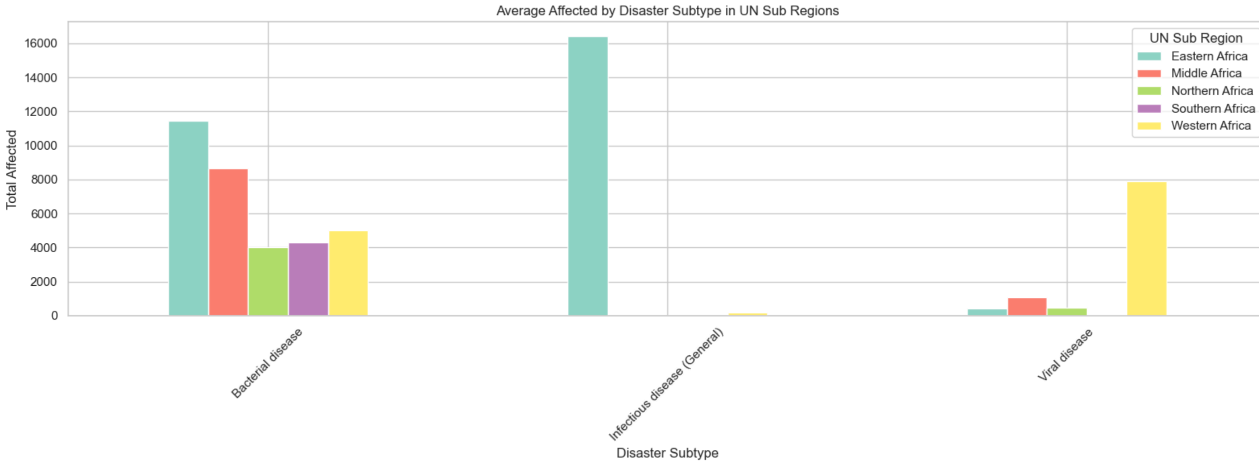
Clustered bars group the bars for each disaster subtype by UN Sub-region, making it easy to compare averages within each subtype across regions, and between subtypes within the same region. This layout makes patterns and disparities visually apparent and gives a clear Representation of Multiple Categories. You can easily identify which UN Sub-region is most affected by each disaster subtype, making it useful for understanding regional vulnerabilities and impact severity.

In [24]:

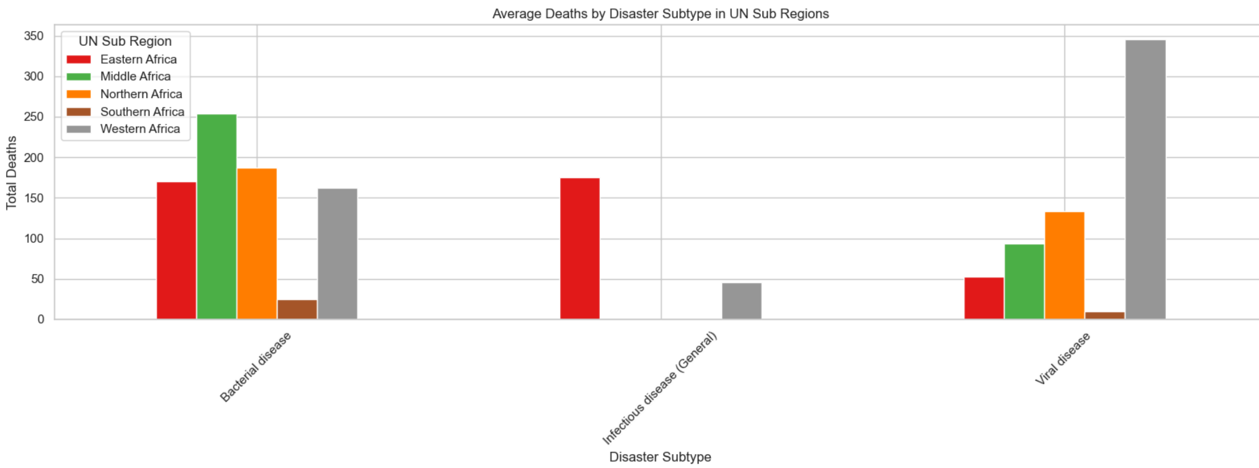
```
# Group and aggregate
Plot_2 = Anita_df.groupby(['UN Sub Region', 'Disaster Subtype'])[['Total Affected', 'Total Deaths']].mear

# Plot for Average Affected
plt.figure(figsize=(14, 6))
affected_pivot = Plot_2.pivot(index='Disaster Subtype', columns='UN Sub Region', values='Total Affected'
affected_pivot.plot(kind='bar', figsize=(16, 6), colormap='Set3')
plt.title('Average Affected by Disaster Subtype in UN Sub Regions')
plt.ylabel('Total Affected')
plt.xlabel('Disaster Subtype')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Plot for Average Deaths
plt.figure(figsize=(14, 6))
deaths_pivot = Plot_2.pivot(index='Disaster Subtype', columns='UN Sub Region', values='Total Deaths')
deaths_pivot.plot(kind='bar', figsize=(16, 6), colormap='Set1')
plt.title('Average Deaths by Disaster Subtype in UN Sub Regions')
plt.ylabel('Total Deaths')
plt.xlabel('Disaster Subtype')
plt.xticks(rotation=45)
```

```
plt.tight_layout()
plt.show()
```

