Global Immigration

The Human Cost of Migration & how to use ML to Save Lives

By Karen Guerrero & Amani Sakif

Data cleaning and feature engineering

```
#Delete rows with all NaN values

conflicts_data.dropna(how='all')
conflicts_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5016 entries, 0 to 5015
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   gwno                    5016 non-null   int64
 1   country                 5016 non-null   object
 2   year                    5016 non-null   int64
 3   ongoing                 4988 non-null   float64
 4   gwarea_tot              3203 non-null   float64
 5   gwsum_bestdeaths        3203 non-null   float64
 6   gwsum_best_sb           3203 non-null   float64
 7   gwshare_confl           449 non-null    float64
 8   gwpop                   3203 non-null   float64
 9   gwpop_confl             3203 non-null   float64
 10  gwpopshare_confl        449 non-null    float64
 11  maxintensity            807 non-null    float64
 12  maxcumulativeintensity  659 non-null    float64
 13  peaceyears              4225 non-null   float64
 14  peaceyearshigh          4225 non-null   float64
 15  postconflict            4174 non-null   float64
 16  pop_affected            449 non-null    float64
 17  area_affected           449 non-null    float64
dtypes: float64(15), int64(2), object(1)
memory usage: 705.5+ KB
```

```
migrants_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6888 entries, 0 to 6887
Data columns (total 21 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Geographic area                                         6888 non-null   object
 1   Indicator                                               6888 non-null   object
 2   Sex                                                     6888 non-null   object
 3   TIME_PERIOD                                             6888 non-null   int64
 4   OBS_VALUE                                               6888 non-null   object
 5   Unit multiplier                                         4342 non-null   object
 6   Unit of measure                                         6888 non-null   object
 7   Observation Status                                      0 non-null      float64
 8   Observation confidentaility                             0 non-null      float64
 9   LOWER_BOUND                                             0 non-null      float64
 10  UPPER_BOUND                                             0 non-null      float64
 11  WGTD_SAMPL_SIZE                                         0 non-null      float64
 12  OBS_FOOTNOTE                                            6888 non-null   object
 13  SERIES_FOOTNOTE                                         0 non-null      float64
 14  DATA_SOURCE                                             6888 non-null   object
 15  SOURCE_LINK                                             6888 non-null   object
 16  CUSTODIAN                                               0 non-null      float64
 17  Time period activity related to when the data are collected  0 non-null  float64
 18  REF_PERIOD                                              0 non-null      float64
 19  COVERAGE_TIME                                           0 non-null      float64
 20  Current age                                             6888 non-null   object
dtypes: float64(10), int64(1), object(10)
memory usage: 1.1+ MB
```

```
# Drop rows with null values and check the values

migrant_deaths_data.dropna(how='all')
migrant_deaths_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16149 entries, 0 to 16148
Data columns (total 25 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Main ID                               16149 non-null  object
 1   Incident ID                           16149 non-null  object
 2   Incident Type                         16149 non-null  object
 3   Region of Incident                    16149 non-null  object
 4   Incident Date                         16134 non-null  object
 5   Incident Year                         16149 non-null  int64
 6   Month                                 16149 non-null  object
 7   Number of Dead                        15340 non-null  float64
 8   Minimum Estimated Number of Missing   1611 non-null   float64
 9   Total Number of Dead and Missing      16149 non-null  int64
 10  Number of Survivors                   2586 non-null   float64
 11  Number of Females                     4016 non-null   float64
 12  Number of Males                       10314 non-null  float64
 13  Number of Children                    2846 non-null   float64
 14  Country of Origin                     16141 non-null  object
 15  Region of Origin                      16148 non-null  object
 16  Cause of Death                        16149 non-null  object
 17  Country of Incident                   16149 non-null  object
 18  Migration Route                       13634 non-null  object
 19  Location of Incident                  16149 non-null  object
 20  Coordinates                           16148 non-null  object
 21  UNSD Geographical Grouping            16148 non-null  object
 22  Information Source                    16136 non-null  object
 23  URL                                   10421 non-null  object
 24  Source Quality                        16149 non-null  int64
dtypes: float64(6), int64(3), object(16)
memory usage: 3.1+ MB
```

Exploratory Data Analysis (EDA) Steps:

Data Inspection:

- Viewing the first few rows of the dataset using .head().
- Checking the data types and non-null counts with .info().
- Descriptive statistics using .describe().

Data Visualization:

- Creating various plots such as scatter plots, bar charts, pie charts, etc.

Data Aggregation and Grouping:

- Grouping data by specific columns and aggregating counts or other statistics.

Value Counts:

- Checking the frequency of different categories in specific columns.

Data Cleaning Steps:

Handling Missing Values:

- Deleting rows with all NaN values using dropna(how='all').
- Replacing NaN values in numeric columns with the mean, except for specified columns like 'ongoing', 'maxintensity', and 'maxcumulativeintensity'.

Data Transformation:

- Preparing data for clustering by selecting relevant features and dropping rows with NaN values in those features.

Data Formatting:

- Ensuring consistent data types and formats across columns, especially after transformations and filling NaN values.

Grouping and Aggregation:

- Calculating conflicts per year and country, and pivoting the DataFrame for further analysis.

Difference Between Emigrate, Immigrate and Migrate

- *Emigrate* is from the point of view of the departure. **Think *exit.***
- *Immigrate* is from the point of view of the destination. **Think *come in.***
- *Migrate* is all about the moving. **Think *move.***

Refugees and Asylum Seekers

Who is a refugee?

A refugee is a person who has fled their own country because they are at risk of serious human rights violations and persecution there. The risks to their safety and life were so great that they felt they had no choice but to leave and seek safety outside their country because their own government cannot or will not protect them from those dangers.

Who is an asylum seeker?

An asylum seeker is a person who has left their country and is seeking protection from persecution and serious human rights violations in another country, but who hasn't yet been legally recognized as a refugee and is waiting to receive a decision on their asylum claim.

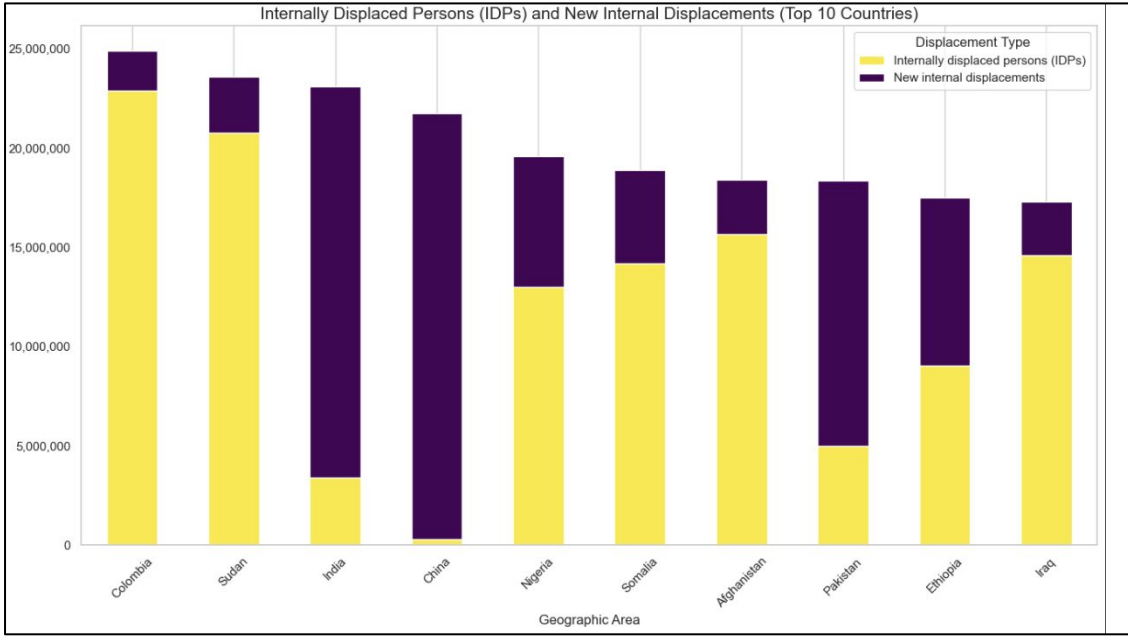
Internally Displaced Persons (IDPs):

- IDPs are individuals who have fled their homes due to conflict, violence, or disasters but remain within their own country.

