# globalmisssingmigrants-1

January 5, 2024

## ***Global Missing Migrants***

This dataset presents a solemn record of individuals who embarked on perilous journeys towards international destinations, only to go missing or tragically lose their lives along the way. The dataset is a result of the ongoing efforts of the Missing Migrants Project, an initiative by the International Organization for Migration (IOM) since 2014.

Migration is a complex and multifaceted phenomenon that touches the lives of millions of people worldwide. This dataset sheds light on the challenges faced by migrants, as well as the immense courage and resilience they display. While the numbers presented here offer a glimpse into the scope of the issue, it's important to acknowledge that the true extent of the problem is likely underestimated due to the inherent difficulties in collecting such data.

The data here shows details such as the date the migrants went missing, the number of migrants that went missing, the region in which the incident occurred, etc. It has the potential to be very helpful in determining the severity of the issue of missing migrants in different regions across the world.

FEATURES:

Incident Type: Type of migration incident

Incident Year: Year when the incident occurred

Reported Month: Month when the incident was reported

Region of Origin: Geographical region where the migrants originated

Region of Incident: Geographical region where the incident occurred

Country of Origin: Country from which the migrants originated

Number of Dead: Number of confirmed deceased migrants

Minimum Estimated Number of Missing: Minimum estimated count of missing migrants

Total Number of Dead and Missing: Total count of both deceased and missing migrants

Number of Survivors: Number of migrants who survived the incident

Number of Females: Number of female migrants involved

Number of Males: Number of male migrants involved

Number of Children: Number of children migrants involved

Cause of Death: Cause of death for the migrants

GOAL

To predict the Total Number of Dead and Missing

[ ]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import figure
df=pd.read_csv('/content/Global Missing Migrants Dataset.csv.zip')
df
```

[ ]:

```
      Incident Type  Incident year Reported Month  \
0          Incident           2014        January
1          Incident           2014        January
2          Incident           2014        January
3          Incident           2014        January
4          Incident           2014        January
…             …            …            …
13015      Incident           2023           July
13016      Incident           2023           July
13017      Incident           2023           July
13018      Incident           2023           July
13019      Incident           2023           July

                    Region of Origin Region of Incident     Country of Origin  \
0                    Central America      North America             Guatemala
1      Latin America / Caribbean (P)      North America               Unknown
2      Latin America / Caribbean (P)      North America               Unknown
3                    Central America      North America                Mexico
4                    Northern Africa             Europe                 Sudan
…                               …                …                   …
13015                   Western Asia       Western Asia  Syrian Arab Republic
13016             Western Africa (P)       Western Asia               Unknown
13017                 Western Africa    Northern Africa               Senegal
13018                          Mixed    Northern Africa               Unknown
13019             Western Africa (P)     Western Africa               Unknown

       Number of Dead  Minimum Estimated Number of Missing  \
0                 1.0                                    0
1                 1.0                                    0
2                 1.0                                    0
3                 1.0                                    0
4                 1.0                                    0
…                …                                  …
13015             4.0                                    0
13016             2.0                                    0
13017            13.0                                    0
```

```
13018             6.0                                    0
13019            16.0                                   37

       Total Number of Dead and Missing  Number of Survivors  \
0                                     1                    0
1                                     1                    0
2                                     1                    0
3                                     1                    0
4                                     1                    2
…                                 …                  …
13015                                 4                    0
13016                                 2                    0
13017                                13                    6
13018                                 6                   48
13019                                53                    2

       Number of Females  Number of Males  Number of Children  \
0                      0                1                   0
1                      0                0                   0
2                      0                0                   0
3                      0                1                   0
4                      0                1                   0
…                  …              …                 …
13015                  0                4                   0
13016                  0                2                   0
13017                  0                0                   0
13018                  0                0                   0
13019                  2                0                   0

                                          Cause of Death  \
0                                       Mixed or unknown
1                                       Mixed or unknown
2                                       Mixed or unknown
3                                               Violence
4      Harsh environmental conditions / lack of adequ…
…                                                  …
13015  Vehicle accident / death linked to hazardous t…
13016  Vehicle accident / death linked to hazardous t…
13017                                           Drowning
13018                                           Drowning
13019                                           Drowning

                                         Migration route  \
0                               US-Mexico border crossing
1                               US-Mexico border crossing
2                               US-Mexico border crossing
3                               US-Mexico border crossing
```

```
4                                                    NaN
…                                                  …
13015                          Türkiye-Europe land route
13016                          Türkiye-Europe land route
13017  Western Africa / Atlantic route to the Canary …
13018  Western Africa / Atlantic route to the Canary …
13019  Western Africa / Atlantic route to the Canary …

                                       Location of death  \
0      Pima Country Office of the Medical Examiner ju…
1      Pima Country Office of the Medical Examiner ju…
2      Pima Country Office of the Medical Examiner ju…
3                             near Douglas, Arizona, USA
4                      Border between Russia and Estonia
…                                                  …
13015  In Ipsala, Edirne province, Türkiye - travelli…
13016  At the Kapıkule Türkiye-Bulgaria Border Gate, …
13017  Off the coasts of Dakhla, Western Sahara - 6 s…
13018  Unspecified location off the coast of Nador, M…
13019           Off the coast of Ouakam, Dakar, Senegal

                                      Information Source  \
0      Pima County Office of the Medical Examiner (PC…
1      Pima County Office of the Medical Examiner (PC…
2      Pima County Office of the Medical Examiner (PC…
3      Ministry of Foreign Affairs Mexico, Pima Count…
4                      EUBusiness (Agence France-Presse)
…                                                  …
13015            Andalou Agency, Son Dakika, Orient News
13016                               Son Dakika, Hurriyet
13017          Barron's News, InfoMigrants, IOM Morrocco
13018          El Nashra, Swiss Info; CGTN, IOM Morrocco
13019                                        IOM Senegal

                   Coordinates UNSD Geographical Grouping
0       31.650259, -110.366453           Northern America
1         31.59713, -111.73756           Northern America
2         31.94026, -113.01125           Northern America
3       31.506777, -109.315632           Northern America
4                  59.1551, 28            Northern Europe
…                        …                        …
13015   40.91271268, 26.369657               Western Asia
13016   41.71697242, 26.351489               Western Asia
13017  23.72836078, -15.901632              Uncategorized
13018   35.17187365, -2.903182              Uncategorized
13019  14.71870705, -17.506255              Uncategorized
```

```
[13020 rows x 19 columns]
```

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

```
[ ]:   Incident Type  Incident year Reported Month                Region of Origin  \
     0      Incident           2014        January                 Central America
     1      Incident           2014        January   Latin America / Caribbean (P)
     2      Incident           2014        January   Latin America / Caribbean (P)
     3      Incident           2014        January                 Central America
     4      Incident           2014        January                 Northern Africa

       Region of Incident Country of Origin  Number of Dead  \
     0      North America         Guatemala             1.0
     1      North America           Unknown             1.0
     2      North America           Unknown             1.0
     3      North America            Mexico             1.0
     4             Europe             Sudan             1.0

        Minimum Estimated Number of Missing  Total Number of Dead and Missing  \
     0                                    0                                 1
     1                                    0                                 1
     2                                    0                                 1
     3                                    0                                 1
     4                                    0                                 1

        Number of Survivors  Number of Females  Number of Males  \
     0                    0                  0                1
     1                    0                  0                0
     2                    0                  0                0
     3                    0                  0                1
     4                    2                  0                1

        Number of Children                                     Cause of Death  \
     0                   0                                   Mixed or unknown
     1                   0                                   Mixed or unknown
     2                   0                                   Mixed or unknown
     3                   0                                           Violence
     4                   0  Harsh environmental conditions / lack of adequ…

                    Migration route  \
     0  US-Mexico border crossing
     1  US-Mexico border crossing
     2  US-Mexico border crossing
     3  US-Mexico border crossing
     4                        NaN

                                       Location of death  \
0  Pima Country Office of the Medical Examiner ju…
1  Pima Country Office of the Medical Examiner ju…
2  Pima Country Office of the Medical Examiner ju…
3                         near Douglas, Arizona, USA
4                  Border between Russia and Estonia

                                  Information Source            Coordinates  \
0  Pima County Office of the Medical Examiner (PC…  31.650259, -110.366453
1  Pima County Office of the Medical Examiner (PC…    31.59713, -111.73756
2  Pima County Office of the Medical Examiner (PC…    31.94026, -113.01125
3  Ministry of Foreign Affairs Mexico, Pima Count…  31.506777, -109.315632
4                  EUBusiness (Agence France-Presse)            59.1551, 28

  UNSD Geographical Grouping
0           Northern America
1           Northern America
2           Northern America
3           Northern America
4            Northern Europe
```

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

```
[ ]:       Incident Type  Incident year Reported Month    Region of Origin   \
     13015      Incident           2023           July        Western Asia
     13016      Incident           2023           July  Western Africa (P)
     13017      Incident           2023           July      Western Africa
     13018      Incident           2023           July               Mixed
     13019      Incident           2023           July  Western Africa (P)

           Region of Incident     Country of Origin  Number of Dead  \
     13015       Western Asia  Syrian Arab Republic             4.0
     13016       Western Asia               Unknown             2.0
     13017    Northern Africa               Senegal            13.0
     13018    Northern Africa               Unknown             6.0
     13019     Western Africa               Unknown            16.0

            Minimum Estimated Number of Missing  Total Number of Dead and Missing  \
     13015                                    0                                 4
     13016                                    0                                 2
     13017                                    0                                13
     13018                                    0                                 6
     13019                                   37                                53

            Number of Survivors  Number of Females  Number of Males  \
     13015                    0                  0                4
     13016                    0                  0                2
     13017                    6                  0                0
```

```
13018                   48                 0                0
13019                    2                 2                0

       Number of Children                                     Cause of Death  \
13015                   0  Vehicle accident / death linked to hazardous t…
13016                   0  Vehicle accident / death linked to hazardous t…
13017                   0                                           Drowning
13018                   0                                           Drowning
13019                   0                                           Drowning

