**✈️ Aircraft Accident Data Analysis Report**

**By Sulaiman Fayzal Gbolahan**

**Data Source:** **https://www.kaggle.com/datasets/jogwums/air-crashes-full-data-1908-2023/data**

**Overview**

This project focuses on the analysis of global aircraft accident data to identify key patterns, trends, and safety insights within the aviation industry. Using a dataset containing information such as accident year, month, country, aircraft manufacturer, fatalities, and operator details, the analysis explores the temporal and geographical distribution of air crashes and their outcomes.

The work was conducted across two major environments — **Jupyter Lab** and **Visual Studio Code (VS Code)**.

* In **Jupyter Lab**, data preprocessing, cleaning, and exploratory analysis were performed using Python libraries like **Pandas**, **Matplotlib**, and **Seaborn**.
* In **VS Code**, interactive data visualization and dashboard development were carried out using **Streamlit** and **Altair**, enabling intuitive exploration of results.

The research addressed ten core analytical questions focusing on accident frequency, fatality rates, manufacturers, and geographical patterns. Overall, the findings show that while accidents have historically fluctuated, there has been a general **decline in frequency and fatality severity over time**, reflecting ongoing improvements in global aviation safety.

## ****1. Introduction****

This report presents an analytical study of aircraft accidents to understand their frequency, severity, and distribution patterns across time, regions, and aircraft types.  
The dataset, consisting of multiple features such as year, month, country, operator, aircraft type, fatalities, and number of people aboard, was processed and analyzed to answer ten defined research questions.

The analysis was initially performed in **Jupyter Lab** for exploratory and data wrangling tasks, while **visualization and interactive dashboards** were implemented in **Visual Studio Code (VS Code)** using **Streamlit** and **Altair**.

The goal of this analysis is to identify meaningful insights from accident trends, highlight safety concerns, and observe changes in accident patterns across time.

## ****2. Environment and Tools Used****

| **Tool / Library** | **Purpose** |
| --- | --- |
| **Jupyter Lab** | Initial data exploration, cleaning, and analysis. |
| **VS Code** | Development environment for Streamlit dashboard. |
| **Python** | Programming language used for the entire analysis. |
| **Pandas** | Data manipulation, grouping, and aggregation. |
| **Altair** | Interactive charting and visualization. |
| **Streamlit** | Building an interactive dashboard for visualization results. |
| **Matplotlib / Seaborn** | Preliminary visualization and trend confirmation. |
|  |  |

## ****3. Methodology****

1. **Data Loading and Inspection:**  
   The dataset was imported using Pandas and explored for missing values, structure, and data consistency.
2. **Cleaning and Preparation:**  
   Columns were standardized (e.g., renaming inconsistent labels, parsing date fields), missing values were handled, and derived columns like Year, Month, and Quarter were created from date attributes.
3. **Exploratory Analysis:**  
   Grouping and aggregation were applied using groupby() functions to answer each research question.
4. **Visualization (Streamlit + Altair):**  
   Interactive line, bar, and pie charts were created for better engagement and interpretability. The Altair library allowed tooltip-based interactivity and filtering across charts.

## ****4. Research Questions and Findings****

### **Question 1:**

**How has the number of aircraft accidents varied by year, quarter, and month over time?**

* A **line chart** displayed using Altair showed general fluctuations in accidents per year, with some years recording significant spikes.
* The **quarterly and monthly breakdowns** revealed that certain quarters, particularly Q2 and Q3, showed relatively higher accident frequencies — possibly due to increased air traffic during travel seasons.

### **Question 2:**

**Which countries or regions record the highest frequency of aircraft accidents, and how do their fatality rates compare?**

* A **bar chart** comparing countries showed that nations with large air traffic volumes (such as the U.S., Russia, and Brazil) recorded higher accident frequencies.
* However, the **fatality rate comparison** indicated that frequency does not always correlate with severity — some countries had fewer accidents but higher average fatalities per crash.

### **Question 3:**

**Top 10 Aircraft Manufacturers with Most Crashes**

* A grouped visualization identified manufacturers like **Boeing, Cessna, and Airbus** among the most frequently involved.
* However, normalization by production volume would provide deeper context to these findings.

### **Question 4:**

**Aircrafts with the Most Fatalities**

* This analysis highlighted specific aircraft models that recorded the highest cumulative fatalities.
* These insights could inform maintenance practices and historical safety reviews for certain aircraft types.

### **Question 5:**

**Which Years Recorded the Highest Number of Air Crashes, and What Are the Top Ten?**

* The **top ten years** showed concentration during certain decades — notably the mid-20th century, aligning with less advanced aviation technology and safety measures.
* Recent years generally reflect improved safety standards and technological evolution in aviation systems.

### **Question 6:**

**How Has the Number of Aircraft Crashes Changed Over Time Per Year?**

* A **yearly trend chart** using Altair line plots demonstrated a steady decline in accident occurrences over recent decades, emphasizing progress in aviation safety and regulation.

### **Question 7:**

**Top 10 Countries with Fatalities**

* Visualization revealed that countries with both high air traffic and extensive geographic coverage recorded the highest fatalities.
* This again reflects not just frequency, but also population exposure and air travel density.

### **Question 8:**

**Which Quarter of the Year Recorded the Highest Fatalities?**

* Aggregated data indicated **Q2 and Q3** (April–September) often recorded higher fatality numbers, aligning with increased flight operations during summer months in many regions.

### **Question 9:**

**Which Aircraft Had the Highest Average Number of People Aboard?**

* Wide-body commercial aircraft showed the highest average passenger count per crash, aligning with fleet size and international operations.
* The insight highlights how aircraft capacity influences potential fatality impact per incident.

### **Question 10:**

**What Proportion of Crashes Involved Fatalities Versus No Fatalities?**

* Using an Altair **pie/donut chart**, fatal crashes accounted for a smaller proportion relative to non-fatal incidents.
* This emphasizes that while accidents occur, modern aviation safety mechanisms significantly reduce the likelihood of fatal outcomes.

## ****5. Key Insights Summary****

* The overall number of aircraft accidents has **declined steadily** over time.
* **Summer months and travel peaks** correspond to higher accident frequencies.
* Some **countries and manufacturers** appear more frequently due to larger operational scales rather than inherent safety issues.
* Despite accident occurrences, **fatal crashes form a minority**, reflecting advances in aviation safety technologies and procedures.

## ****6. Conclusion****

This Project, leveraging **Streamlit** and **Altair** for interactive visualization, provides a clearer understanding of aircraft accident trends and their evolution.  
Through combining **Jupyter Lab** for analysis and **VS Code** for dashboard development, the project effectively demonstrates a complete workflow — from raw data to actionable insight.

The findings suggest consistent improvement in aviation safety globally, though specific trends related to regions, aircraft types, and travel seasons merit continued monitoring.

## ****7. Recommendations****

* Regulatory bodies should continue enhancing safety inspections during peak travel seasons.
* Future analyses could incorporate **flight volume data** to normalize accident counts per flight hour or distance.
* Implementing **machine learning predictive models** could assist in identifying risk factors for potential accident likelihood.