New Internal Displacements:

- The number of new instances of displacement within a country over a specific period, typically due to conflict, violence, or disasters.

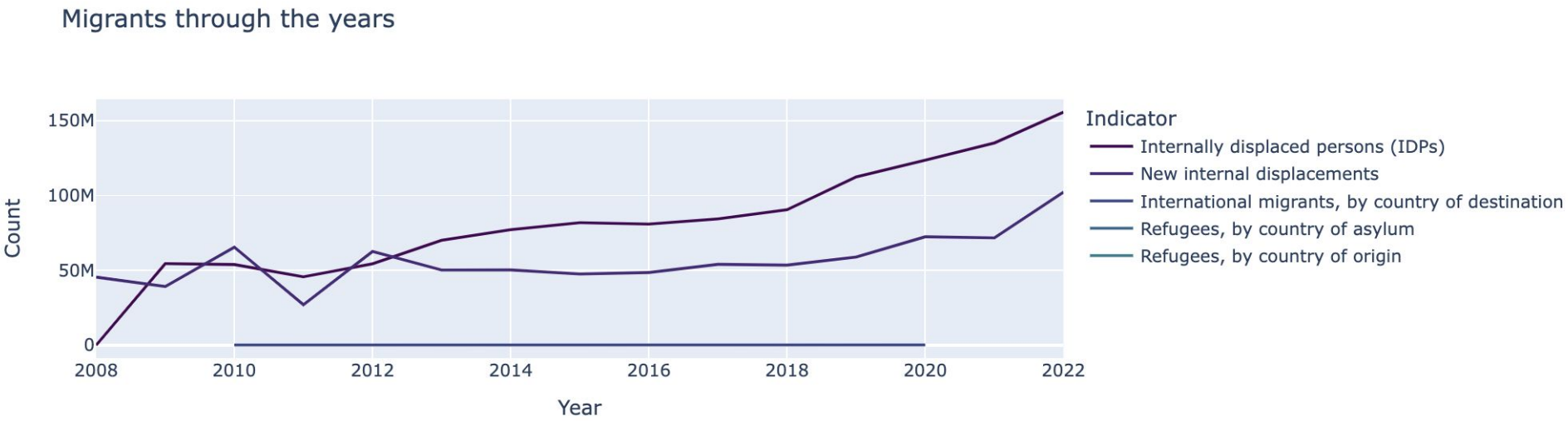
```python
from matplotlib.ticker import FuncFormatter

# Set 'Geographic area' as the index
final_df.set_index('Geographic area', inplace=True)

# Columns to visualize
columns_to_visualize = ['Internally displaced persons (IDPs)', 'New internal displacements']

# Calculate total displacements and select top 10 countries
final_df['Total Displacements'] = final_df[columns_to_visualize].sum(axis=1)
top_10_countries = final_df.nlargest(10, 'Total Displacements')

# Define a formatter function to remove scientific notation e
def plain_formatter(x, pos):
    return f'{int(x):,}'

# Get the reversed viridis colormap
reversed_viridis = plt.cm.viridis(np.linspace(0, 1, len(columns_to_visualize))[::-1])

# Plot the stacked bar plot
top_10_countries[columns_to_visualize].plot(kind='bar', stacked=True, figsize=(14, 8),
                                            color=reversed_viridis)

plt.title('Internally Displaced Persons (IDPs) and New Internal Displacements (Top 10 Countries)'
plt.xlabel('Geographic Area', fontsize=12)
plt.ylabel('Number of Persons', fontsize=12)
plt.legend(title='Displacement Type')
plt.grid(axis='y')
plt.gca().yaxis.set_major_formatter(FuncFormatter(plain_formatter))