<Figure size 1400x600 with 0 Axes>



<Figure size 1400x600 with 0 Axes>



6. Declaration, appeals and OFDA_BHA_Response_counts across UN Sub-region

Using a table to present the counts of Declarations, Appeals, and OFDA/BHA Responses across various UN Sub-regions is a strategic and effective choice because it allows for easy parallel comparison of different support/response indicators across each UN Sub-region. This format provides a clear breakdown of how each region was engaged in terms of international support or intervention.

In [25]:
```
# Strip column names and values (removes extra spaces)
Anita_df.columns = Anita_df.columns.str.strip()
for col in ['OFDA/BHA Response', 'Appeal', 'Declaration']:
    Anita_df[col] = Anita_df[col].astype(str).str.strip().str.title()  # Normalize to "Yes"/"No"

# Create crosstabs for each response type
OFDA_BHA_Response_counts = pd.crosstab(Anita_df['UN Sub Region'], Anita_df['OFDA/BHA Response'])
appeal_counts = pd.crosstab(Anita_df['UN Sub Region'], Anita_df['Appeal'])
declaration_counts = pd.crosstab(Anita_df['UN Sub Region'], Anita_df['Declaration'])

# Combine the tables
summary_table = OFDA_BHA_Response_counts.add_prefix('OFDA_') \
    .join(appeal_counts.add_prefix('Appeal_')) \
    .join(declaration_counts.add_prefix('Declaration_'))

# Display the table
(summary_table)
```

Out[25]:

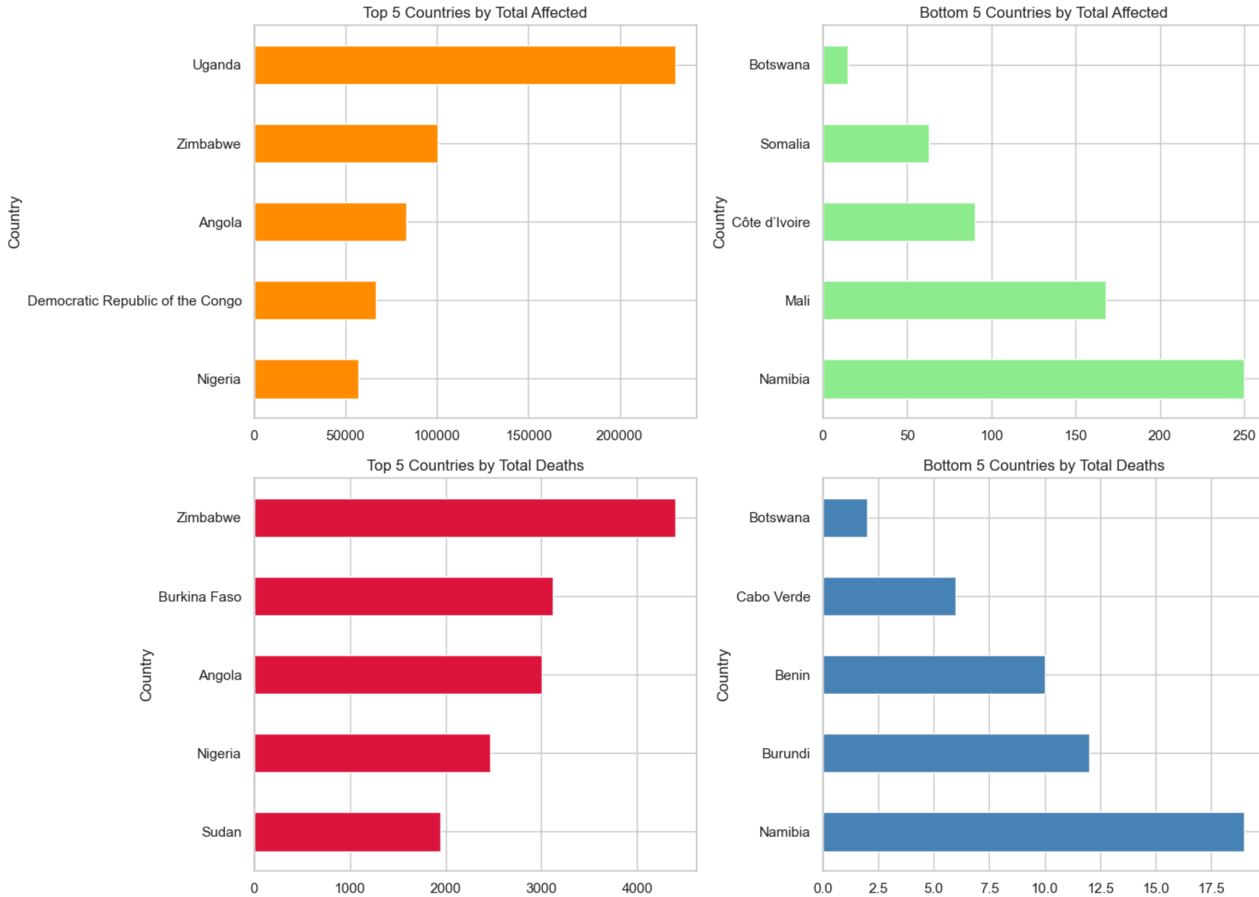
UN Sub Region	OFDA_No	OFDA_Yes	Appeal_No	Appeal_Yes	Declaration_No	Declaration_Yes
Eastern Africa	41	1	42	0	41	1
Middle Africa	22	2	23	1	22	2
Northern Africa	11	0	11	0	11	0
Southern Africa	4	0	4	0	3	1
Western Africa	37	1	38	0	38	0

7. Top and Bottom 5 Affected / Death for countries

Bar charts are ideal for ranking data showing the top 5 and bottom 5 countries and makes it easy to quickly identify countries with unusually high or low impact. Each bar represents a single country, so the viewer can immediately associate the length of the bar with the magnitude of the metric (affected or deaths). It emphasizes magnitude differences, supports ranking-based insights, and is ideal for summary visuals.

In [26]:

```
# Group by country and sum values
summary = Anita_df.groupby('Country')[['Total Affected', 'Total Deaths']].sum()

# Top and bottom 5 by Total Affected
Top_affected = summary.sort_values(by='Total Affected', ascending=False).head(5)
Bottom_affected = summary.sort_values(by='Total Affected', ascending=True).head(5)

# Top and bottom 5 by Total Deaths
Top_deaths = summary.sort_values(by='Total Deaths', ascending=False).head(5)
Bottom_deaths = summary.sort_values(by='Total Deaths', ascending=True).head(5)

# Plotting
fig, axes = plt.subplots(2, 2, figsize=(14, 10))

# Top Affected
Top_affected['Total Affected'].plot(kind='barh', ax=axes[0, 0], color='darkorange')
axes[0, 0].set_title('Top 5 Countries by Total Affected')
axes[0, 0].invert_yaxis()

# Bottom Affected
Bottom_affected['Total Affected'].plot(kind='barh', ax=axes[0, 1], color='lightgreen')
axes[0, 1].set_title('Bottom 5 Countries by Total Affected')
axes[0, 1].invert_yaxis()

# Top Deaths
Top_deaths['Total Deaths'].plot(kind='barh', ax=axes[1, 0], color='crimson')
axes[1, 0].set_title('Top 5 Countries by Total Deaths')
axes[1, 0].invert_yaxis()

# Bottom Deaths
Bottom_deaths['Total Deaths'].plot(kind='barh', ax=axes[1, 1], color='steelblue')
axes[1, 1].set_title('Bottom 5 Countries by Total Deaths')
axes[1, 1].invert_yaxis()

plt.tight_layout()
plt.show()
```



