Death Age Analysis

By : Mariam Hossam Diab

Gmail:
marimhosamdiab@gmail.com

All task files :

https://drive.google.com/drive/folders/1hCpvFqrtQcfuHc3m6ZYrzMFVdAY2Dd7F

Merged dataset:

https://drive.google.com/file/d/19ix3oN6I135DAwpmQOQ5GdirMQcnMMIr/view?usp=drive_link

Eda link :

https://drive.google.com/file/d/1dfQ_gldrmlcqy5B3oXdh9_PXZzl32Vt8/view?usp=drive_link

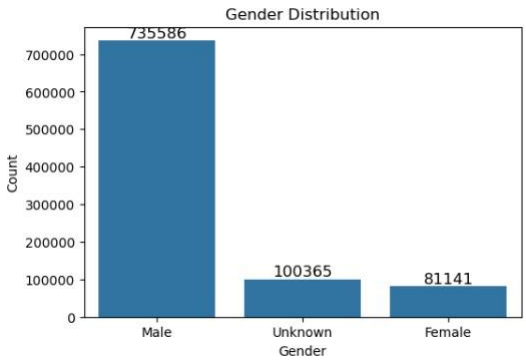
Answer part 4

Analysis and Insights

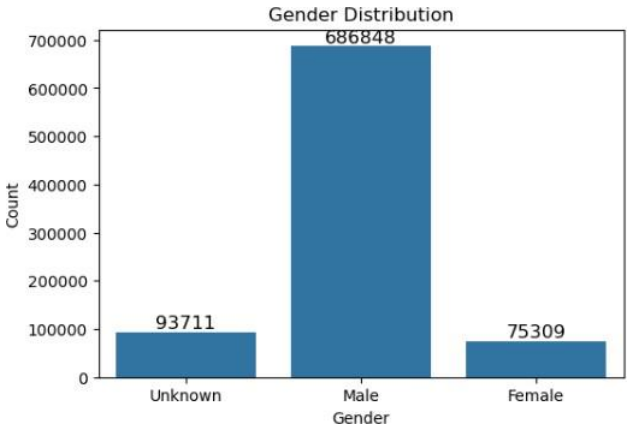
1. Gender Distribution:

○ What is the distribution of genders (Male and Female) in the combined dataset? How does it differ between the two datasets before matching?