                                         Migration route  \
13015                          Türkiye-Europe land route
13016                          Türkiye-Europe land route
13017  Western Africa / Atlantic route to the Canary …
13018  Western Africa / Atlantic route to the Canary …
13019  Western Africa / Atlantic route to the Canary …

                                       Location of death  \
13015  In Ipsala, Edirne province, Türkiye - travelli…
13016  At the Kapıkule Türkiye-Bulgaria Border Gate, …
13017  Off the coasts of Dakhla, Western Sahara - 6 s…
13018  Unspecified location off the coast of Nador, M…
13019           Off the coast of Ouakam, Dakar, Senegal

                              Information Source              Coordinates  \
13015    Andalou Agency, Son Dakika, Orient News   40.91271268, 26.369657
13016                       Son Dakika, Hurriyet   41.71697242, 26.351489
13017  Barron's News, InfoMigrants, IOM Morrocco  23.72836078, -15.901632
13018  El Nashra, Swiss Info; CGTN, IOM Morrocco   35.17187365, -2.903182
13019                                IOM Senegal  14.71870705, -17.506255

      UNSD Geographical Grouping
13015               Western Asia
13016               Western Asia
13017              Uncategorized
13018              Uncategorized
13019              Uncategorized
```

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

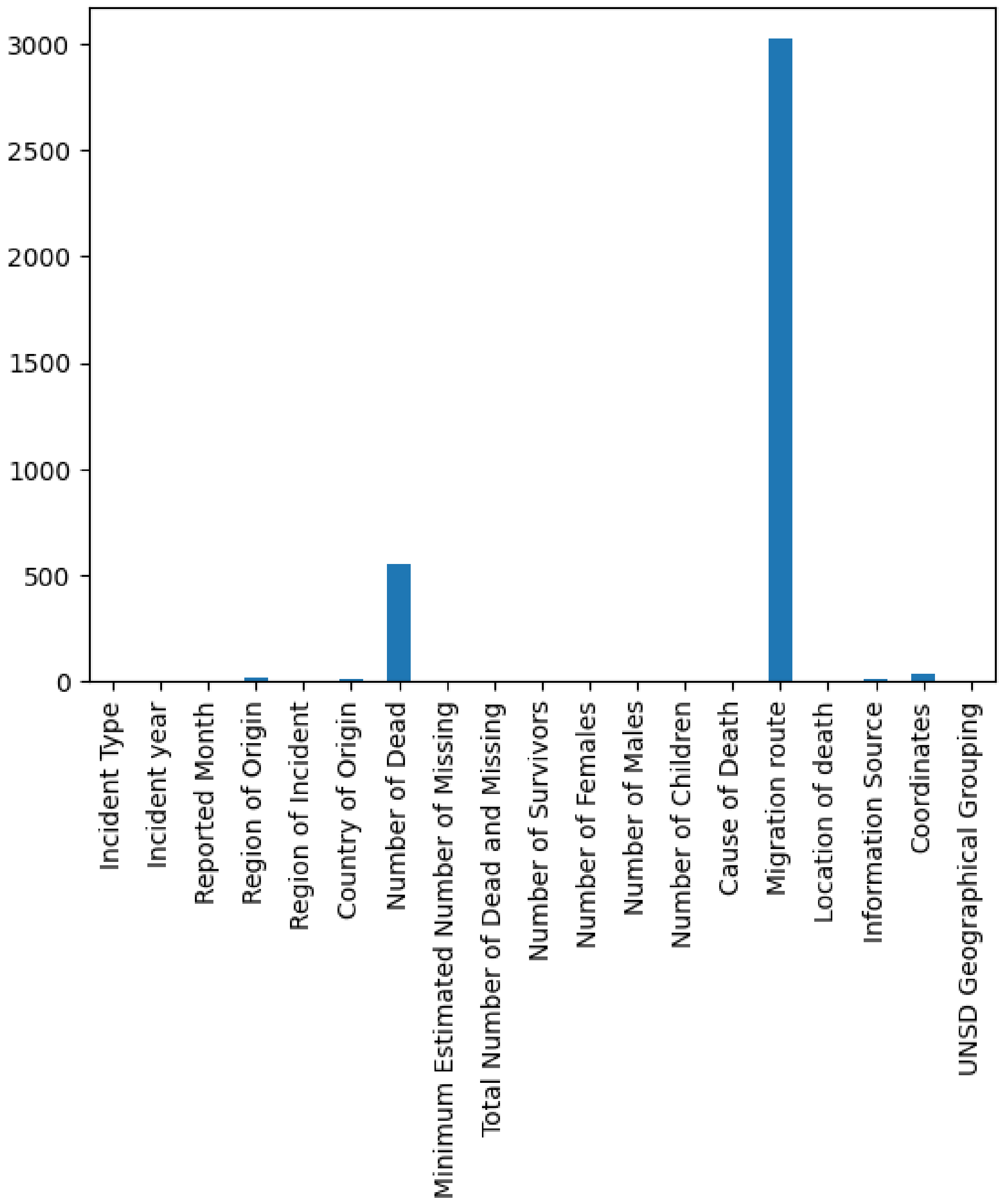
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13020 entries, 0 to 13019
Data columns (total 19 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   Incident Type                        13020 non-null  object
 1   Incident year                        13020 non-null  int64
 2   Reported Month                      13020 non-null  object
 3   Region of Origin                    12998 non-null  object
 4   Region of Incident                  13020 non-null  object
 5   Country of Origin                   13012 non-null  object
 6   Number of Dead                      12470 non-null  float64
 7   Minimum Estimated Number of Missing  13020 non-null  int64
 8   Total Number of Dead and Missing    13020 non-null  int64
 9   Number of Survivors                 13020 non-null  int64
 10  Number of Females                   13020 non-null  int64
 11  Number of Males                     13020 non-null  int64
 12  Number of Children                  13020 non-null  int64
 13  Cause of Death                      13020 non-null  object
 14  Migration route                     9999 non-null   object
 15  Location of death                   13020 non-null  object
 16  Information Source                  13012 non-null  object
 17  Coordinates                         12984 non-null  object
 18  UNSD Geographical Grouping          13019 non-null  object
dtypes: float64(1), int64(7), object(11)
memory usage: 1.9+ MB
```

```
[ ]: df.isna().sum()
```

```
[ ]: Incident Type                            0
     Incident year                            0
     Reported Month                           0
     Region of Origin                        22
     Region of Incident                       0
     Country of Origin                        8
     Number of Dead                         550
     Minimum Estimated Number of Missing      0
     Total Number of Dead and Missing         0
     Number of Survivors                      0
     Number of Females                        0
     Number of Males                          0
     Number of Children                       0
     Cause of Death                           0
     Migration route                       3021
     Location of death                        0
     Information Source                       8
     Coordinates                             36
     UNSD Geographical Grouping               1
     dtype: int64
```

```
[ ]: df.isna().sum().plot(kind='bar')
```

```
[ ]: <Axes: >
```



[ ]:
```
df.dtypes
```

[ ]:
```
Incident Type                           object
Incident year                            int64
Reported Month                          object
Region of Origin                        object
Region of Incident                      object
Country of Origin                       object
```

```
Number of Dead                                 float64
Minimum Estimated Number of Missing              int64
Total Number of Dead and Missing                 int64
Number of Survivors                              int64
Number of Females                                int64
Number of Males                                  int64
Number of Children                               int64
Cause of Death                                  object
Migration route                                 object
Location of death                               object
Information Source                              object
Coordinates                                     object
UNSD Geographical Grouping                      object
dtype: object
```

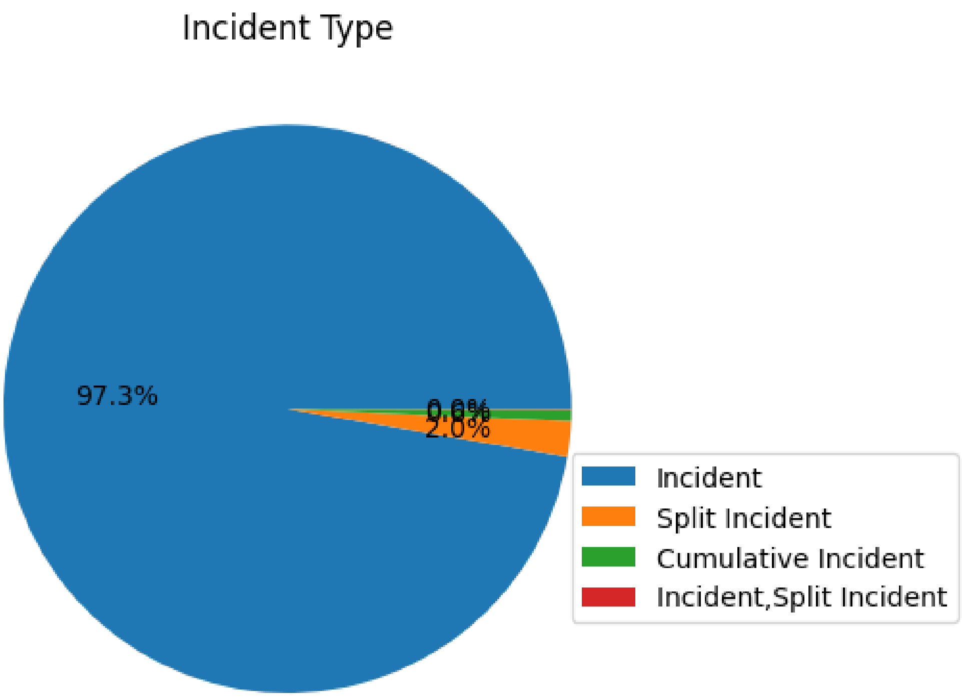
***Visualization of object Columns***

```
[ ]: incident=df['Incident Type'].value_counts()
     incident
```

```
[ ]: Incident                    12670
     Split Incident                261
     Cumulative Incident            84
     Incident,Split Incident         5
     Name: Incident Type, dtype: int64
```

```
[ ]: plt.pie(incident,autopct='%2.1f%%')
     plt.legend(incident.index,loc=(0.9,0.2))
     plt.title("Incident Type")
```

```
[ ]: Text(0.5, 1.0, 'Incident Type')
```