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



Immigration Analysis

Installation (US-Mexican border) by Ronald Rael and Virginia San Fratello

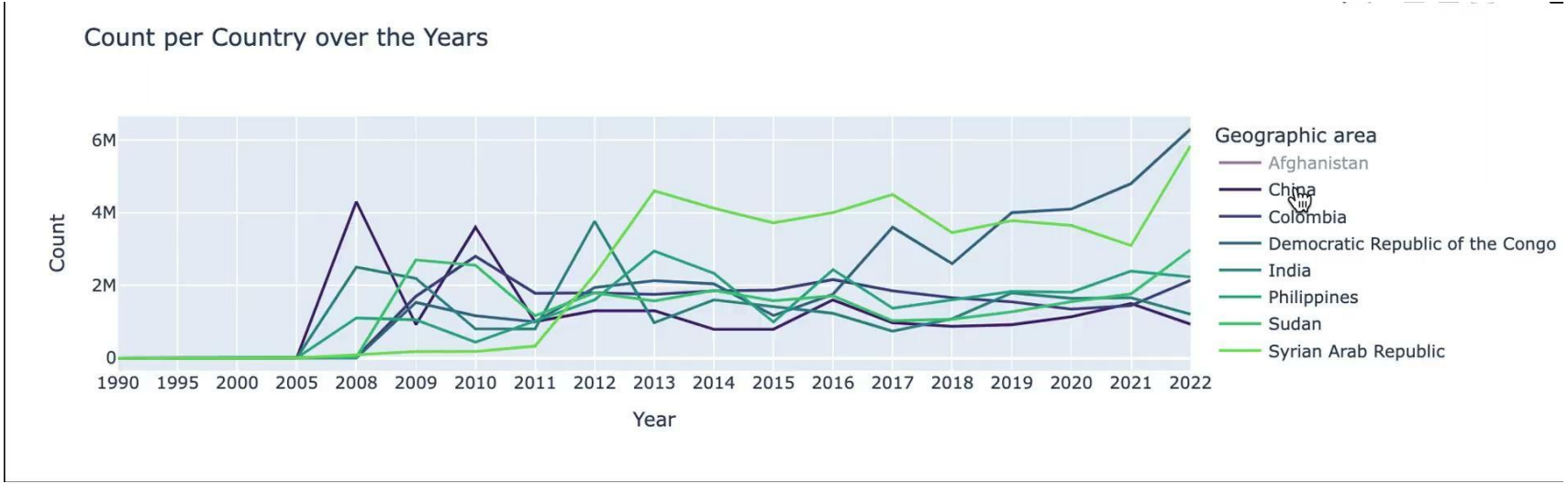
Migrants per Year



Observations:

- **Post-2014 Increase**: of IDPs and international migrants after 2014, likely due to major global conflicts and economic migrations.
- **Consistent Growth**: of new internal displacements and international migrants highlights ongoing global instability and economic opportunities driving migration.
- **Impact of Global Events**: Peaks in the refugee data correspond with major global events, such as conflicts in Syria, the Middle East, and parts of Africa.

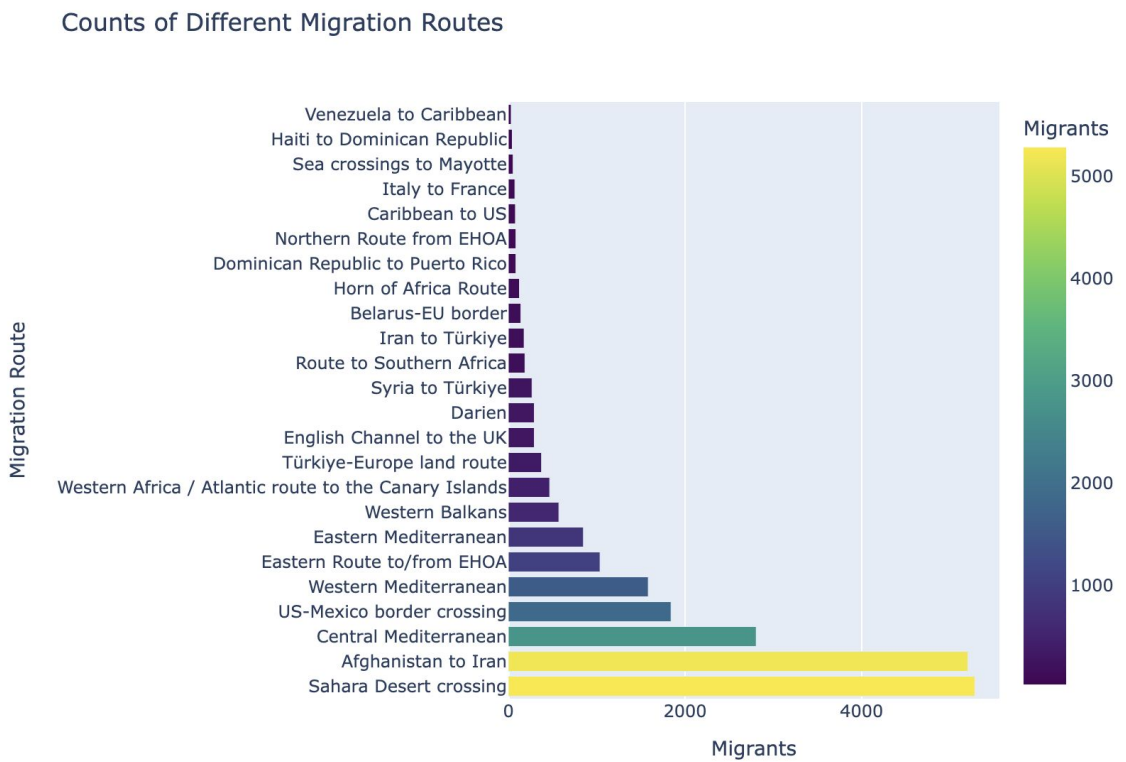
Migrants per Country of Origin

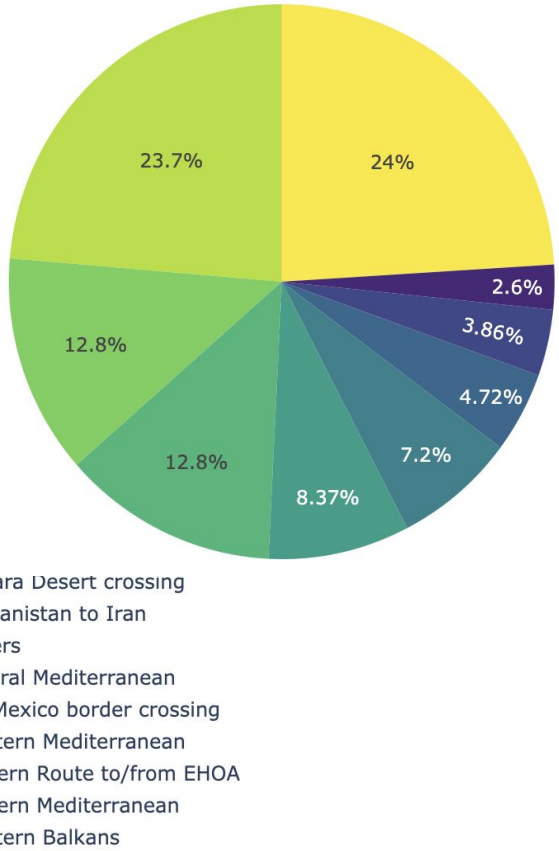


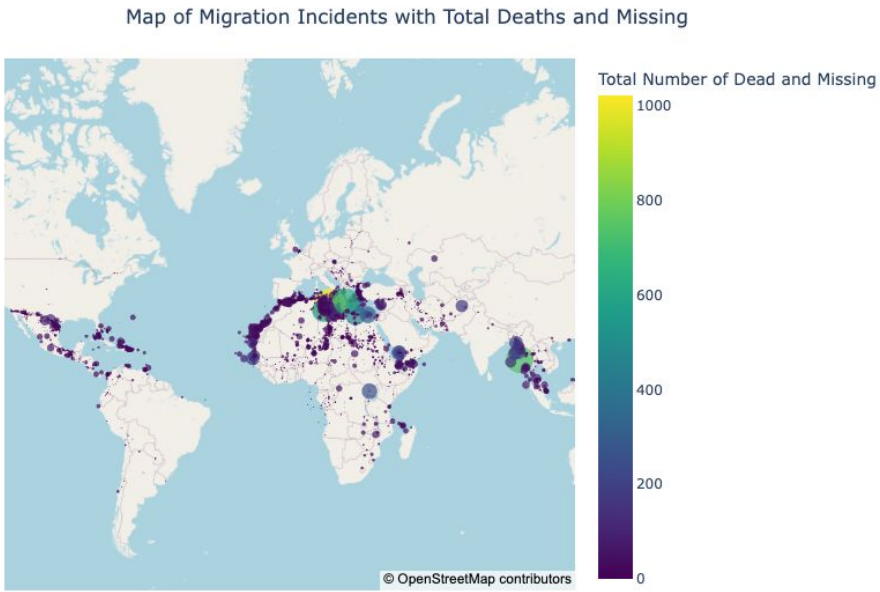
Observations:

- **Post-2010 Spike**: migrants from Afghanistan and Syria after 2010, corresponding with escalating conflicts.
- **Economic Migration**: Steady upward trends in migrants from India and the Philippines, driven by sustained economic migration.
- **Conflict-Driven Migration**: Peaks in migration from Colombia and the Democratic Republic of the Congo during periods of heightened conflict.

Most Popular Migration Routes



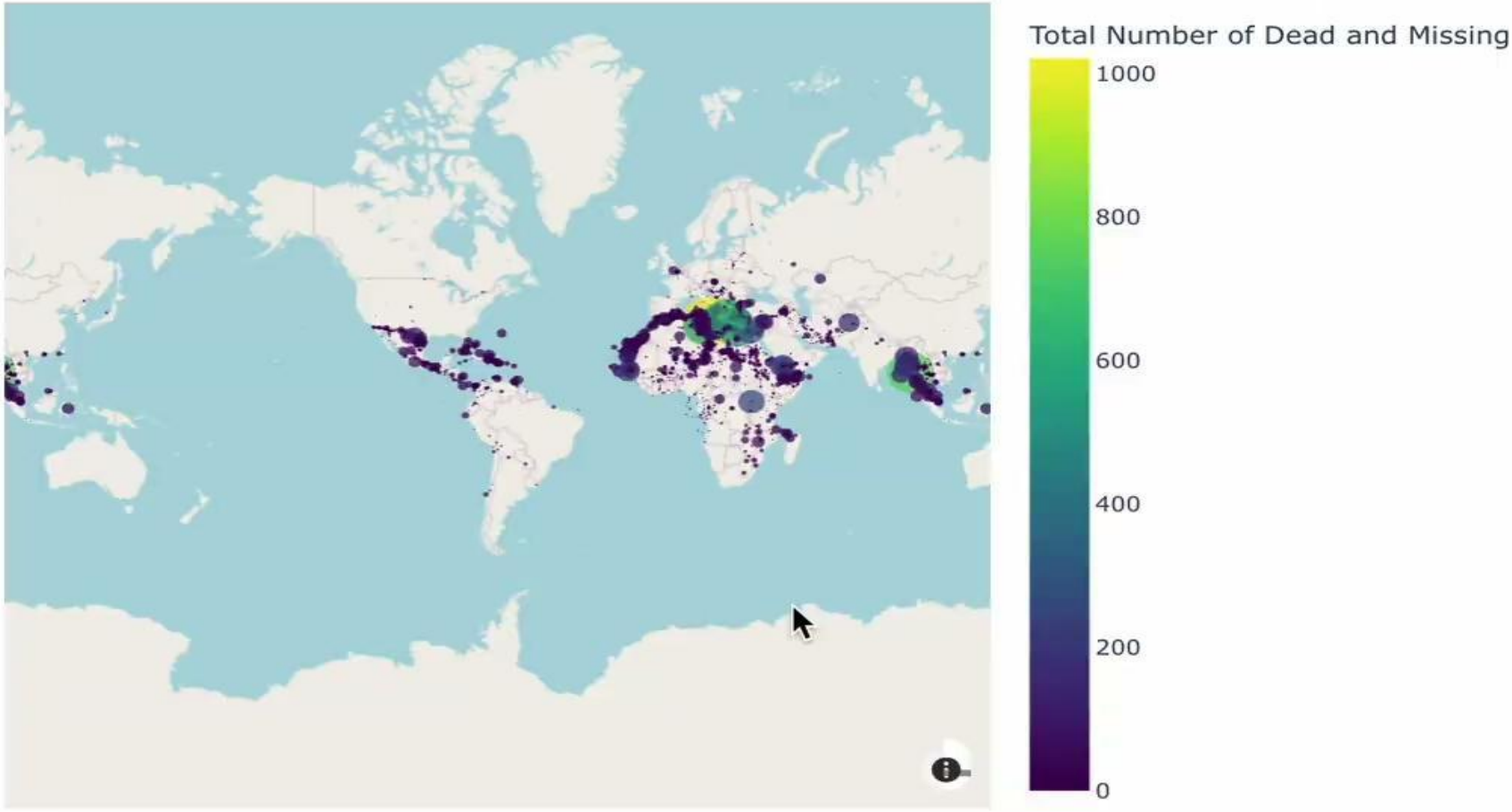




Entrée [15]