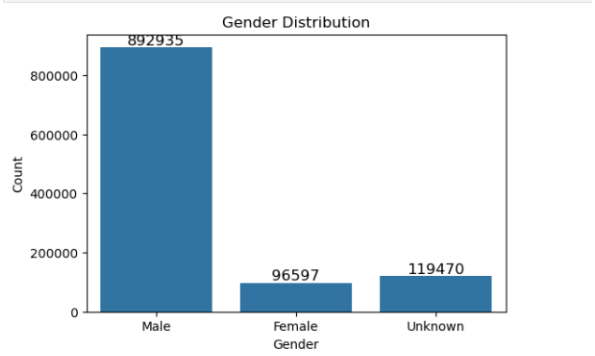
Dataset1



Data set 2

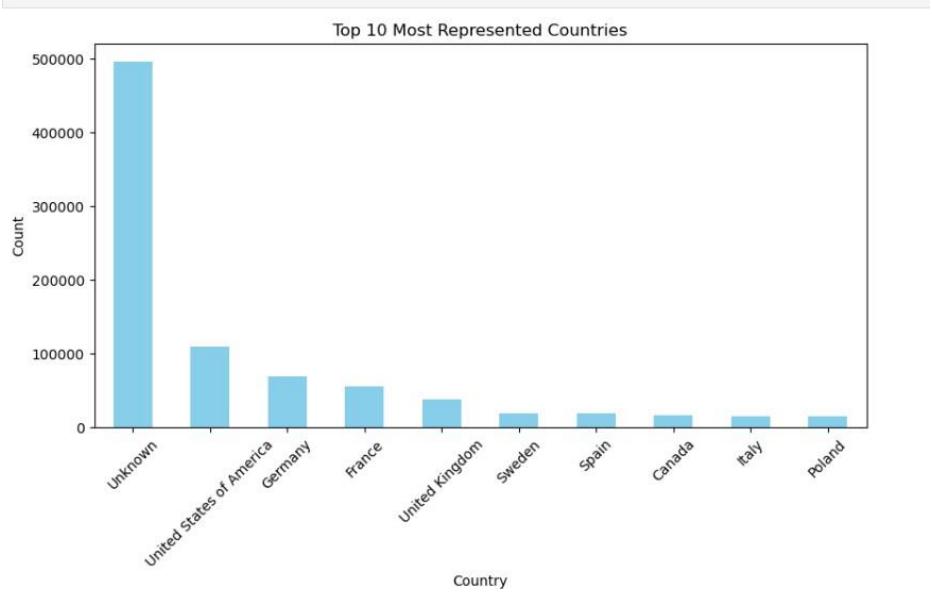


After merge



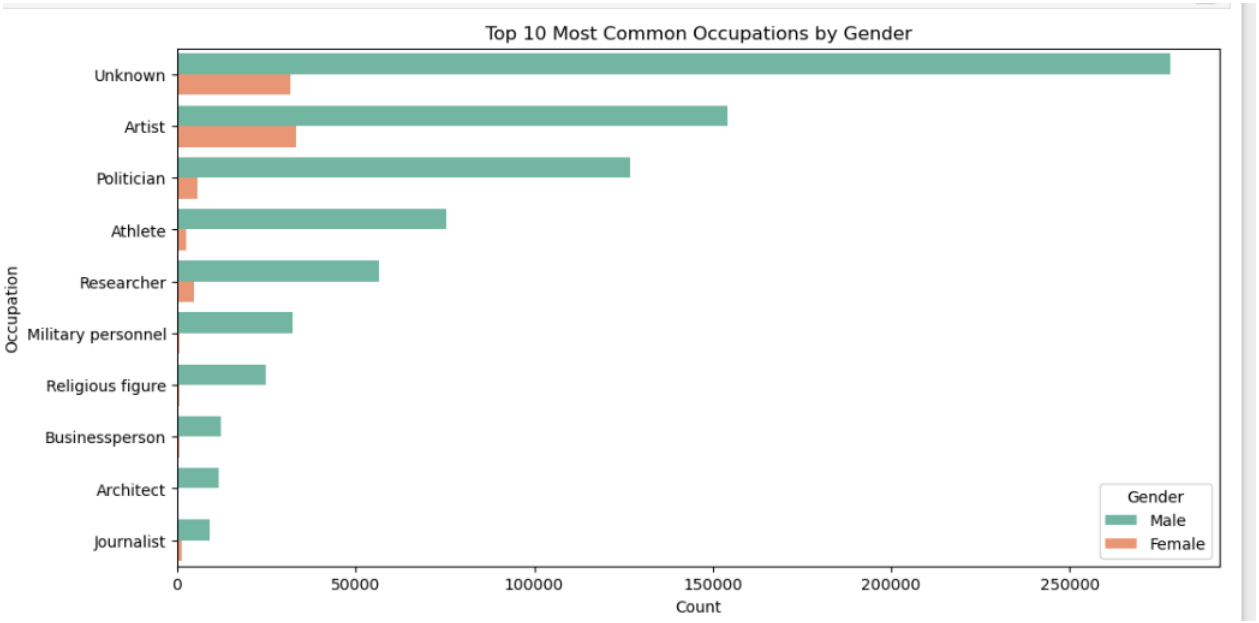
2. Geographic Representation:

○ Which countries are most represented in the dataset, and are there any discrepancies in country distributions between Dataset 1 and Dataset 2?



3. Occupation Trends:

○ What are the most common occupations, and how do they vary by gender?



