Aircraft Crash Data Analysis Report

By Timilehin OnasanyaS

# Overview

This report provides a summary of the analysis conducted on historical aircraft crash data. The goal of the project was to clean, filter, and visualize the dataset to uncover key insights about accident patterns, fatalities, survival rates, and other important metrics. The application developed uses Streamlit to create an interactive dashboard for real-time exploration of the data.

# Data Preparation and Cleaning

The raw dataset contained inconsistencies, missing values, and spelling errors. The following steps were taken to prepare the data for analysis:  
- Standardized column names by removing spaces and special characters.  
- Filled missing values in key fields such as Country/Region and Operator with 'Unspecified'.  
- Applied correction dictionaries to fix typos and inconsistent entries in columns like Country/Region, Aircraft Manufacturer, Aircraft, Location, and Operator.  
- Removed invalid records where fatalities were greater than the number aboard.  
- Reset the dataset index to maintain a clean structure.

# Interactive Filters

To allow for targeted exploration, the application includes sidebar filters. Users can filter the dataset based on:  
- Location  
- Country/Region  
- Aircraft Manufacturer  
- Aircraft  
- Operator  
  
These filters dynamically update the data and visualizations displayed in the dashboard.

# Key Metrics and Summary Statistics

The dashboard displays high-level metrics to give a quick overview of the dataset, including:  
- Year with the highest number of accidents.  
- Total fatalities aboard all aircraft.  
- Total fatalities on the ground.  
- Country/Region with the most accidents.  
- Location with the most crashes.  
- Aircraft manufacturer involved in the most accidents.

# Research Questions and Findings

The analysis addressed several key research questions using tables and visualizations:  
1. Top 10 countries/regions with the highest number of accidents.  
2. Top 10 aircraft manufacturers with the most accidents.  
3. Average fatalities aboard per aircraft type.  
4. Accident trends by year.  
5. Comparison of fatalities aboard versus on the ground.  
6. Changes in survival rates over time.  
7. Top 10 locations with the most accidents.  
8. Top 10 operators involved in crashes.  
9. Accident trends by quarter.  
10. Total accidents and unique manufacturers analyzed by decade.

# Conclusion

The analysis provided valuable insights into historical aircraft crash data. By cleaning and preparing the dataset, applying filters, and creating interactive visualizations, users can easily explore trends and patterns.