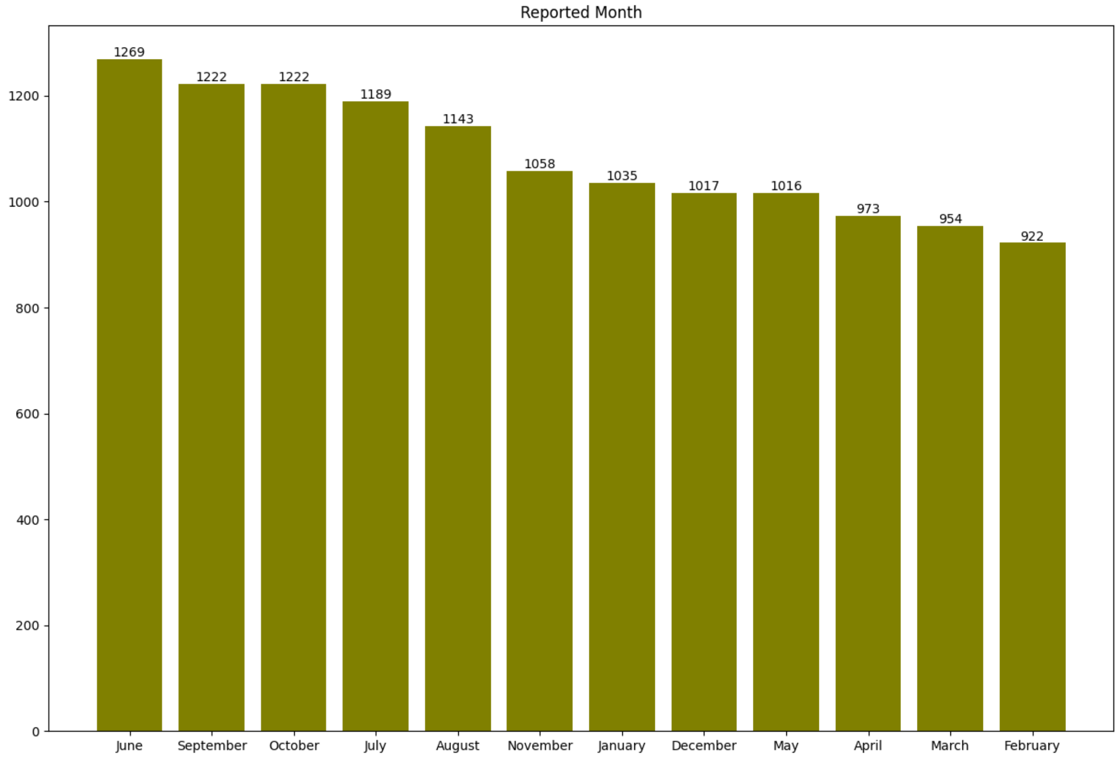


```
[ ]: Report=df['Reported Month'].value_counts()
     Report
```

```
[ ]: June         1269
     September    1222
     October      1222
     July         1189
     August       1143
     November     1058
     January      1035
     December     1017
     May          1016
     April         973
     March         954
     February      922
     Name: Reported Month, dtype: int64
```

```
[ ]: plt.figure(figsize=(15,10))
     plt.bar(Report.index,Report,color="olive")
     for i,count in enumerate(Report):
       plt.text(Report.index[i],count,str(count),ha='center',va='bottom')
     plt.title('Reported Month')
```

```
[ ]: Text(0.5, 1.0, 'Reported Month')
```



```
[ ]: Origin=df['Region of Origin'].value_counts()
     Origin
```

```
[ ]: Latin America / Caribbean (P)    2164
     Southern Asia                    1904
     Unknown                          1737
     Central America                  1565
     Sub-Saharan Africa (P)           1528
     Eastern Africa (P)               1133
     Northern Africa                   452
     Western Asia                      432
     South America                     322
     Eastern Africa                    298
     Caribbean                         278
     Western / Southern Asia (P)       245
     Western Africa                    229
     Northern Africa (P)               122
     Sub-Saharan Africa                116
     Mixed                             111
     Southern Asia (P)                  90
     South-eastern Asia                 68
```

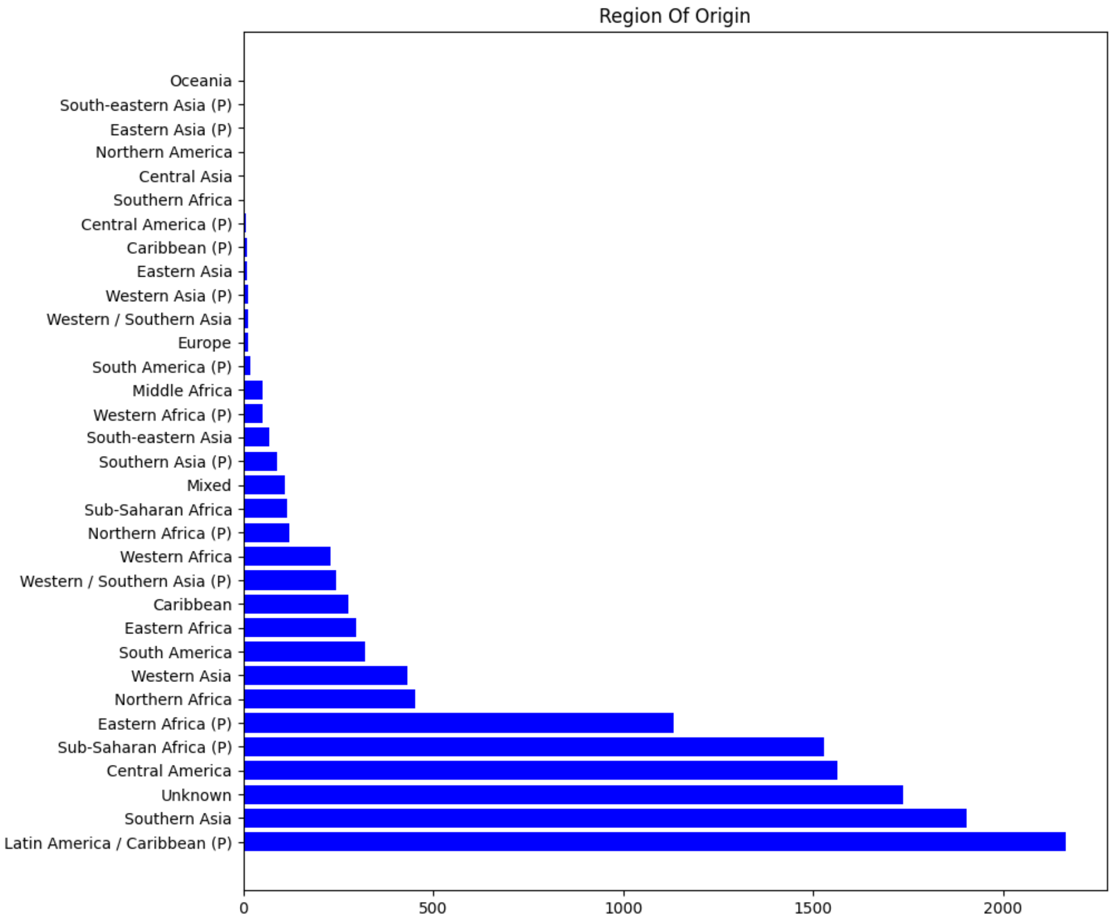
```
Western Africa (P)              52
Middle Africa                   51
South America (P)               18
Europe                          14
Western / Southern Asia         14
Western Asia (P)                12
Eastern Asia                    11
Caribbean (P)                   10
Central America (P)              8
Southern Africa                  5
Central Asia                     4
Northern America                 2
Eastern Asia (P)                 1
South-eastern Asia (P)           1
Oceania                          1
Name: Region of Origin, dtype: int64
```

[ ]:

```
plt.figure(figsize=(10,10))
plt.barh(Origin.index,Origin,color="blue")
plt.title('Region Of Origin')
```

[ ]:

```
Text(0.5, 1.0, 'Region Of Origin')
```



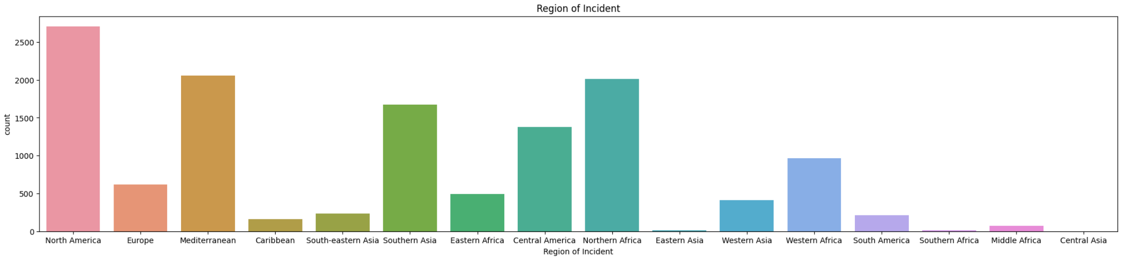
```
[ ]: Incident1=df["Region of Incident"].value_counts()
     Incident1
```

```
[ ]: North America         2706
     Mediterranean         2055
     Northern Africa       2014
     Southern Asia         1673
     Central America       1375
     Western Africa         967
     Europe                 619
     Eastern Africa         489
     Western Asia           414
     South-eastern Asia     237
     South America          209
     Caribbean              160
     Middle Africa           75
     Southern Africa         16
     Eastern Asia            10
```

```
Central Asia              1
Name: Region of Incident, dtype: int64
```

```
[ ]: plt.figure(figsize=(25,5))
     sns.countplot(x='Region of Incident',data=df)
     plt.title("Region of Incident")
```

```
[ ]: Text(0.5, 1.0, 'Region of Incident')
```



```
[ ]: country=df['Country of Origin'].value_counts()
     country
```

```
[ ]: Unknown                                                            7220
     Afghanistan                                                        1702
     Mexico                                                              709
     Syrian Arab Republic                                                308
     Honduras                                                            307
                                                                        …
     Nigeria,Sudan                                                         1
     Mali,Senegal,Unknown                                                  1
     Somalia,Unknown                                                       1
     Cameroon,Côte d'Ivoire,Democratic Republic of the Congo,Tunisia       1
     Gambia,Mali,Nigeria,Senegal,Unknown                                   1
     Name: Country of Origin, Length: 335, dtype: int64
```