4. Historical Patterns:

○ What is the average age at death for individuals in the dataset? Are there differences based on gender or country?

```
merged['Age of death'].mean()
```

```
70.30995255193407
```

```
merged.groupby('Country')['Age of death'].mean()
```

```
Country
Abbasid Caliphate     70.220930
Abbasids              58.000000
Achaemenid Empire     66.000000
Afghan                57.000000
Afghanistan           63.343284
                          ...
historical country    59.500000
history of Genoa      50.000000
island nation         79.500000
nationality           89.000000
statelessness         73.000000
Name: Age of death, Length: 950, dtype: float64
```

```
merged.groupby('Gender')['Age of death'].mean()
```

```
Gender
Female     72.905059
Male       70.077497
Unknown    69.949092
Name: Age of death, dtype: float64
```

5. Analysis of Missing Data:

○ After matching, what proportion of records still contain missing values in key fields such as Country or Manner of death? What strategies could be used to impute or address these gaps?

```
print(f"Number of null values in 'gender': {unknown_count_gender}")
```

```
Number of null values in 'Country': 495789
Number of null values in 'Manner of death': 1073697
Number of null values in 'gender': 119470
```

Handling Missing gender

Use Forward/Backward Fill (if data is sequential) by name if the gender was known

Handling Missing manner of death

this column has many nan value we may drop it or filter data , that value is exist

Handling Missing country

Replace with "Unknown" or "Other"

```
#merged.to_csv('output1.csv', index=False)
```

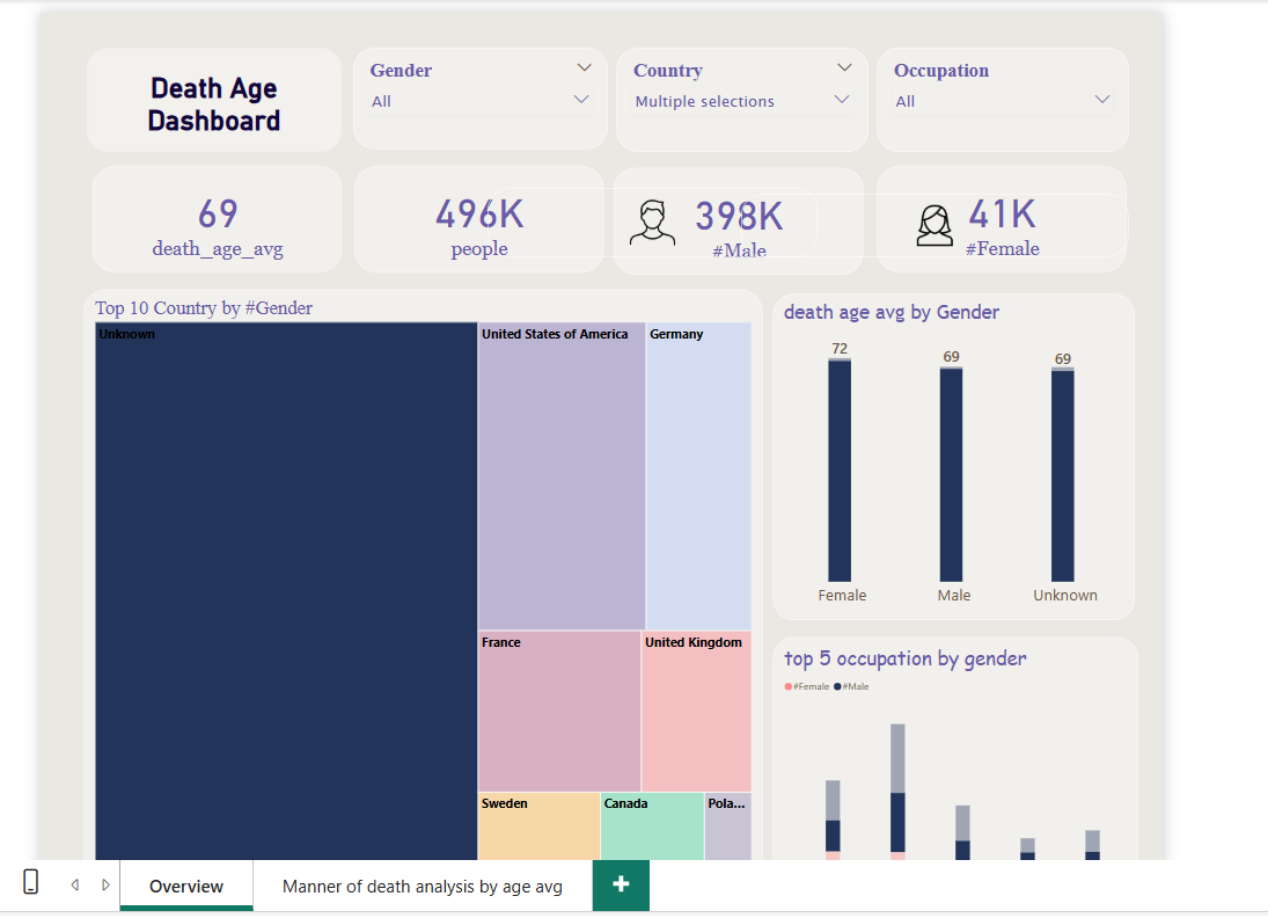
Handling Missing Gender Using Forward/Backward Fill

```
# Replace "Unknown" with None (NaN)
merged['Gender'] = merged['Gender'].replace("Unknown", None)

# Forward fill missing values
merged['Gender'] = merged.groupby('Name')['Gender'].ffill()

print("\nAfter filling missing gender values:")
print(merged["Gender"].value_counts())
```

```
After filling missing gender values:
Gender
Male      896386
Female     96805
Name: count, dtype: int64
```