```
# Split the Coordinates column into Latitude and Longitude
migrant_deaths_data[['Latitude', 'Longitude']] = migrant_deaths_data['Coordinates'].str.split(','
                                                                    expand=True)
migrant_deaths_data['Latitude'] = pd.to_numeric(migrant_deaths_data['Latitude'])
migrant_deaths_data['Longitude'] = pd.to_numeric(migrant_deaths_data['Longitude'])

# Create the map
fig = px.scatter_mapbox(migrant_deaths_data,
                        lat='Latitude',
                        lon='Longitude',
                        hover_name='Country of Incident',
                        hover_data=['Incident Type', 'Total Number of Dead and Missing',
                                    'Incident Date'],
                        color='Total Number of Dead and Missing',
                        color_continuous_scale='Viridis',
                        size='Total Number of Dead and Missing',
                        zoom=1,
                        height=600)

fig.update_layout(mapbox_style="open-street-map")
fig.update_layout(title='Map of Migration Incidents with Total Deaths and Missing',title_x=0.5)
fig.show()
```

Map of Migration Incidents with Total Deaths and Missing



Mass grave of at least 65 migrants found in Libya desert, UN agency says

At least 65 migrants' bodies have been discovered in a mass grave in southwest Libya, the International Organization for Migration said on social media platform X on Friday.

Issued on: 22/03/2024 - 17:46 Modified: 22/03/2024 - 21:29 1 min

Home News Sport Business Innovation Culture Travel Earth Video Live

Ferromex: Mexican railway operator suspends routes amid migrant deaths

20 September 2023

Share

By Vanessa Buschschlüter, BBC News

Clambering up and down the wagons is very risky, and many migrants have been maimed or killed

A company which operates some of Mexico's railways has suspended 60 of its routes after a surge of migrants hitching rides on freight trains.

'Bloody policies': MSF recovers 11 bodies from Mediterranean off Libya

European countries accused of complicity in deaths due to migration policies after charity vessel retrieves 11 corpses in Mediterranean, rescues more than 160.