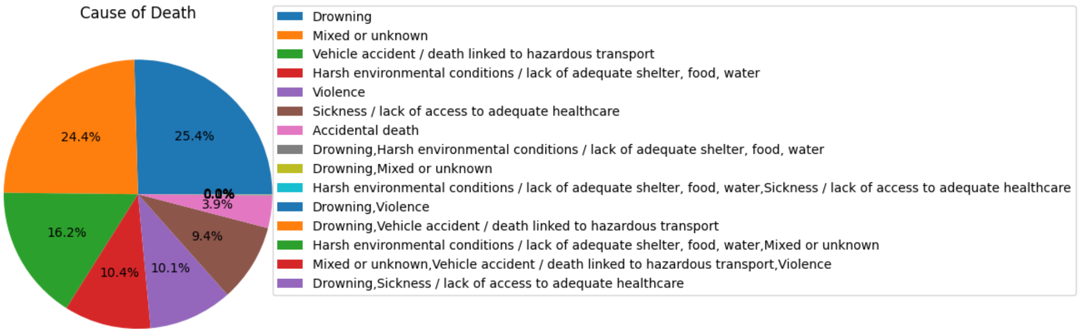
```
[ ]: Death=df['Cause of Death'].value_counts()
     Death
```

```
[ ]: Drowning
     3313
     Mixed or unknown
     3175
     Vehicle accident / death linked to hazardous transport
     2112
     Harsh environmental conditions / lack of adequate shelter, food, water
     1360
     Violence
```

```
1313
Sickness / lack of access to adequate healthcare
1219
Accidental death
507
Drowning,Harsh environmental conditions / lack of adequate shelter, food, water
8
Drowning,Mixed or unknown
4
Harsh environmental conditions / lack of adequate shelter, food, water,Sickness
/ lack of access to adequate healthcare       3
Drowning,Violence
2
Drowning,Vehicle accident / death linked to hazardous transport
1
Harsh environmental conditions / lack of adequate shelter, food, water,Mixed or
unknown                                       1
Mixed or unknown,Vehicle accident / death linked to hazardous transport,Violence
1
Drowning,Sickness / lack of access to adequate healthcare
1
Name: Cause of Death, dtype: int64
```

```
[ ]: plt.pie(Death,autopct='%1.1f%%')
     plt.legend(Death.index,loc=(0.9,0.2))
     plt.title("Cause of Death")
```

```
[ ]: Text(0.5, 1.0, 'Cause of Death')
```

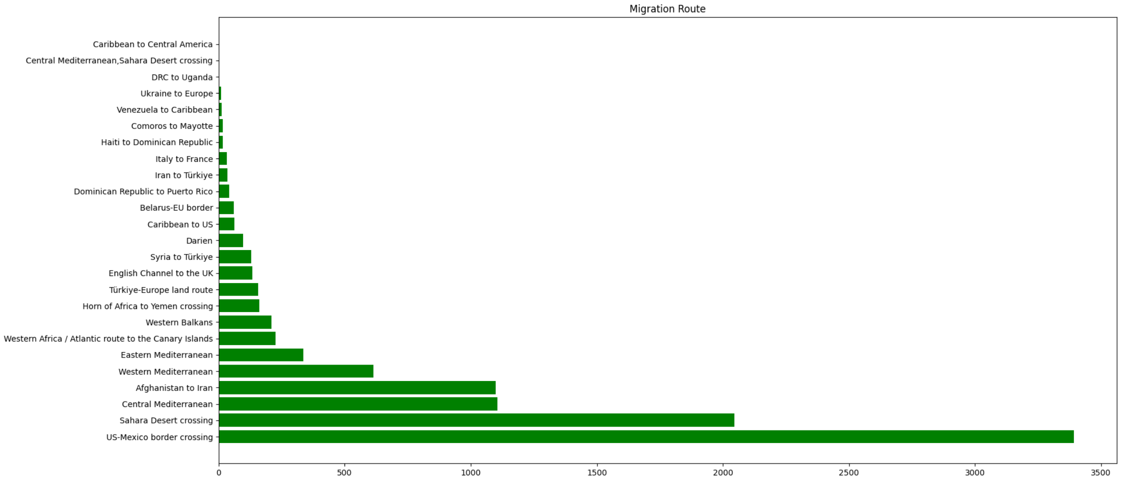


```
[ ]: Migration=df['Migration route'].value_counts()
     Migration
```

```
[ ]: US-Mexico border crossing                                3392
     Sahara Desert crossing                                   2046
     Central Mediterranean                                    1106
     Afghanistan to Iran                                      1099
     Western Mediterranean                                     614
     Eastern Mediterranean                                     336
     Western Africa / Atlantic route to the Canary Islands     226
     Western Balkans                                           210
     Horn of Africa to Yemen crossing                          161
     Türkiye-Europe land route                                 157
     English Channel to the UK                                 134
     Syria to Türkiye                                          129
     Darien                                                     98
     Caribbean to US                                            63
     Belarus-EU border                                          61
     Dominican Republic to Puerto Rico                          42
     Iran to Türkiye                                            34
     Italy to France                                            33
     Haiti to Dominican Republic                                17
     Comoros to Mayotte                                         16
     Venezuela to Caribbean                                     11
     Ukraine to Europe                                           9
     DRC to Uganda                                               3
     Central Mediterranean,Sahara Desert crossing                1
     Caribbean to Central America                                1
     Name: Migration route, dtype: int64
```

```
[ ]: plt.figure(figsize=(20,10))
     plt.barh(Migration.index,Migration,color="green")
     plt.title("Migration Route")
```

```
[ ]: Text(0.5, 1.0, 'Migration Route')
```



```
[ ]: Location=df['Location of death'].value_counts()
     Location
```

```
[ ]: Pima Country Office of the Medical Examiner jurisdiction, Arizona, USA (see
     coordinates for exact location)    1061
     Pima County Office of the Medical Examiner jurisdiction, Arizona, USA (see
     coordinates for exact location)      404
     Reported at Milak border crossing, Iran
     200
     Agadez, Niger
     121
     Sahara desert, Libya
     116
                                                                                …
     Evros River, near Orestiada, Greece
     1
     Bodies recovered near Plage de Trougout, Nador, Morocco
     1
     Namanga, Tanzania-Kenya border
     1
     Ndola, Zambia, near border with Democratic Republic of the Congo
     1
     Off the coast of Ouakam, Dakar, Senegal
     1
     Name: Location of death, Length: 7460, dtype: int64
```

```
[ ]: Info=df["Information Source"].value_counts()
     Info
```

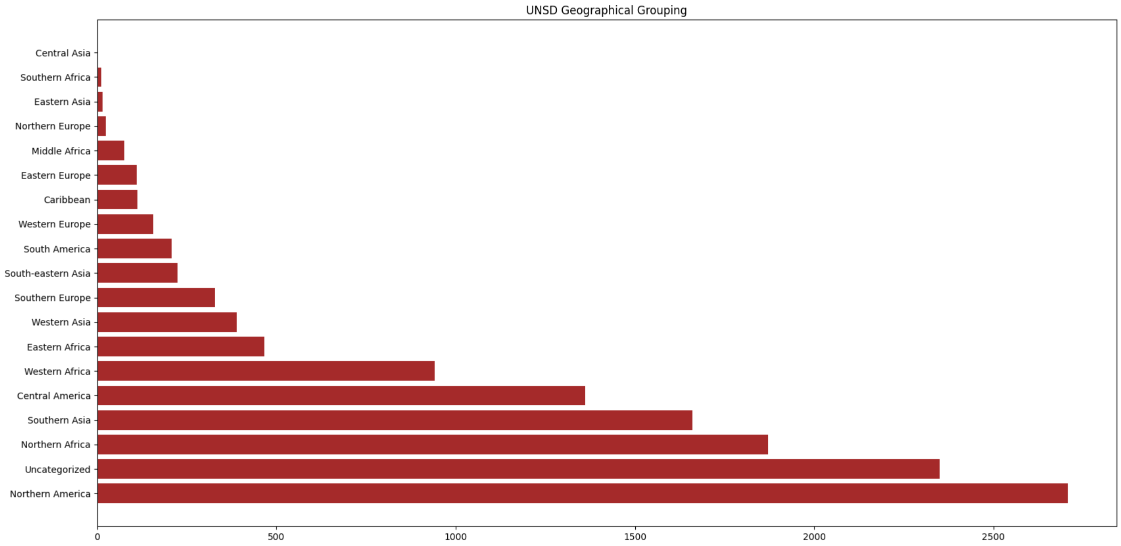
```
[ ]: IOM Afghanistan                                          1538
     Pima County Office of the Medical Examiner (PCOME)       1480
     Mixed Migration Monitoring Mechanism Initiative (4Mi)    1089
     Mixed Migration Monitoring Mechanism Initative (4mi)      992
     Mixed Migration Monitoring Mechanism Initiative (4mi)     673
                                                                …
     El Faro de Ceuta, Europa Press                              1
     Buzzfeed News, Al Jazeera                                   1
     Eagle Pass Texas News, Zócalo                               1
     Posto                                                       1
     El Nashra, Swiss Info; CGTN, IOM Morrocco                   1
     Name: Information Source, Length: 3803, dtype: int64
```

```
[ ]: Grouping=df['UNSD Geographical Grouping'].value_counts()
     Grouping
```

```
[ ]: Northern America      2708
     Uncategorized         2351
     Northern Africa       1872
     Southern Asia         1660
     Central America       1362
     Western Africa         941
     Eastern Africa         467
     Western Asia           389
     Southern Europe        329
     South-eastern Asia     225
     South America          207
     Western Europe         156
     Caribbean              113
     Eastern Europe         111
     Middle Africa           75
     Northern Europe         25
     Eastern Asia            15
     Southern Africa         12
     Central Asia             1
     Name: UNSD Geographical Grouping, dtype: int64
```

```
[ ]: plt.figure(figsize=(20,10))
     plt.barh(Grouping.index,Grouping,color='brown')
     plt.title("UNSD Geographical Grouping")
```

```
[ ]: Text(0.5, 1.0, 'UNSD Geographical Grouping')
```