Dashboard link :

https://drive.google.com/file/d/1_41xkigk8J2oENlkUIZjUkR9e0KPIMtv/view?usp=drive_link



1. Filters (Top Section)

The green section at the top contains three dropdown filters:

- **Gender** – Select "All" or filter by "Male" or "Female."
- **Country** – Choose one or multiple countries to filter the data.
- **Occupation** – Select an occupation to see relevant insights.

🔹 *How to Use:*

- Click on a dropdown menu.
- Select one or multiple options.
- The dashboard updates automatically based on your selection.

---

2. Key Metrics (Top Center)

Below the filters, four key metrics provide a quick summary:

- **Death Age Average** – Displays the average age at death.
- **Total People** – Shows the total number of records in the dataset.
- **Number of Males** – Represents the total count of male records.
- **Number of Females** – Represents the total count of female records.

- *How to Use:*

- These are static indicators that change dynamically based on your selected filters.

---

3. Data Visualizations (Main Section)

A. Top 10 Countries by Gender (Treemap)

- This chart visualizes the top 10 countries with data distribution by gender.
- Larger blocks represent a higher proportion of records.
- The "Unknown" category may include missing or unidentified country data.

🔹 *How to Use:*

- Hover over a country to see additional details.
- The size of each block indicates the relative proportion.

B. Top 5 Occupations by Gender (Bar Chart)

- Displays the top 5 occupations with gender representation.
- Darker sections represent male data, while lighter sections represent female data.

🔹 *How to Use:*

- Compare gender distribution across different occupations.
- Hover over bars for precise values.

C. Death Age Average by Gender (Bar Chart)

- Shows the average death age for each gender.
- "Unknown" represents records where gender is unspecified.