More than 20,000 deaths and disappearances have been registered in the central Mediterranean since 2014 [File: Nora Boerding/Sea-Watch via AP]

WORLD NEWS

Out of options, Rohingya are fleeing Myanmar and Bangladesh by boat despite soaring death toll

IMMIGRATION · GLOBAL ISSUES

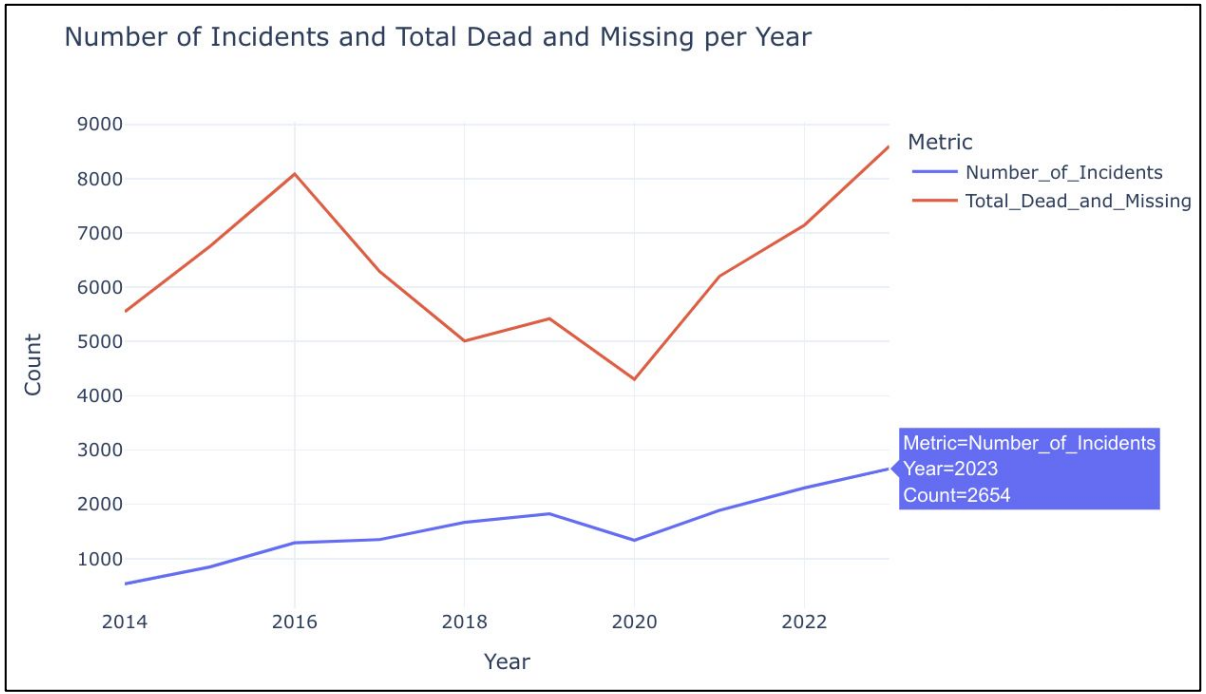
16 dead, dozens missing in shipwrecks off Tunisia, Western Sahara

Migrant deaths have surged in recent years as thousands flee war or crushing poverty, seeking to cross the Mediterranean in the hopes of finding better lives in Europe.

Le Monde with AFP

Published on August 7, 2023, at 6:47 pm (Paris) · 2 min read

Migrant Incidents and Deaths



The year **2016** stands out with the highest number of total dead and missing, **around 9,000**, indicating a particularly severe period in terms of human impact. Following a decline, the numbers rise again, reflecting ongoing and possibly increasing challenges in managing and preventing such incidents.

Migrant Incidents and Deaths

Key Events in 2016:

Conflict and War:

- **Syrian Civil War**
- **Iraq Conflict**
- **Yemen Civil War**

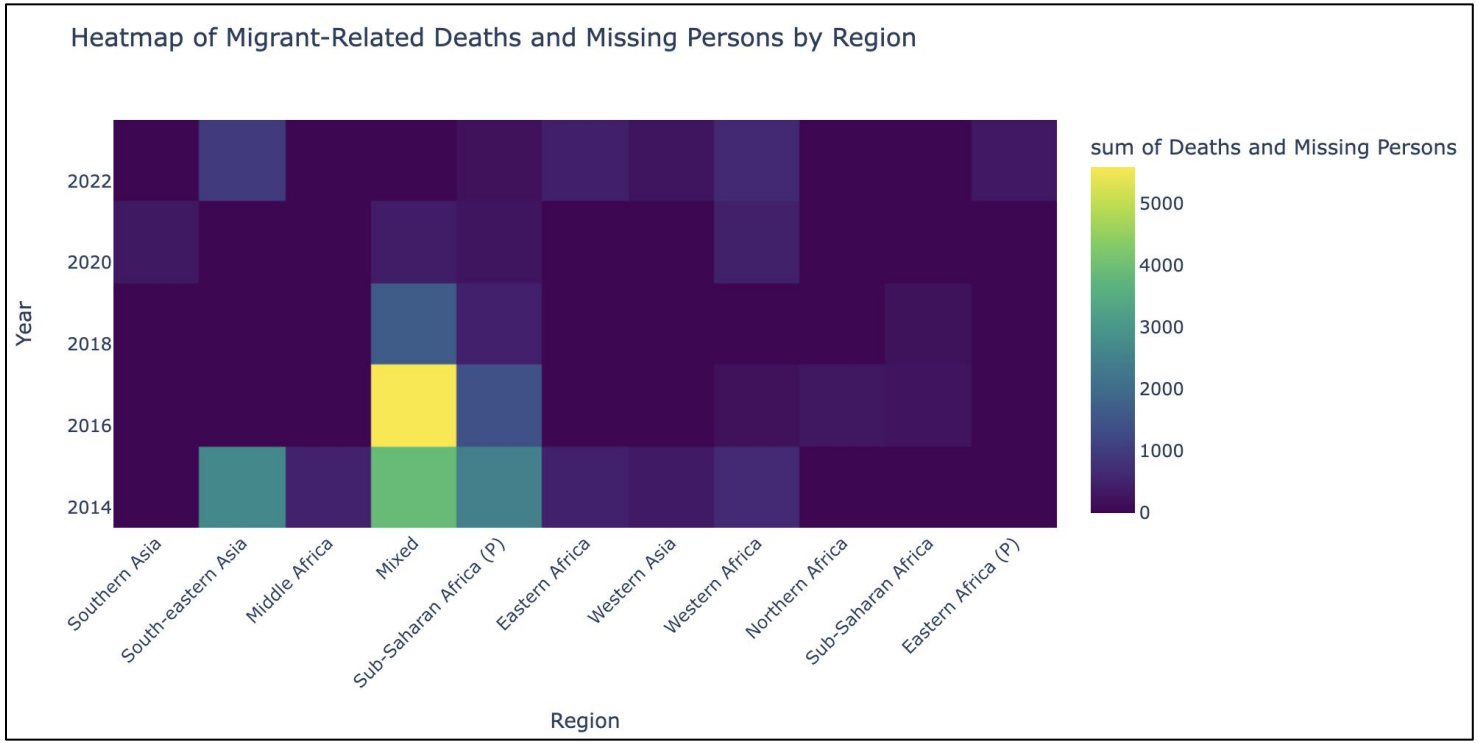
Natural Disasters:

- **Earthquakes**: Several devastating earthquakes occurred in 2016, including a significant one in Ecuador (7.8 magnitude) in April, resulting in considerable loss of life and property.
- **Hurricane Matthew**: This powerful hurricane caused extensive damage and loss of life in the Caribbean and the southeastern United States, particularly in Haiti where it exacerbated an already dire situation following the 2010 earthquake.

Migration and Refugee Crisis:

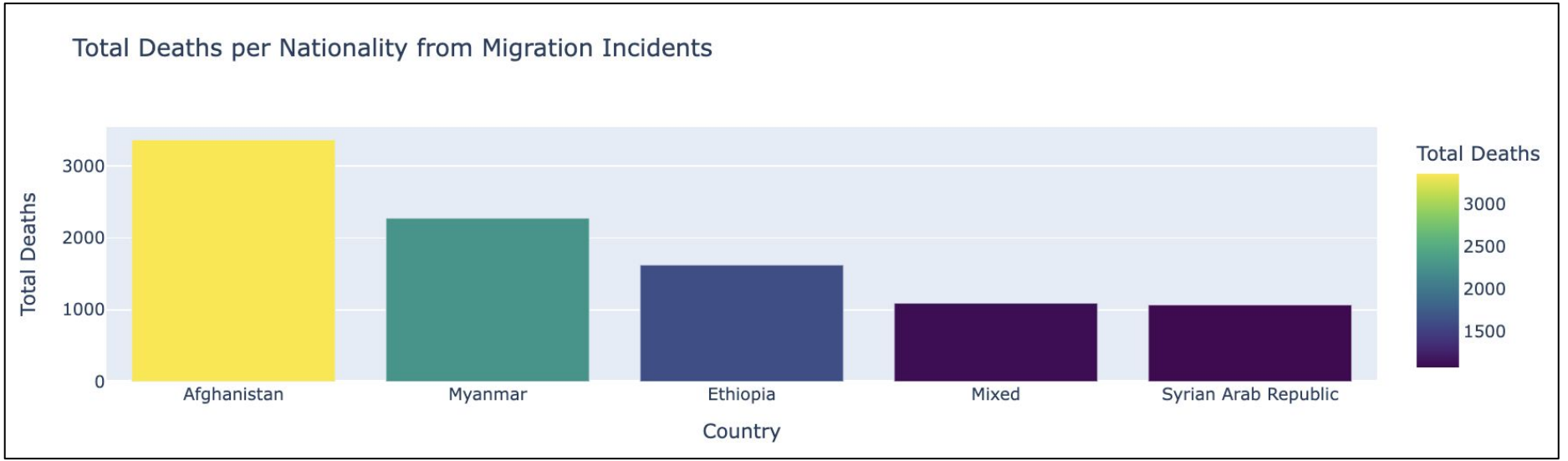
- The Syrian conflict, along with instability in other regions such as Afghanistan and parts of Africa, contributed to one of the largest migration crises in recent history. Europe saw a significant influx of refugees and migrants, leading to political and social tensions.

Migrant Deaths per Region per Year



*Heat Map depicts 90 or more deaths to improve readability

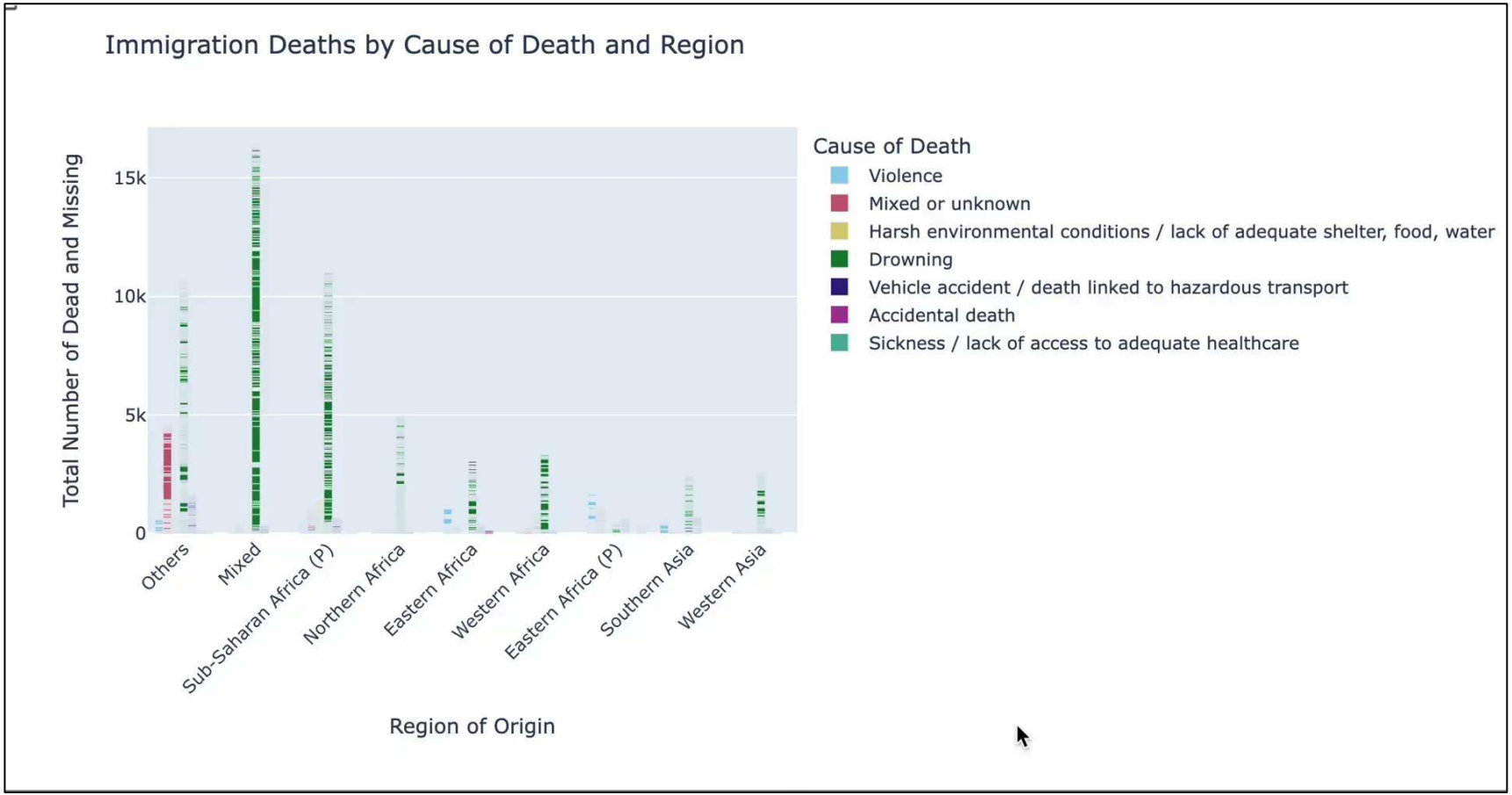
Most Deaths and Missing per Country due to Migration



Observations:

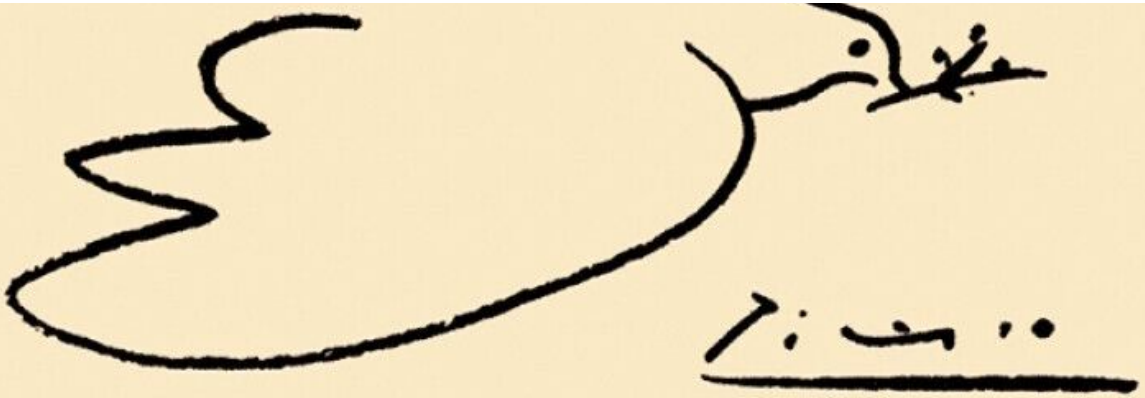