### *Filling*

```
[ ]: df['Number of Dead'].unique()
```

```
[ ]: array([  1.,  12.,   5.,  15.,   2.,   8.,  11.,   7., 251.,  17.,  10.,
              4.,   0.,   6.,  22.,  44.,  13.,  62.,   3.,   9.,  45.,  29.,
             20., 170.,  18.,  24.,  42.,  64.,  70.,  41.,  27.,  21.,  16.,
            111.,  nan,  26., 750.,  14.,  36.,  47., 106.,  30., 100.,  40.,
             49.,  52.,  71.,  37.,  61.,  34.,  95.,  28.,  43.,  57.,  19.,
             23., 123.,  35.,  39.,  25.,  51., 133., 120., 204.,  97.,  87.,
             54.,  32.,  74.,  33.,  31.,  48.,  84.,  46.,  38.,  83.,  75.,
             53.,  55., 167.,  56.,  50., 117., 160.,  60.,  86.,  80.])
```

```
[ ]: ##Filling missing Values
     df['Number of Dead'].fillna(0, inplace=True)
```

```
[ ]: df['Region of Origin'].fillna('Unknown', inplace=True)
     df['Country of Origin'].fillna('Unknown', inplace=True)
     df['Migration route'].fillna('Unknown', inplace=True)
     df['Information Source'].fillna('Unknown', inplace=True)
     df['UNSD Geographical Grouping'].fillna('Unknown', inplace=True)
     df['Coordinates']=df['Coordinates'].fillna(df['Coordinates'].mode()[0])
```

```
[ ]: df.isna().sum()
```

```
[ ]: Incident Type                          0
     Incident year                          0
     Reported Month                         0
     Region of Origin                       0
     Region of Incident                     0
     Country of Origin                      0
     Number of Dead                         0
     Minimum Estimated Number of Missing    0
     Total Number of Dead and Missing       0
     Number of Survivors                    0
     Number of Females                      0
     Number of Males                        0
     Number of Children                     0
     Cause of Death                         0
     Migration route                        0
     Location of death                      0
     Information Source                     0
     Coordinates                            0
     UNSD Geographical Grouping             0
     dtype: int64
```

### ***Encoding***

```
[ ]: from sklearn.preprocessing import LabelEncoder
     le=LabelEncoder()
     df['Incident Type']=le.fit_transform(df['Incident Type'])
     df['Reported Month']=le.fit_transform(df['Reported Month'])
     df['Region of Origin']=le.fit_transform(df['Region of Origin'])
     df['Region of Incident']=le.fit_transform(df['Region of Incident'])
     df['Country of Origin']=le.fit_transform(df['Country of Origin'])
     df['Cause of Death']=le.fit_transform(df['Cause of Death'])
     df['Migration route']=le.fit_transform(df['Migration route'])
     df['Location of death']=le.fit_transform(df['Location of death'])
     df['Information Source']=le.fit_transform(df['Information Source'])
     df['UNSD Geographical Grouping']=le.fit_transform(df['UNSD Geographical␣
      ↪Grouping'])
```

```
[ ]: df[['Latitude', 'Longitude']] = df['Coordinates'].str.split(',', expand=True).
      ↪astype(float)
     df
```

```
[ ]:        Incident Type  Incident year  Reported Month  Region of Origin  \
     0                  1           2014               4                 2
     1                  1           2014               4                10
     2                  1           2014               4                10
     3                  1           2014               4                 2
     4                  1           2014               4                13
     …              …            …             …               …
     13015              1           2023               5                31
     13016              1           2023               5                30
     13017              1           2023               5                29
     13018              1           2023               5                12
     13019              1           2023               5                30

            Region of Incident  Country of Origin  Number of Dead  \
     0                       8                195             1.0
     1                       8                326             1.0
     2                       8                326             1.0
     3                       8                259             1.0
     4                       5                315             1.0
     …                   …                …              …
     13015                  15                318             4.0
     13016                  15                326             2.0
     13017                   9                303            13.0
     13018                   9                326             6.0
     13019                  14                326            16.0

            Minimum Estimated Number of Missing  Total Number of Dead and Missing  \
     0                                        0                                 1
     1                                        0                                 1
```

```
2                                        0                                 1
3                                        0                                 1
4                                        0                                 1
…                                    …                               …
13015                                    0                                 4
13016                                    0                                 2
13017                                    0                                13
13018                                    0                                 6
13019                                   37                                53

       Number of Survivors  …  Number of Males  Number of Children  \
0                        0  …                1                   0
1                        0  …                0                   0
2                        0  …                0                   0
3                        0  …                1                   0
4                        2  …                1                   0
…                    …  …              …                 …
13015                    0  …                4                   0
13016                    0  …                2                   0
13017                    6  …                0                   0
13018                   48  …                0                   0
13019                    2  …                0                   0

       Cause of Death  Migration route  Location of death  Information Source  \
0                  10               19               4562                2784
1                  10               19               4562                2784
2                  10               19               4562                2784
3                  14               19               7373                2521
4                   7               21               1362                 890
…               …              …                 …                 …
13015              13               18               2506                 277
13016              13               18                678                3127
13017               1               23               4281                 415
13018               1               23               7056                1114
13019               1               23               4165                1840

                   Coordinates UNSD Geographical Grouping   Latitude  \
0       31.650259, -110.366453                          8  31.650259
1         31.59713, -111.73756                          8  31.597130
2         31.94026, -113.01125                          8  31.940260
3       31.506777, -109.315632                          8  31.506777
4                  59.1551, 28                          9  59.155100
…                        …                        …        …
13015   40.91271268, 26.369657                         18  40.912713
13016   41.71697242, 26.351489                         18  41.716972
13017  23.72836078, -15.901632                         15  23.728361
13018   35.17187365, -2.903182                         15  35.171874
```

```
13019  14.71870705, -17.506255                        15  14.718707

        Longitude
0     -110.366453
1     -111.737560
2     -113.011250
3     -109.315632
4       28.000000
…             …
13015   26.369657
13016   26.351489
13017  -15.901632
13018   -2.903182
13019  -17.506255

[13020 rows x 21 columns]
```

```
[ ]: df.dtypes
```

```
[ ]: Incident Type                                int64
     Incident year                                int64
     Reported Month                               int64
     Region of Origin                             int64
     Region of Incident                           int64
     Country of Origin                            int64
     Number of Dead                             float64
     Minimum Estimated Number of Missing          int64
     Total Number of Dead and Missing             int64
     Number of Survivors                          int64
     Number of Females                            int64
     Number of Males                              int64
     Number of Children                           int64
     Cause of Death                               int64
     Migration route                              int64
     Location of death                            int64
     Information Source                           int64
     Coordinates                                 object
     UNSD Geographical Grouping                   int64
     Latitude                                   float64
     Longitude                                  float64
     dtype: object
```

### *Correlation*

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

```
<ipython-input-32-2f6f6606aa2c>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
```

```
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.corr()
```

[ ]:

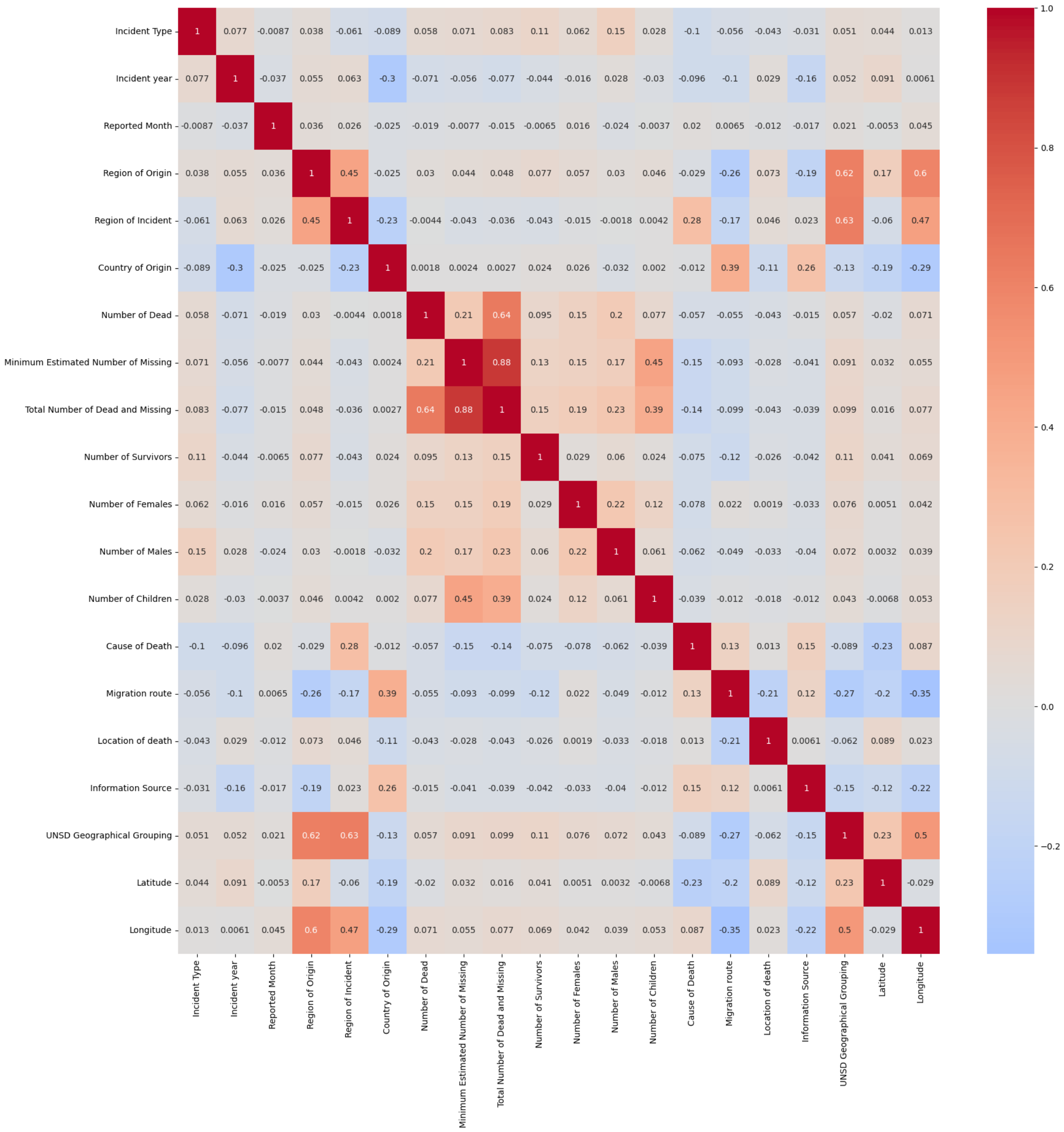
```
                                     Incident Type  Incident year  \
Incident Type                             1.000000       0.076628
Incident year                             0.076628       1.000000
Reported Month                           -0.008714      -0.036519
Region of Origin                          0.038305       0.055448
Region of Incident                       -0.060845       0.063038
Country of Origin                        -0.089466      -0.301706
Number of Dead                            0.058494      -0.070679
Minimum Estimated Number of Missing       0.070826      -0.055558
Total Number of Dead and Missing          0.083497      -0.077348
Number of Survivors                       0.109269      -0.044353
Number of Females                         0.062180      -0.015575
Number of Males                           0.148464       0.027849
Number of Children                        0.027818      -0.030011
Cause of Death                           -0.099913      -0.095503
Migration route                          -0.055576      -0.099958
Location of death                        -0.042855       0.029271
Information Source                       -0.030788      -0.159215
UNSD Geographical Grouping                0.050894       0.052416
Latitude                                  0.043708       0.090723
Longitude                                 0.013392       0.006074

                                     Reported Month  Region of Origin  \
Incident Type                             -0.008714          0.038305
Incident year                             -0.036519          0.055448
Reported Month                             1.000000          0.035904
Region of Origin                           0.035904          1.000000
Region of Incident                         0.026476          0.447338
Country of Origin                         -0.025246         -0.025367
Number of Dead                            -0.019174          0.029656
Minimum Estimated Number of Missing       -0.007721          0.043659
Total Number of Dead and Missing          -0.015219          0.048411
Number of Survivors                       -0.006513          0.076922
Number of Females                          0.015927          0.056673
Number of Males                           -0.024489          0.030489
Number of Children                        -0.003678          0.045628
Cause of Death                             0.020247         -0.028682
Migration route                            0.006460         -0.261433
Location of death                         -0.012376          0.072978
Information Source                        -0.016946         -0.187777
UNSD Geographical Grouping                 0.021245          0.617495
Latitude                                  -0.005279          0.172293
Longitude                                  0.044589          0.599155
```

```
                                     Region of Incident  Country of Origin  \
Incident Type                                 -0.060845          -0.089466   
Incident year                                  0.063038          -0.301706   
Reported Month                                 0.026476          -0.025246   
Region of Origin                               0.447338          -0.025367   
Region of Incident                             1.000000          -0.226761   
Country of Origin                             -0.226761           1.000000   
Number of Dead                                -0.004398           0.001795   
Minimum Estimated Number of Missing           -0.043498           0.002398   
Total Number of Dead and Missing              -0.036213           0.002738   
Number of Survivors                           -0.042863           0.024192   
Number of Females                             -0.014861           0.026061   
Number of Males                               -0.001823          -0.032380   
Number of Children                             0.004213           0.001961   
Cause of Death                                 0.283421          -0.011881   
Migration route                               -0.168905           0.385098   
Location of death                              0.045629          -0.110229   
Information Source                             0.023146           0.264236   
UNSD Geographical Grouping                     0.628374          -0.130228   
Latitude                                      -0.060237          -0.187679   
Longitude                                      0.469840          -0.294024   

                                     Number of Dead  \
Incident Type                              0.058494   
Incident year                             -0.070679   
Reported Month                            -0.019174   
Region of Origin                           0.029656   
Region of Incident                        -0.004398   
Country of Origin                          0.001795   
Number of Dead                             1.000000   
Minimum Estimated Number of Missing        0.208926   
Total Number of Dead and Missing           0.641773   
Number of Survivors                        0.094863   
Number of Females                          0.147181   
Number of Males                            0.199589   
Number of Children                         0.077084   
Cause of Death                            -0.056879   
Migration route                           -0.055341   
Location of death                         -0.042649   
Information Source                        -0.015144   
UNSD Geographical Grouping                 0.057422   
Latitude                                  -0.019662   
Longitude                                  0.071367   

                                     Minimum Estimated Number of Missing  \
Incident Type                                                   0.070826   
Incident year                                                  -0.055558
Reported Month                                                 -0.007721
Region of Origin                                                0.043659
Region of Incident                                             -0.043498
Country of Origin                                               0.002398
Number of Dead                                                  0.208926
Minimum Estimated Number of Missing                             1.000000
Total Number of Dead and Missing                                0.884053
Number of Survivors                                             0.131206
Number of Females                                               0.147635
Number of Males                                                 0.172177
Number of Children                                              0.453512
Cause of Death                                                 -0.147592
Migration route                                                -0.092862
Location of death                                              -0.028315
Information Source                                             -0.040671
UNSD Geographical Grouping                                      0.090670
Latitude                                                        0.032465
Longitude                                                       0.054571

                                     Total Number of Dead and Missing  \
Incident Type                                                0.083497
Incident year                                               -0.077348
Reported Month                                              -0.015219
Region of Origin                                             0.048411
Region of Incident                                          -0.036213
Country of Origin                                            0.002738
Number of Dead                                               0.641773
Minimum Estimated Number of Missing                          0.884053
Total Number of Dead and Missing                             1.000000
Number of Survivors                                          0.148230
Number of Females                                            0.186118
Number of Males                                              0.230411
Number of Children                                           0.392485
Cause of Death                                              -0.142927
Migration route                                             -0.099272
Location of death                                           -0.042588
Information Source                                          -0.039132
UNSD Geographical Grouping                                   0.098547
Latitude                                                     0.016062
Longitude                                                    0.076903

                                     Number of Survivors  Number of Females  \
Incident Type                                   0.109269           0.062180
Incident year                                  -0.044353          -0.015575
Reported Month                                 -0.006513           0.015927
Region of Origin                                0.076922           0.056673
Region of Incident                             -0.042863          -0.014861
Country of Origin                               0.024192           0.026061
Number of Dead                                  0.094863           0.147181
Minimum Estimated Number of Missing             0.131206           0.147635
Total Number of Dead and Missing                0.148230           0.186118
Number of Survivors                             1.000000           0.029244
Number of Females                               0.029244           1.000000
Number of Males                                 0.059640           0.217927
Number of Children                              0.024462           0.120186
Cause of Death                                 -0.074721          -0.078089
Migration route                                -0.121244           0.021933
Location of death                              -0.026061           0.001855
Information Source                             -0.041555          -0.033250
UNSD Geographical Grouping                      0.107444           0.075876
Latitude                                        0.040866           0.005138
Longitude                                       0.068794           0.041674

                                     Number of Males  Number of Children  \
Incident Type                               0.148464            0.027818
Incident year                               0.027849           -0.030011
Reported Month                             -0.024489           -0.003678
Region of Origin                            0.030489            0.045628
Region of Incident                         -0.001823            0.004213
Country of Origin                          -0.032380            0.001961
Number of Dead                              0.199589            0.077084
Minimum Estimated Number of Missing         0.172177            0.453512
Total Number of Dead and Missing            0.230411            0.392485
Number of Survivors                         0.059640            0.024462
Number of Females                           0.217927            0.120186
Number of Males                             1.000000            0.061169
Number of Children                          0.061169            1.000000
Cause of Death                             -0.062073           -0.039434
Migration route                            -0.049471           -0.011727
Location of death                          -0.033049           -0.017800
Information Source                         -0.039937           -0.012138
UNSD Geographical Grouping                  0.071890            0.043456
Latitude                                    0.003240           -0.006756
Longitude                                   0.038784            0.053277

                                     Cause of Death  Migration route  \
Incident Type                             -0.099913        -0.055576
Incident year                             -0.095503        -0.099958
Reported Month                             0.020247         0.006460
Region of Origin                          -0.028682        -0.261433
Region of Incident                         0.283421        -0.168905
Country of Origin                         -0.011881         0.385098
Number of Dead                            -0.056879        -0.055341
```

```
Minimum Estimated Number of Missing       -0.147592        -0.092862
Total Number of Dead and Missing          -0.142927        -0.099272
Number of Survivors                       -0.074721        -0.121244
Number of Females                         -0.078089         0.021933
Number of Males                           -0.062073        -0.049471
Number of Children                        -0.039434        -0.011727
Cause of Death                             1.000000         0.133395
Migration route                            0.133395         1.000000
Location of death                          0.013312        -0.213810
Information Source                         0.148667         0.120960
UNSD Geographical Grouping                -0.088946        -0.270369
Latitude                                  -0.233722        -0.203507
Longitude                                  0.087025        -0.354839

                                     Location of death  Information Source  \
Incident Type                                -0.042855           -0.030788
Incident year                                 0.029271           -0.159215
Reported Month                               -0.012376           -0.016946
Region of Origin                              0.072978           -0.187777
Region of Incident                            0.045629            0.023146
Country of Origin                            -0.110229            0.264236
Number of Dead                               -0.042649           -0.015144
Minimum Estimated Number of Missing          -0.028315           -0.040671
Total Number of Dead and Missing             -0.042588           -0.039132
Number of Survivors                          -0.026061           -0.041555
Number of Females                             0.001855           -0.033250
Number of Males                              -0.033049           -0.039937
Number of Children                           -0.017800           -0.012138
Cause of Death                                0.013312            0.148667
Migration route                              -0.213810            0.120960
Location of death                             1.000000            0.006075
Information Source                            0.006075            1.000000
UNSD Geographical Grouping                   -0.061893           -0.151213
Latitude                                      0.088876           -0.120756
Longitude                                     0.023146           -0.215984

                                     UNSD Geographical Grouping  Latitude  \
Incident Type                                          0.050894  0.043708
Incident year                                          0.052416  0.090723
Reported Month                                         0.021245 -0.005279
Region of Origin                                       0.617495  0.172293
Region of Incident                                     0.628374 -0.060237
Country of Origin                                     -0.130228 -0.187679
Number of Dead                                         0.057422 -0.019662
Minimum Estimated Number of Missing                    0.090670  0.032465
Total Number of Dead and Missing                       0.098547  0.016062
Number of Survivors                                    0.107444  0.040866
Number of Females                                        0.075876  0.005138
Number of Males                                          0.071890  0.003240
Number of Children                                       0.043456 -0.006756
Cause of Death                                          -0.088946 -0.233722
Migration route                                         -0.270369 -0.203507
Location of death                                       -0.061893  0.088876
Information Source                                      -0.151213 -0.120756
UNSD Geographical Grouping                               1.000000  0.231927
Latitude                                                 0.231927  1.000000
Longitude                                                0.504979 -0.028584

                                     Longitude
Incident Type                         0.013392
Incident year                         0.006074
Reported Month                        0.044589
Region of Origin                      0.599155
Region of Incident                    0.469840
Country of Origin                    -0.294024
Number of Dead                        0.071367
Minimum Estimated Number of Missing   0.054571
Total Number of Dead and Missing      0.076903
Number of Survivors                   0.068794
Number of Females                     0.041674
Number of Males                       0.038784
Number of Children                    0.053277
Cause of Death                        0.087025
Migration route                      -0.354839
Location of death                     0.023146
Information Source                   -0.215984
UNSD Geographical Grouping            0.504979
Latitude                             -0.028584
Longitude                             1.000000
```

