Data Biography and EDA Findings

Resources:

Raw data:
Cleaned data:

Purpose and Context of the Data

- To quantify and document the number of migrant fatalities, disappearances, and survivals.
- To identify patterns related to causes of death, incident locations, and migration routes.
- To create awareness about the human cost of migration.
- To inform policy-making and humanitarian response strategies.

Data Overview:

- Total number of rows: **177**
- Total number of columns: **26**
- The dataset records incidents related to migrant journeys, including details on the incident type, location, casualties, survivors, and other contextual data.

Data types:

1. Main ID: object
2. Incident ID: object
3. Incident Type: object
4. Region of Incident: object
5. Incident Date: object
6. Incident Year: int64
7. Month: object
8. Number of Dead: float64
9. Minimum Estimated Number of Missing: float64
10. Total Number of Dead and Missing: int64
11. Number of Survivors: float64
12. Number of Females: float64
13. Number of Males: float64
14. Number of Children: float64

15. Country of Origin: object
16. Region of Origin: object
17. Cause of Death: object
18. Country of Incident: object
19. Migration Route: object
20. Location of Incident: object
21. Coordinates: object
22. UNSD Geographical Grouping: object
23. Information Source: object
24. URL: object
25. Source Quality: int64
26. DistanceFromCapital: float64

Potential Biases and Limitations:

- Gender and age data are frequently missing or incomplete.
- Incidents involving groups of migrants may underreport individual details.
- Incidents in remote or dangerous locations may be underreported.
- Media bias toward specific countries or types of incidents may skew data.

Outliers and Anomalies

- Highest Fatality Incident: 40 deaths in a single incident in Mexico.
- Lowest Fatality Incident: 0 fatalities recorded in 12 incidents.
- Highest Number of Survivors: 43 survivors in a single incident in Honduras.

Ethical Considerations and Sensitivity

- Reporting discrepancies between governmental sources and humanitarian organizations could lead to mistrust or conflict.
- Human lives are represented as numerical data, which could minimize their personal significance.
- Balancing between raising awareness and respecting the dignity of victims is essential.

Key Takeaways

- The dataset reveals the complex human toll of migration journeys, with significant loss of life and a high level of survivor **trauma**.
- Patterns of incidents suggest systemic dangers linked to geographical and political factors.
- Data gaps reflect reporting biases and the complexity of gathering comprehensive information on migrant journeys.