- **Afghanistan and Myanmar**: These countries have the highest number of migration-related deaths, pointing to the extreme risks and hardships faced by migrants from these regions.
- **Significant Numbers for Ethiopia and Syria**: Both countries also show substantial numbers, reflecting ongoing conflicts and crises contributing to dangerous migration conditions.
- **Mixed Nationalities**: The presence of a significant number of deaths where nationalities are mixed or unknown highlights the complexity and breadth of migration challenges.

Cause of Death



Cause of Death

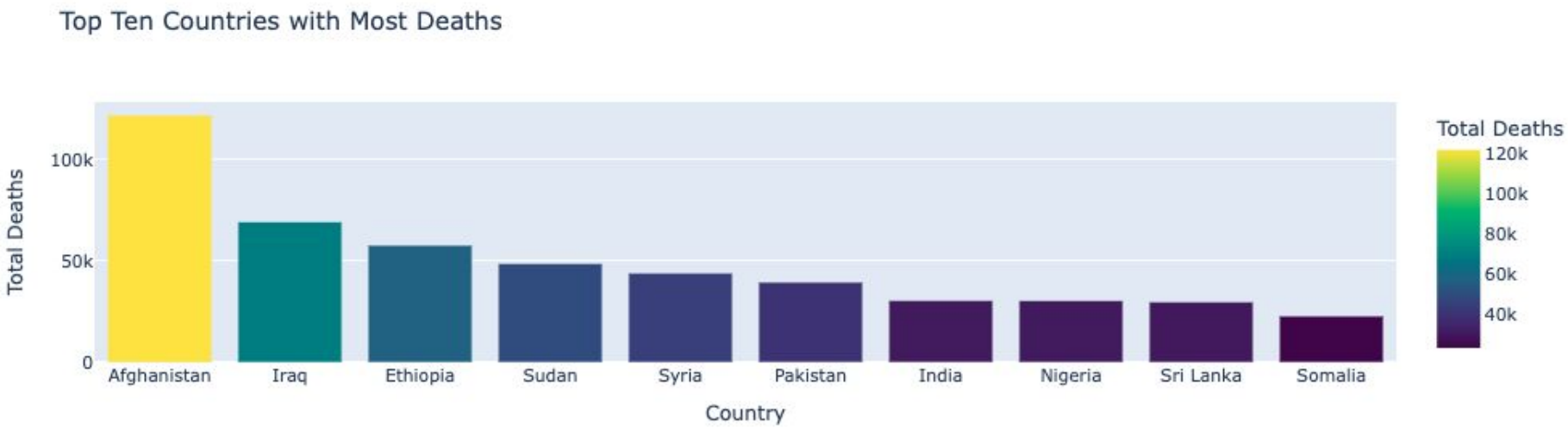
World Conflicts Analysis



"Dove" by Pablo Picasso

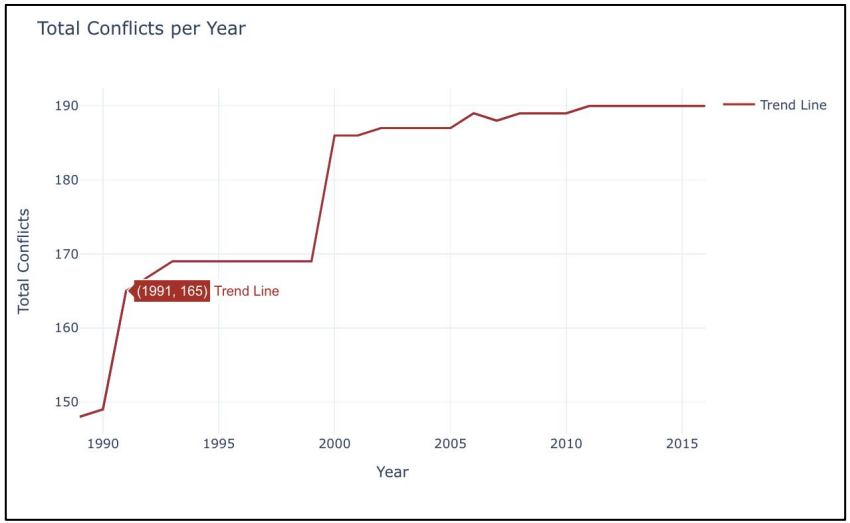
Map of World Conflicts

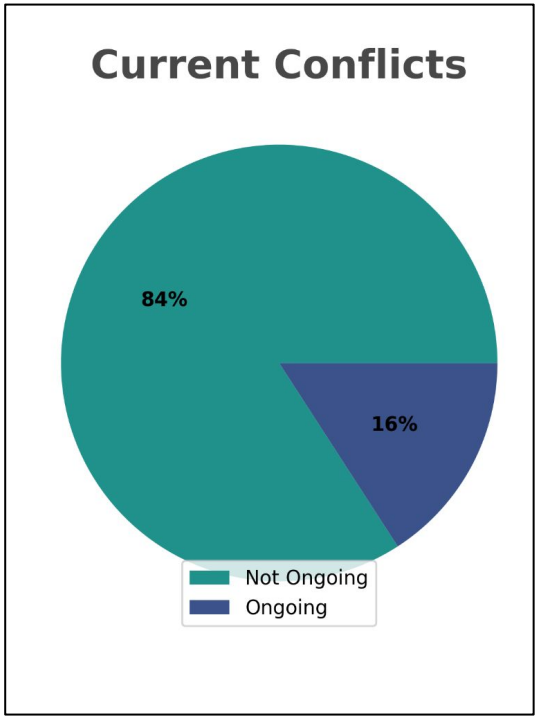
Most Deaths per Country due to Conflicts



The visualization highlights the devastating impact of conflicts, with Afghanistan, Iraq, and Ethiopia showing the highest death tolls. These regions face severe and prolonged violence, resulting in significant loss of life. The data underscores the urgent need for conflict resolution and humanitarian aid to mitigate these tragedies. It emphasizes the critical importance of international intervention to support affected populations and promote peace. The stark figures call for concerted global efforts to address the root causes of conflict and protect the most vulnerable communities.

Increase of World Conflicts Yearly





ongoing: dummy variable coded 1 for more than 25 deaths in intrastate conflict and 0 otherwise according to UCDP/PRIO Armed Conflict Dataset 17.1.

- **coded 1:** If a conflict results in more than 25 deaths in an intrastate conflict, this variable is set to 1.
- **coded 0:** If a conflict results in 25 or fewer deaths in an intrastate conflict, this variable is set to 0.

Classification: Severity of migrant incidents to effectively deploy rescue e

"Human Flow" Documentary by Ai Weiwei

1 Define thresholds for the severity of incidents:

< 500 = low

< 1000 = medium

Else = high

2 Prepare data and run Random Forest Classifier

3 Evaluate results

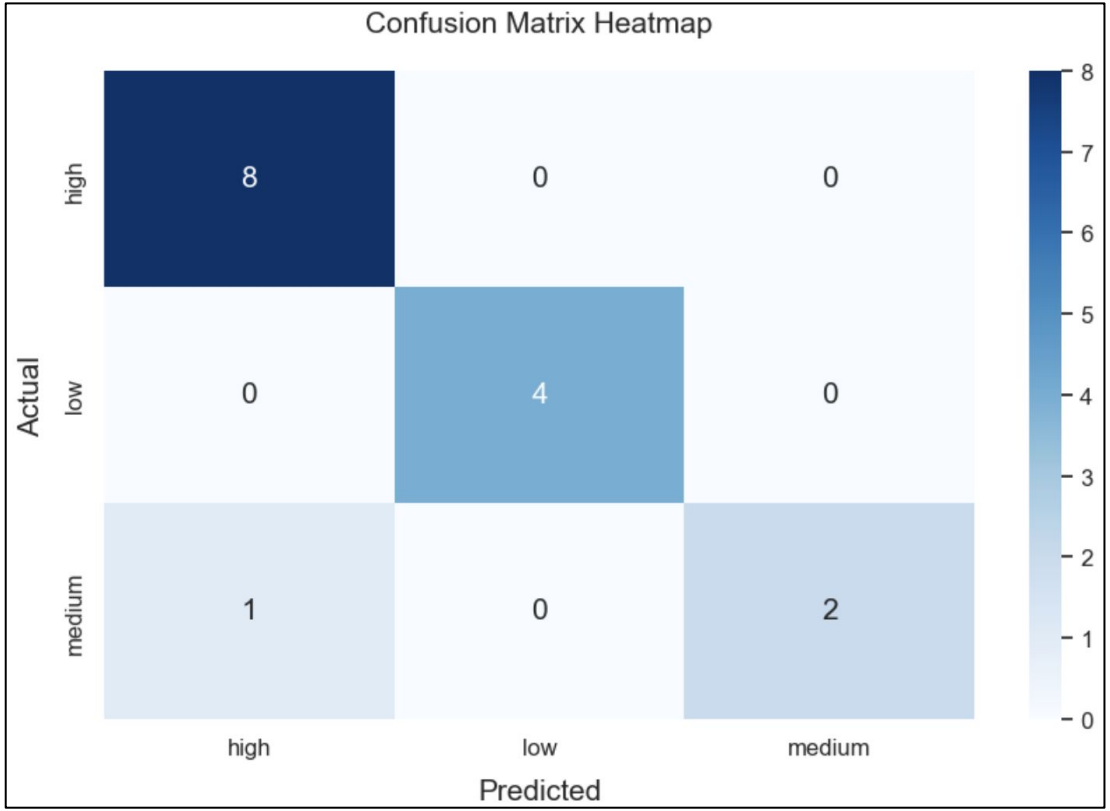
```python
# Select features and target variable
features = ['Geographic area', 'Number of Dead', 'Number of Survivors', 'Total Number of Dead and Missing']
target = 'Severity'