[ ]:
```
Corr_Matrix = df.corr()

# Set up the figure and plot the heatmap
plt.figure(figsize=(20, 20))
sns.heatmap(Corr_Matrix, annot=True, cmap='coolwarm', center=0)
plt.show()
```

```
<ipython-input-33-e7c2484fc002>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  Corr_Matrix = df.corr()
```



***Top 5 most positively Correlated***

```
[ ]: print('Top 5 Most Positively Correlated to the Total Number of Dead and␣
     ↪Missing')
     Corr_Matrix['Total Number of Dead and Missing'].sort_values(ascending=False).
     ↪head(5)
```

```
Top 5 Most Positively Correlated to the Total Number of Dead and Missing
```

```
[ ]: Total Number of Dead and Missing       1.000000
     Minimum Estimated Number of Missing    0.884053
```

```
Number of Dead                             0.641773
Number of Children                         0.392485
Number of Males                            0.230411
Name: Total Number of Dead and Missing, dtype: float64
```

***Top 5 most Negatively Correlated***

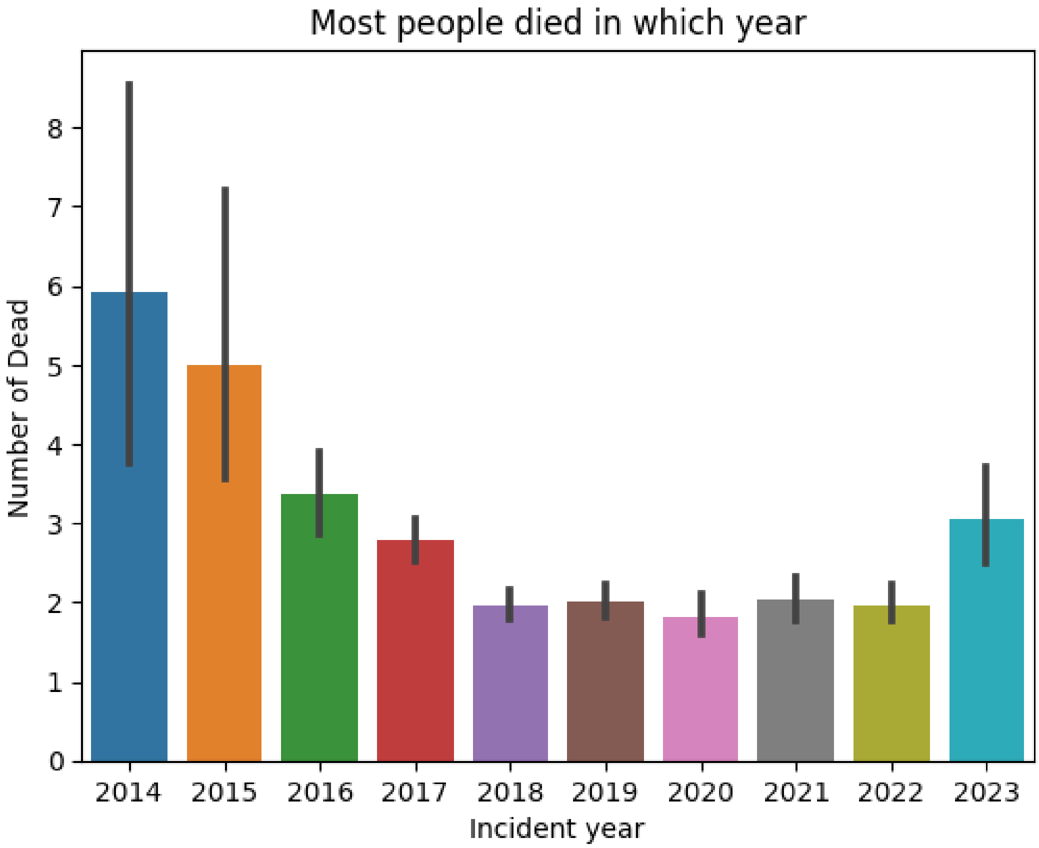
```
[ ]: print('Top 5 Most Negatively Correlated to Total Number of Dead and Missing')
     Corr_Matrix['Total Number of Dead and Missing'].sort_values(ascending=True).
      ↪head(5)
```

```
Top 5 Most Negatively Correlated to Total Number of Dead and Missing
```

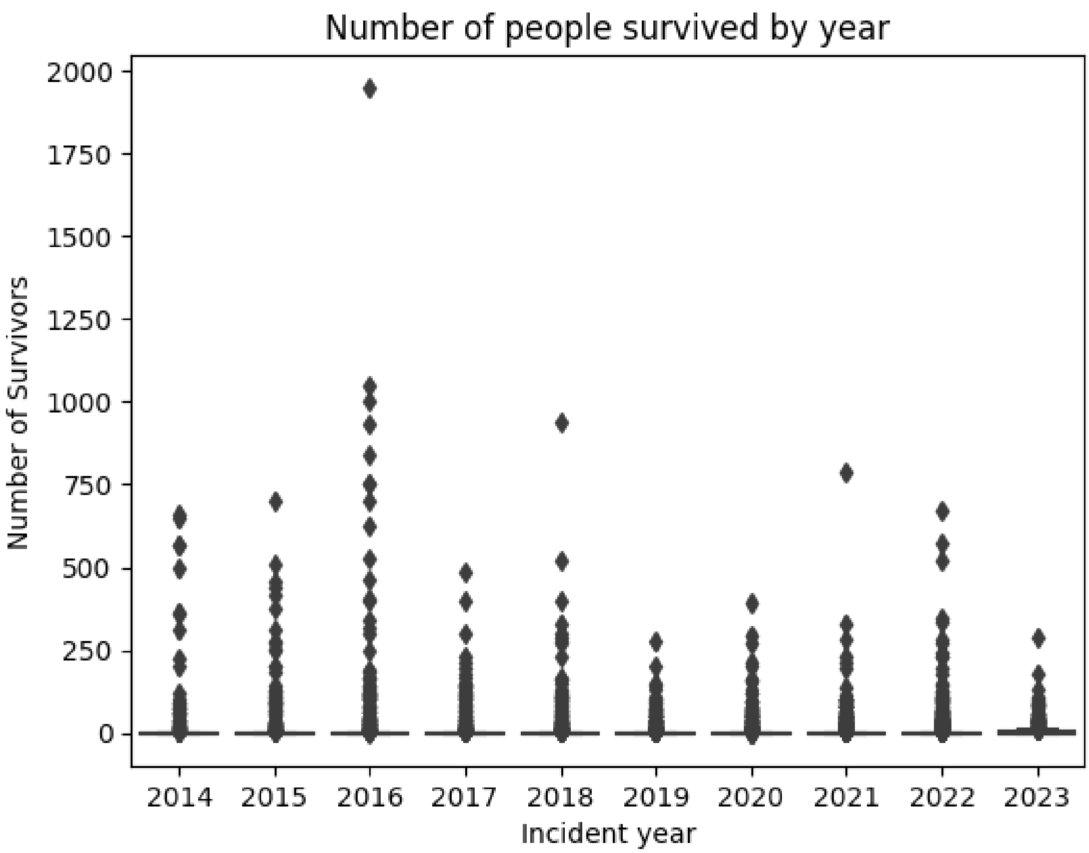
```
[ ]: Cause of Death       -0.142927
     Migration route      -0.099272
     Incident year        -0.077348
     Location of death    -0.042588
     Information Source   -0.039132
     Name: Total Number of Dead and Missing, dtype: float64
```

***Exploratory analysis & Visualization***

```
[ ]: sns.barplot(data= df, x='Incident year',y= 'Number of Dead')
     plt.title('Most people died in which year');
```



```
[ ]: sns.boxplot(data = df, x= 'Incident year', y = 'Number of Survivors')
     plt.title('Number of people survived by year');
```