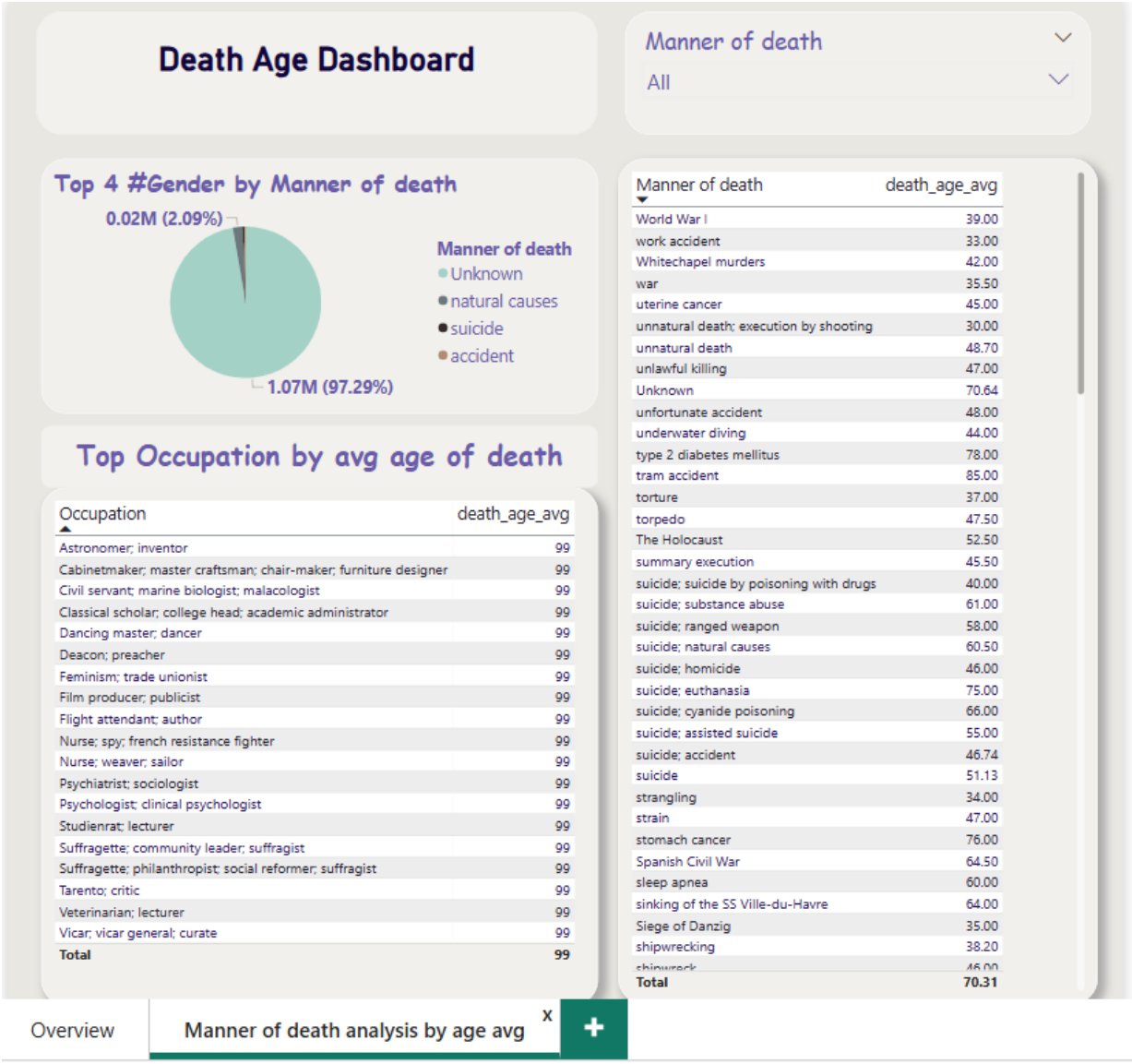
🔹 *How to Use:*

- Compare life expectancy trends by gender.
- Hover over bars to see specific values.

---

4. Navigation & Additional Features

- The bottom panel has tabs for different analysis views.
- Click **"Manner of death analysis by age avg"** to explore another detailed analysis.
- Use the **"+" button** to add new reports or custom insights (if enabled).



5. Filter Section (Top Right)

🔹 **Manner of Death Dropdown** – Allows users to filter data based on the cause of death, including:

- **Natural causes**
- **Suicide**
- **Accident**
- **Unknown**

✅ *How to Use:*

- Click the dropdown menu.
- Select a specific manner of death to refine the results.
- The dashboard updates automatically based on your selection.

---

6. Pie Chart: "Top 4 #Gender by Manner of Death"

- Shows the distribution of deaths by manner.
- Different colors represent different categories (natural causes, suicide, accident, and unknown).
- The **percentage** and **total count** are displayed.

✅*How to Use:*

- Hover over the chart to see detailed percentages.
- Use the manner of death filter to observe category-specific changes.

---

7. Table: "Manner of Death & Average Death Age" (Top Right Panel)

- Displays various causes of death alongside their **average age at death**.
- Sorted in descending order of frequency or impact.
- Causes include **work accidents, suicides, war-related deaths, illnesses, and other fatal incidents**.

✅*How to Use:*

- Scroll down to explore different causes of death.
- Identify causes with significantly high or low average death ages.

---

8. Table: "Top Occupations by Average Age of Death" (Bottom Panel)

- Lists occupations alongside their **average age at death**.
- Some professions (e.g., **Vicar, Lecturer, Astronomer**) have an average death age of **99**, suggesting longevity.
- Other professions may show significantly lower life expectancy based on risk factors.

✅*How to Use:*

- Analyze which occupations are associated with higher or lower death ages.
- Compare trends between different job roles.