8. Epidemic Severity in Africa by Country

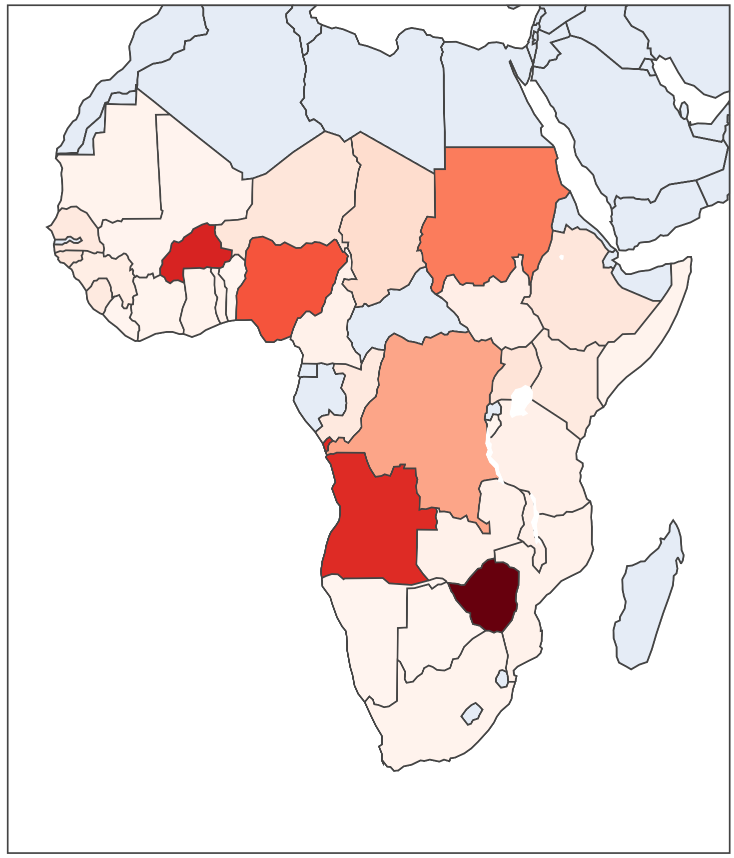
The Natural Earth map presents landmasses, terrain, and physical geography in a more realistic and recognizable form. This is valuable for public health visualization, where understanding regional proximity, environmental influences, and cross-border risks is important (e.g., how an epidemic in West Africa could spread to neighboring countries). This allows viewers to quickly identify which countries are more or less affected. The map format works well to view different levels of epidemic severity using a color gradient.

In [27]:

```
import plotly.express as px

# Filter for Epidemic disasters in Africa only
Epidemic_Africa = Anita_df[
    (Anita_df['Disaster Type'] == 'Epidemic') &
    (Anita_df['Region'] == 'Africa')
]

# Group to get total deaths per country
death_df = Epidemic_Africa.groupby(['Country', 'ISO', 'UN Sub Region'])['Total Deaths'].sum().reset_inde

# Group to get epidemic count per country
count_df = Epidemic_Africa.groupby(['Country', 'ISO', 'UN Sub Region']).size().reset_index(name='Epidemi

# Merge both metrics into one DataFrame
Africa_epidemic_summary = pd.merge(death_df, count_df, on=['Country', 'ISO', 'UN Sub Region'])

# Create choropleth map using Total Deaths as color, and show both values in hover
fig = px.choropleth(
    Africa_epidemic_summary,
    locations='ISO',
    color='Total Deaths',  # Color by total deaths
    hover_name='Country',
    hover_data={
        'UN Sub Region': True,
        'Total Deaths': True,
```

```
        'Epidemic Count': True,
        'ISO': False
    },
    color_continuous_scale='Reds',
    title='Epidemic Severity in Africa by Country (Total Deaths and Event Count)',
    projection='natural earth'
)

# Zoom to Africa
fig.update_geos(
    showcountries=True,
    fitbounds="locations",
    lataxis_range=[-40, 40],
    lonaxis_range=[-20, 60]
)

# Step 4: Center the map and adjust layout
fig.update_layout(
    autosize=False,
    width=1000,     # adjust as needed (try 800-1200 for balance)
    height=600,
    margin=dict(l=40, r=40, t=60, b=40),
    title_x=0.5     # Center the title
)

fig.show()

%pip install -U kaleido

# Save map as static PNG
import plotly.io as pio
pio.write_image(fig, 'newplot.png')

# Step 4: Display image so it appears in the PDF
from IPython.display import Image
Image('newplot.png')
```