***Dropping unwanted columns and rows***

```
[ ]: # dropping Coordinates  along with unwanted columns found by correlation matrix
     df=df.drop(['Coordinates','Country of Origin'],axis=1)
     df
```

```
[ ]:        Incident Type  Incident year  Reported Month  Region of Origin  \
     0                  1           2014               4                 2
     1                  1           2014               4                10
     2                  1           2014               4                10
     3                  1           2014               4                 2
     4                  1           2014               4                13
     …              …             …              …                …
     13015              1           2023               5                31
     13016              1           2023               5                30
     13017              1           2023               5                29
     13018              1           2023               5                12
     13019              1           2023               5                30

            Region of Incident  Number of Dead  \
     0                       8             1.0
     1                       8             1.0
     2                       8             1.0
     3                       8             1.0
     4                       5             1.0
     …                   …             …
     13015                  15             4.0
     13016                  15             2.0
     13017                   9            13.0
     13018                   9             6.0
     13019                  14            16.0

            Minimum Estimated Number of Missing  Total Number of Dead and Missing  \
     0                                        0                                 1
     1                                        0                                 1
     2                                        0                                 1
     3                                        0                                 1
     4                                        0                                 1
     …                                    …                               …
     13015                                    0                                 4
     13016                                    0                                 2
     13017                                    0                                13
     13018                                    0                                 6
     13019                                   37                                53

            Number of Survivors  Number of Females  Number of Males  \
     0                        0                  0                1
     1                        0                  0                0
     2                        0                  0                0
     3                        0                  0                1
     4                        2                  0                1
     …                    …                …              …
     13015                    0                  0                4
     13016                    0                  0                2
     13017                    6                  0                0
     13018                   48                  0                0
     13019                    2                  2                0

            Number of Children  Cause of Death  Migration route  Location of death  \
     0                       0              10               19               4562
     1                       0              10               19               4562
     2                       0              10               19               4562
     3                       0              14               19               7373
     4                       0               7               21               1362
     …                   …             …              …                …
     13015                   0              13               18               2506
     13016                   0              13               18                678
```

```
13017                   0               1               23               4281
13018                   0               1               23               7056
13019                   0               1               23               4165

       Information Source  UNSD Geographical Grouping   Latitude   Longitude
0                    2784                           8  31.650259 -110.366453
1                    2784                           8  31.597130 -111.737560
2                    2784                           8  31.940260 -113.011250
3                    2521                           8  31.506777 -109.315632
4                     890                           9  59.155100   28.000000
…                   …                         …        …         …
13015                 277                          18  40.912713   26.369657
13016                3127                          18  41.716972   26.351489
13017                 415                          15  23.728361  -15.901632
13018                1114                          15  35.171874   -2.903182
13019                1840                          15  14.718707  -17.506255

[13020 rows x 19 columns]
```

#### ***Finding Duplicates***

```
[ ]: df.duplicated().sum()
```

```
[ ]: 644
```

```
[ ]: # dropping duplicate rows
     df.drop_duplicates(keep='first', inplace=True)
     df
```

```
[ ]:        Incident Type  Incident year  Reported Month  Region of Origin  \
     0                  1           2014               4                 2
     1                  1           2014               4                10
     2                  1           2014               4                10
     3                  1           2014               4                 2
     4                  1           2014               4                13
     …              …            …             …             …
     13015              1           2023               5                31
     13016              1           2023               5                30
     13017              1           2023               5                29
     13018              1           2023               5                12
     13019              1           2023               5                30

            Region of Incident  Number of Dead  \
     0                       8             1.0
     1                       8             1.0
     2                       8             1.0
     3                       8             1.0
```

```
4                        5             1.0
…                   …             …
13015                   15             4.0
13016                   15             2.0
13017                    9            13.0
13018                    9             6.0
13019                   14            16.0

       Minimum Estimated Number of Missing  Total Number of Dead and Missing  \
0                                        0                                 1
1                                        0                                 1
2                                        0                                 1
3                                        0                                 1
4                                        0                                 1
…                                    …                               …
13015                                    0                                 4
13016                                    0                                 2
13017                                    0                                13
13018                                    0                                 6
13019                                   37                                53

       Number of Survivors  Number of Females  Number of Males  \
0                        0                  0                1
1                        0                  0                0
2                        0                  0                0
3                        0                  0                1
4                        2                  0                1
…                    …                …              …
13015                    0                  0                4
13016                    0                  0                2
13017                    6                  0                0
13018                   48                  0                0
13019                    2                  2                0

       Number of Children  Cause of Death  Migration route  Location of death  \
0                       0              10               19               4562
1                       0              10               19               4562
2                       0              10               19               4562
3                       0              14               19               7373
4                       0               7               21               1362
…                   …             …               …                …
13015                   0              13               18               2506
13016                   0              13               18                678
13017                   0               1               23               4281
13018                   0               1               23               7056
13019                   0               1               23               4165
```

```
       Information Source  UNSD Geographical Grouping   Latitude   Longitude
0                    2784                           8  31.650259 -110.366453
1                    2784                           8  31.597130 -111.737560
2                    2784                           8  31.940260 -113.011250
3                    2521                           8  31.506777 -109.315632
4                     890                           9  59.155100   28.000000
…                       …                         …          …         …
13015                 277                          18  40.912713   26.369657
13016                3127                          18  41.716972   26.351489
13017                 415                          15  23.728361  -15.901632
13018                1114                          15  35.171874   -2.903182
13019                1840                          15  14.718707  -17.506255

[12376 rows x 19 columns]
```

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12376 entries, 0 to 13019
Data columns (total 19 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Incident Type                      12376 non-null  int64
 1   Incident year                      12376 non-null  int64
 2   Reported Month                     12376 non-null  int64
 3   Region of Origin                   12376 non-null  int64
 4   Region of Incident                 12376 non-null  int64
 5   Number of Dead                     12376 non-null  float64
 6   Minimum Estimated Number of Missing  12376 non-null  int64
 7   Total Number of Dead and Missing   12376 non-null  int64
 8   Number of Survivors                12376 non-null  int64
 9   Number of Females                  12376 non-null  int64
 10  Number of Males                    12376 non-null  int64
 11  Number of Children                 12376 non-null  int64
 12  Cause of Death                     12376 non-null  int64
 13  Migration route                    12376 non-null  int64
 14  Location of death                  12376 non-null  int64
 15  Information Source                 12376 non-null  int64
 16  UNSD Geographical Grouping         12376 non-null  int64
 17  Latitude                           12376 non-null  float64
 18  Longitude                          12376 non-null  float64
dtypes: float64(3), int64(16)
memory usage: 1.9 MB
```

*Importing Libraries*

```
[ ]: from sklearn.model_selection import train_test_split
     from sklearn.linear_model import LinearRegression
     from sklearn.ensemble import RandomForestRegressor
     from sklearn.metrics import r2_score, mean_absolute_error,␣
      ↪mean_squared_error,mean_absolute_percentage_error
```

*Splitting the Dataset*

```
[ ]: x = df.drop(columns=['Total Number of Dead and Missing'])
     y = df['Total Number of Dead and Missing']

     x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,␣
      ↪random_state=42)
     # Display the shapes of the resulting datasets
     print("X_train shape:", x_train.shape)
     print("X_test shape:", x_test.shape)
     print("y_train shape:", y_train.shape)
     print("y_test shape:", y_test.shape)
```

```
X_train shape: (9900, 18)
X_test shape: (2476, 18)
y_train shape: (9900,)
y_test shape: (2476,)
```

*Model Building and Analysis*

```
[ ]: models = {
         'Random Forest': RandomForestRegressor(random_state=42),
         'Linear Regression':LinearRegression()
     }
     best_model = None
     best_r2 = 0

     for model_name, model in models.items():
         model.fit(x_train, y_train)
         y_pred= model.predict(x_test)

         # Evaluate the model
         r2 = r2_score(y_test, y_pred)
         mape=mean_absolute_percentage_error(y_test,y_pred)
         mae = mean_absolute_error(y_test, y_pred)
         mse = mean_squared_error(y_test, y_pred)
         submit = pd.DataFrame()
         submit['Actual Number of Dead'] = y_test
         submit['Predict_Number of Dead'] = y_pred
         submit = submit.reset_index()
         r2 = r2_score(y_test, y_pred)
         if r2 > best_r2:
             best_r2 = r2
             best_model = model_name
```

```
    print(f'{model_name}:')                 # the f-string formatting is used to␣
 ↪embed variables (r2, mape, mae, mse) directly into the strings.
    print(f'R2 Score: {r2:.2f}')
    print(f'Mean Absolute Percentage Error(MAPE):{mape:.2f}')
    print(f'Mean Absolute Error (MAE): {mae:.2f}')
    print(f'Mean Squared Error (MSE): {mse:.2f}')          # 2f ==>till 2␣
 ↪decimals value will be displayed
    print(submit.head(5))

    print('----------------------------------------')
print(f"The best performing model is: {best_model} with accuracy: {best_r2:.
 ↪2f}")
```

```
Random Forest:
R2 Score: 0.95
Mean Absolute Percentage Error(MAPE):0.00
Mean Absolute Error (MAE): 0.23
Mean Squared Error (MSE): 17.38
   index  Actual Number of Dead  Predict_Number of Dead
0   3794                      1                    1.00
1   5020                      6                    6.11
2   7631                      1                    1.00
3   3234                      2                    2.00
4   4528                    114                  114.38
----------------------------------------
Linear Regression:
R2 Score: 1.00
Mean Absolute Percentage Error(MAPE):0.00
Mean Absolute Error (MAE): 0.00
Mean Squared Error (MSE): 0.00
   index  Actual Number of Dead  Predict_Number of Dead
0   3794                      1                     1.0
1   5020                      6                     6.0
2   7631                      1                     1.0
3   3234                      2                     2.0
4   4528                    114                   114.0
----------------------------------------
The best performing model is: Linear Regression with accuracy: 1.00
```

### *Conclusion*

Best Model:

The Linear Regression model appears to outperform the Random Forest model based on the evaluation metrics provided. It achieved perfect predictions on the test data, indicating an exact match between predicted and actual values. The Linear Regression model has an R2 score of 1.00, meaning it explains all the variance in the target variable based on the features.

Overfitting:

The perfect performance of the Linear Regression model (R2 score of 1.00) on the test set might suggest overfitting, especially when the training and testing datasets are the same. On the other hand, the Random Forest model, although slightly lower in R2 score (0.95), exhibits a reasonable level of performance without perfect accuracy on the test set. This might suggest that it is not overfitting as much as the Linear Regression model.