# Prepare the data
df = final_df[features].dropna()

# Create the severity column
df[target] = df.apply(classify_severity, axis=1)

# One-hot encode the 'Geographic area' column
one_hot_encoder = OneHotEncoder()
encoded_geographic_area = one_hot_encoder.fit_transform(df[['Geographic area']]).toarray()
encoded_geographic_area_df = pd.DataFrame(encoded_geographic_area, columns=one_hot_encoder.get_feature_names_out(['Geographic area']))

# Merge the encoded columns with the original dataframe
df = pd.concat([df.reset_index(drop=True), encoded_geographic_area_df.reset_index(drop=True)], axis=1).drop(columns=['Geographic area'])

# Split the data into training and testing sets
X = df.drop(columns=[target])
y = df[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train a Random Forest Classifier
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

```
Classification Report:
              precision    recall  f1-score

        high       0.89      1.00      0.94
         low       1.00      1.00      1.00
      medium       1.00      0.67      0.80

    accuracy                           0.93
   macro avg       0.96      0.89      0.91
weighted avg       0.94      0.93      0.93
```

Very accurate for the "high" class, with 8 true positives and no false negatives or false positives.

Good performance for the "low" class, with 4 true positives and no false negatives or false positives.

Some issues with the "medium" class, with only 2 true positives and 1 false negative. Additionally, there is 1 instance of a "medium" being predicted as "high".



Predict Migrant Deaths and Missing

Installation “Law of the Journey” by Ai Weiwei

```python
# Merge the datasets on 'Country of Incident' and 'Country of Destination'
merged_data = pd.merge(migrant_deaths_data, int_migrants_data, left_on='Country of Incident', right_on='Geographic area', how='left')

# Drop rows with missing values in the target column
merged_data = merged_data.dropna(subset=['Total Number of Dead and Missing'])

# Fill missing values in other columns with suitable values (e.g., 0, mean, median)
merged_data = merged_data.fillna(0)

# Select features and target variable
features = ['Incident Year', 'Number of Females', 'Number of Males', 'Number of Children', 'Number of Survivors',
            'Minimum Estimated Number of Missing', 'Number of Dead']
target = 'Total Number of Dead and Missing'

X = merged_data[features]
y = merged_data[target]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```python
# Train a RandomForestRegressor model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

1

Merge data frames, clean data, and fill missing values

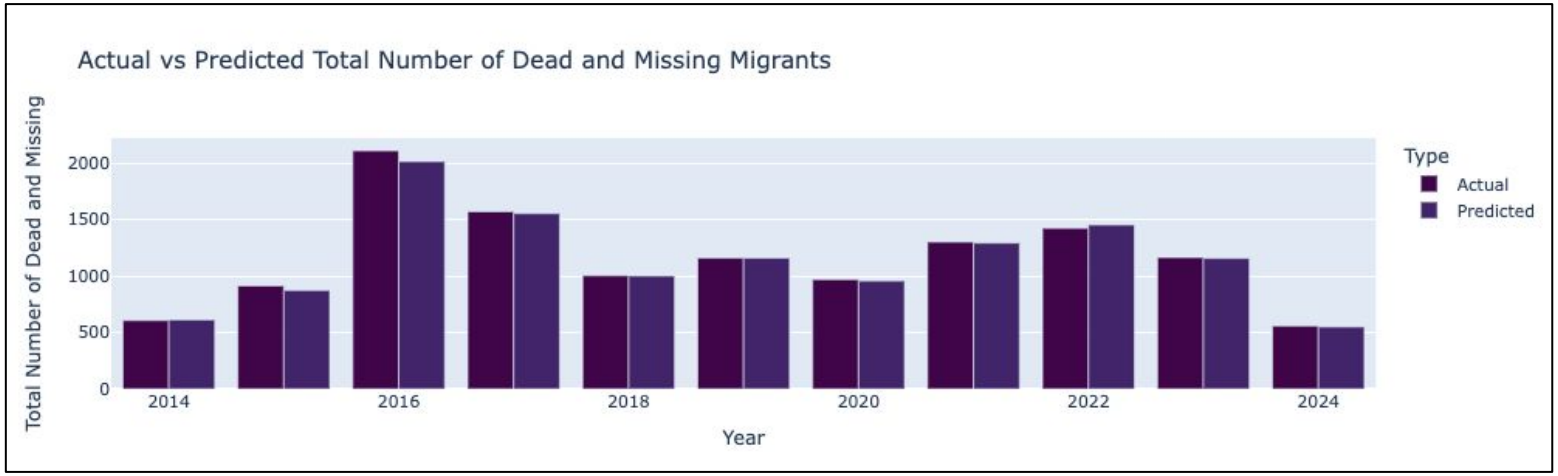
2

Select features to train model

3

Evaluate the performance of the model

Prediction of IDPs based on conflicts

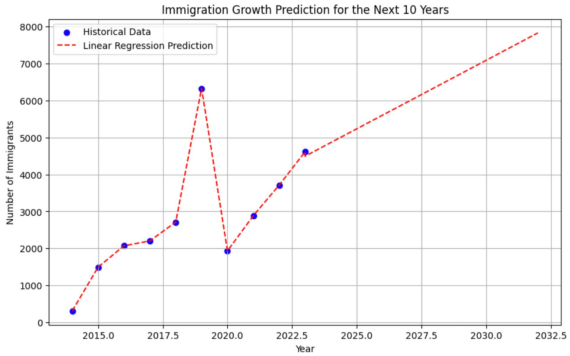