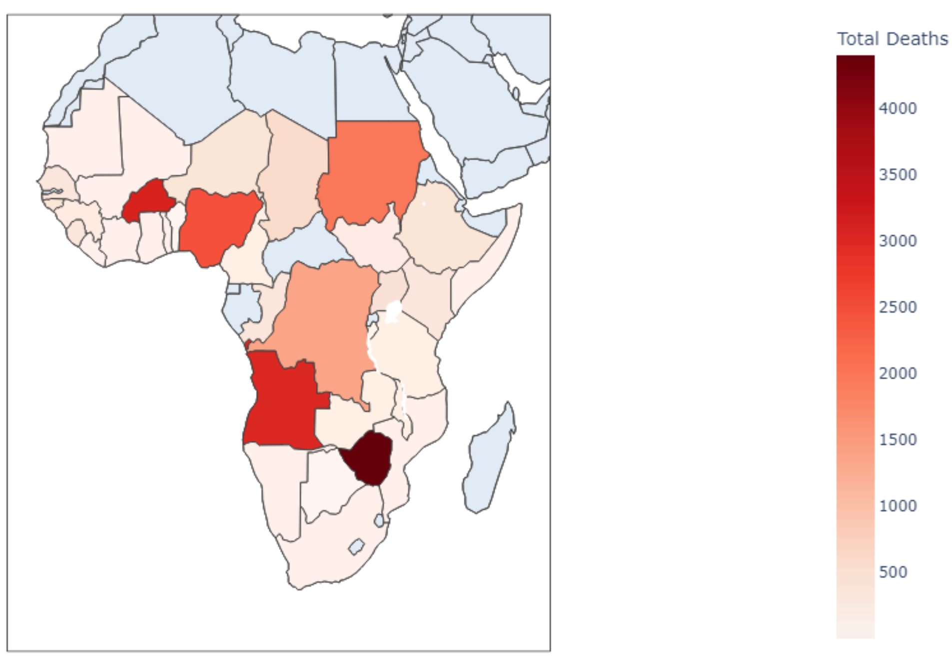
Epidemic Severity in Africa by Country (Total Deaths and Event C



```
Requirement already satisfied: kaleido in c:\users\user\anaconda3\lib\site-packages (0.2.1)
Note: you may need to restart the kernel to use updated packages.
```

Out[27]:

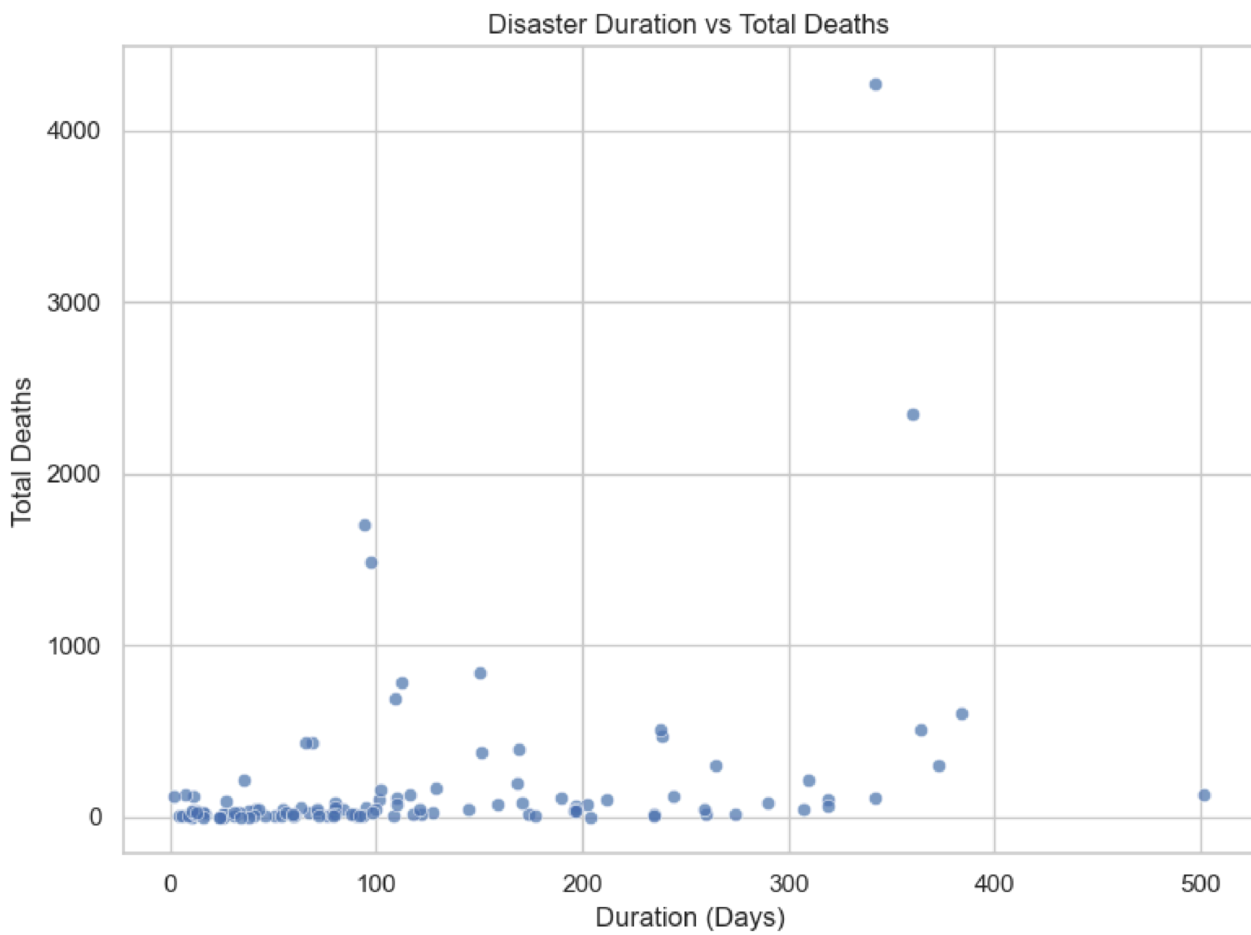
Epidemic Severity in Africa by Country (Total Deaths and Event Count)



9. Disaster Duration vs Total Deaths

A scatter plot is ideal for showing whether a relationship exists between two quantitative variables. in this case, how duration of a disaster may influence the number of deaths. It Identifies Outliers and clearly exposes extreme values (e.g., very long disasters with unusually low or high death tolls).

In [28]:
```
# Clean data
df = Anita_df[['Disaster duration', 'Total Deaths']]
df = df[(df['Disaster duration'] > 0) & (df['Total Deaths'] > 0)]


# Plot
plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x='Disaster duration', y='Total Deaths', alpha=0.7)
plt.title('Disaster Duration vs Total Deaths')
plt.xlabel('Duration (Days)')
plt.ylabel('Total Deaths')
plt.tight_layout()
plt.show()
```



📊 📌 Overview and Insights from the Analysis

In this section, we explain how various **Disaster subtypes** are distributed across different **UN Sub-regions and years**. This analysis helps in identifying which subtype of disease are more common in particular regions and countries.

🔍 Regional Interpretation

🌍 **Eastern Africa:** They face a mix of **Cholera, Meningococcal diseases,Rift valley fever, typhoid fever,Acute watery diarrhoeal syndrome**. **Bacterial infections** and **zoonotic diseases**. Cholera and typhoid fever suggest issues with **water sanitation and hygiene**, while Rift Valley fever, is a mosquito-borne illness, points to **vector control weaknesses**. Meningococcal outbreaks are often seasonal, tied to dry periods.

🌍 **Middle Africa:** Middle Africa experienced a high incidence of **Cholera, Hepatitis E, Meningitis, Meningococcal disease**, including **Ebola**, which requires urgent containment strategies. Hepatitis E and Cholera imply **unsafe water sources**, while meningitis and meningococcal diseases remain a recurring public health threat, especially in the **African meningitis belt**.

🌍 **Southern Africa:** Southern Africa shows a **narrower disease spread** dominated by **Cholera and Polio**. While Polio has been nearly eradicated globally, its appearance here suggests **immunization gaps** or resurgence. Cholera remains a threat due to **periodic outbreaks in vulnerable communities**.