```
Mean Absolute Error (MAE): 0.10713312693498452
Mean Squared Error (MSE): 3.5895617337461307
Root Mean Squared Error (RMSE): 1.8946138745787044
R² Score: 0.9868936872172935
```

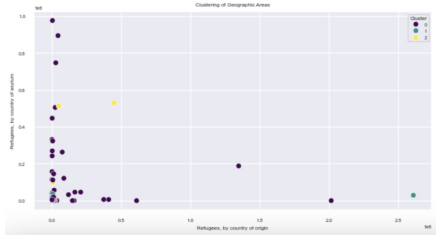
Project Learnings

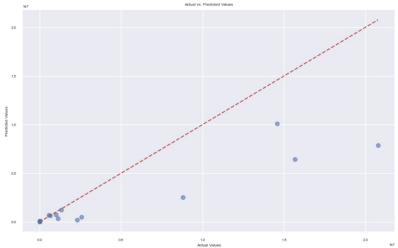
"Thank God For Immigrants" by Jeremy Deller (Projected during the COVID 19 Pandemic)

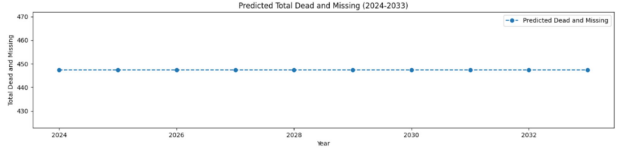
What we tried and how we can improve it?

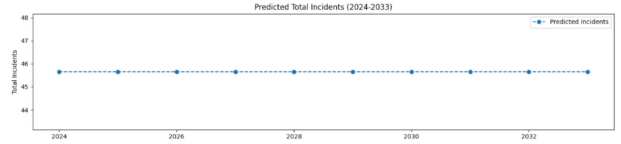



```
Mean Absolute Error: 5466.01 degrees.
Accuracy: -236.34 %.
```

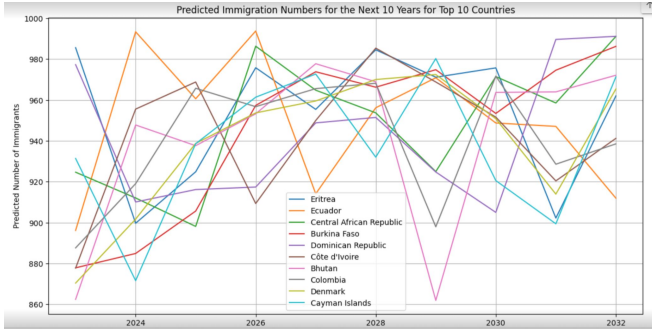








precision	recall	f1-score
1.00	1.00	1.00
1.00	1.00	1.00
		1.00
1.00	1.00	1.00
1.00	1.00	1.00




```
RMSE: 686.0581088484172
MAE: 542.7153333333333
R-squared: 0.004874751651025688
```

	Geographic area	Predicted International migrants
32	India	2227.270
68	Ukraine	2061.550
47	Mexico	1979.040
54	Nigeria	1956.930
24	Ethiopia	1894.380
38	Lebanon	1742.900
44	Malaysia	1741.360
0	Afghanistan	1707.820
56	Pakistan	1685.080
63	South Sudan	1607.535

What we tried and how we can improve it?

Linear regression:
No linear relationship between variables as well as outliers (countries and years with very little data)

Neural networks:
Difficult to interpret results (ie. Denmark as a country with high number of migrants)

Time Series:
The data time series didn’t coincide (some data until 2015, others until 2022)

Clustering:
Too many outliers (small values from certain countries and years)

Possible improvements:

- Analyze the size of boats and capacity to improve parameter definition so estimate the resources needed.
- Data collection:
  - Incomplete data on conflicts analysis (only until 2015).
  - Not enough data on international migrants (constant value from 2010-2020)

Conclusions

Darien Gap by the NYT

Strategies to Reduce Migration-Related Issues

1. Enhance Human Security and Safety

- **Safe Migration Pathways**: Governments and international organizations must collaborate to create safe and legal migration routes to reduce the reliance on dangerous, irregular pathways.
- **Humanitarian Assistance**: Increase funding and support for humanitarian aid in conflict zones and areas affected by natural disasters to prevent forced migration.

2. Address Root Causes of Displacement

- **Conflict Resolution**: Invest in peacebuilding efforts and conflict resolution initiatives in regions with high levels of displacement.
- **Climate Adaptation**: Implement strategies to mitigate the impacts of climate change, such as building resilient infrastructure and providing support for affected communities.

3. Support Economic Stability and Development

- **Economic Opportunities**: Create job opportunities and improve economic conditions in migrants' home countries to reduce the economic pressure to migrate.
- **Education and Skills Training**: Provide education and vocational training to improve the employability of individuals in their home countries and abroad.

Recommendations for Associations and Governments

1. Policy and Legal Frameworks

- **Comprehensive Migration Policies**: Develop and implement comprehensive migration policies that address all aspects of migration, from pre-departure to integration.
- **Legal Protections**: Strengthen legal protections for migrants, ensuring their rights and safety are upheld in transit and destination countries.

2. Integration Programs

- **Community Integration**: Promote programs that facilitate the integration of migrants into host communities, such as language courses, cultural orientation, and employment assistance.
- **Social Services**: Ensure access to healthcare, education, and social services for migrants to support their well-being and integration.

3. Utilizing Migrants as Investments

- **Leveraging Skills**: Recognize and utilize the skills and talents of migrants to contribute to the economy. Implement recognition of foreign qualifications and provide pathways for skilled migrants.
- **Entrepreneurship Support**: Encourage and support migrant entrepreneurship through access to credit, business training, and mentorship programs.

Further Reading and Resources

Darien Gap by the NYT

Links to Articles and Publications

- Using Machine Learning to Help Refugees Succeed - Stanford University
- Migrant children in U.S. detention face physical, mental harms: report - Harvard University
- Global Conflict Tracker - CFR
- Migration to Advanced Economies Can Raise Growth - International Monetary Fund
- Human Flow Documentary by Ai Weiwei
- Source for definitions: Amnesty International (https://www.amnesty.org/en/what-we-do/refugees-asylum-seekers-and-migrants/)

Data Sources (with links)

- UNICEF - Immigration
- World Bank - World Conflicts
- Missing Migrants Project - Deaths in immigration