🌍 **Northern Africa:** Northern Africa also shows a **narrower disease spread** dominated by **bacterial infections** and **arboviruses** like **Meningococcal diseases, Cholera, Acute watery diarrhoeal syndrome, Rift Valley fever and Yellow fever**. The presence of these diseases signals **vector presence**, possibly worsened by **climatic conditions** and **transborder livestock movement**.

🌍 **Western Africa:** Western Africa's recurring **Meningococcal diseases, Acute watery diarrhoeal syndrome, Cholera** and **meningitis outbreaks** align with patterns in other sub-regions, but the presence of **Dengue**, a mosquito-borne virus, is notable and may reflect **urban crowding and stagnant water**. The region remains vulnerable to both **bacterial and viral outbreaks**.

📌Insights

Epidemic subtypes vary across sub-regions, **Meningitis** events are common in the Sahel belt (West/Middle Africa), **Cholera and Meningococcal diseases** are the most frequently occurring epidemic types. This suggests that **water sanitation**, **vaccination**, and **early warning systems** are critical priorities. **Ebola** dominated reports in Middle Africa, but occurred less frequently elsewhere. Preventable diseases continue to pose major challenges, indicating gaps in vaccination, health education, and infrastructure. Least affected and least deaths occurred in more stable nations like Mali, senegal, Cote d'ivoire, ghana, benin republic etc. Some countries face frequent epidemics with lower deaths (indicating better control), while others suffer infrequent but deadly outbreaks (like Ebola). Majority of epidemic events are of short duration and result in relatively low fatalities. Some Short Epidemics Were Extremely Deadly. One or two points show very high death tolls (above 2,000) under 300-400 days long. These could represent cholera, Meningococcal diseases outbreaks known for rapid spread and high mortality even in short periods. Events lasting over 200 days (even up to 500) mostly stay under 1,000 deaths, suggesting long duration doesn't guarantee high fatality.

Epidemic events increased with visible peaks between 2005 to 2006 due to bacterial diseases like Cholera, Meningococcal diseases. Later years (2007-2019) show decline. The initial peak in 2005 likely reflects high vulnerability, limited infrastructure, and major outbreaks of cholera and Meningococcal diseases. The decline by 2019 suggests improvements in epidemic prevention, control, and public health investments.

The sub-regions had little or No declarations, appeals or even International US humanitarian Aids, and Africa is a less developed Region. Africa's sub-regions most likely received little or no declarations, appeals, or US humanitarian aid during epidemics due to various reasons such as mix of underreporting, aid prioritization criteria, chronic crisis normalization, limited global attention or political interest. Countries or sub-regions that are not geo politically significant can most likely not be prioritized, even during health emergencies. This reflects deep inequities in global health response systems.

The analysis of top epidemic events by UN Sub-region provides **vital insights for region-specific planning**. Each area faces **unique health challenges**, but common themes like poor water access, inadequate vaccination coverage, and climate-sensitive vectors are shared across borders. This analysis reveals a clear pattern: **Epidemic threats are regional and often recurring**.

By knowing all these informations and data, the diseases that are most frequent in each areas are known and certain focus should be on how:

- Governments can tailor public health policies and build resilient healthcare systems across the African continent and also focus on early detection and rapid response.
- Humanitarian organizations can Invest and prioritize stockpiling vaccines and treatments.
- Researchers can focus on locally relevant disease dynamics.
- Government can Strengthen water safety programs, vector control and public sanitation and hygiene, routine immunization and Vaccination

📊 Conclusion

Bacterial and viral diseases, are the most frequently reported disasters. While duration plays a role in the impact of an epidemic, it is not the primary predictor of death toll. Short but deadly outbreaks emphasize the need for rapid response and containment, while longer, less fatal outbreaks highlight the importance of sustained public health infrastructure and prevention. A few data points stand out with exceptionally high deaths (2000–4000+), highlighting severe and high-impact outbreaks that warrant deeper investigation.

Tools and Libraries used

- Pandas
- Numpy
- matplotlib
- seaborn
- plotly.express
- -U kaleido
- plotly.io
- IPython.display

Short Epidemics, Deep Scars□

🌍 "The deadliest outbreaks weren't always the longest, they were the fastest and the most unprepared for". Epidemics don't wait, Neither should we.

```
In [ ]:
```