Table of Contents

[Air Crash Analysis Report 2](#_Toc212030914)

[Executive Summary 3](#_Toc212030915)

[Introduction 3](#_Toc212030916)

[Research Questions 3](#_Toc212030917)

[Data & Scope 4](#_Toc212030918)

[Methodology — Data Cleaning & Processing 4](#_Toc212030919)

[Dashboard & Visualizations 5](#_Toc212030920)

[Analysis & Findings 6](#_Toc212030921)

[Discussion 6](#_Toc212030922)

[Recommendations 6](#_Toc212030923)

[Conclusion 6](#_Toc212030924)

[References 6](#_Toc212030925)

[Appendix — Data Summaries 6](#_Toc212030926)

Air Crash Analysis Report

Angel Okoye  
Dataset: Air Crashes — Full Data (1908–2024)  
Course Project: Excel Data Analysis & Dashboard  
Date: October 2025

## Executive Summary

This project analyzes worldwide air crash incidents between 1908 and 2024 using a dataset obtained from Kaggle titled Air Crashes Full Data (1908–2023).  
The initial dataset contained over 5,000 records. After thorough cleaning; correcting misspellings, filtering out invalid or misplaced data, and trimming extra spaces, the final working dataset contained 4,347 accurate and consistent records ready for analysis.  
  
The study investigates trends in crashes over time, affected regions, aircraft manufacturers, and fatalities.  
The dashboard (created in Excel) includes interactive visuals showing crash trends per year, top affected countries, crashes by aircraft manufacturer, and fatalities by year.  
  
Key findings show that Douglas and Boeing aircraft appear most frequently in historical crash records, and Russia, Brazil, and Colombia have the highest country-level counts.  
  
The analysis provides insights that can help aviation authorities and operators improve safety practices and focus on high-risk categories.

## Introduction

Air crash analysis is an important part of aviation safety and risk management. By studying past crash records, analysts can uncover recurring causes, operational weaknesses, and manufacturer- or region-specific risk trends.  
  
This report forms part of an Excel-based data analysis project where the dataset was cleaned, analyzed, and summarized into an interactive Excel dashboard. The purpose is to:  
- Identify long-term patterns in air crash occurrences.  
- Highlight major manufacturers and countries with high crash frequencies.  
- Examine trends in fatalities and flight safety over time.  
- Draw meaningful recommendations for aviation safety improvements.

## Research Questions

1. How has the total number of air crashes changed over the years from 1908 to 2024?
2. What is the proportion of fatal versus non-fatal crashes, and how has this changed over time?
3. Which aircraft types are most frequently involved in crashes?
4. Which regions or countries record the highest number of crashes according to the dataset?
5. How do fatalities vary over time, and what periods show the highest fatality counts?
6. Which operators or airlines have recorded the most crashes in the dataset?
7. What are the observed monthly or seasonal patterns in air crash frequency?
8. How do air fatalities compare to ground fatalities, and what does this indicate about crash impact?
9. What is the overall trend in crash outcomes (fatal vs non-fatal) as aviation technology has advanced?
10. How does the distribution of crashes by region and aircraft type reflect global aviation safety patterns?

## Data & Scope

Data Source: Kaggle: Air Crashes Full Data (1908–2023)  
File Used: Excel workbook with multiple sheets (archive 3, pivot tables, dashboard).  
Initial Record Count: Over 5,000 rows (before cleaning).  
Final Record Count: 4,347 rows after cleaning.  
Columns: 15 (including Year, Month, Country/Region, Aircraft Manufacturer, Aircraft, Operator, Fatalities (air), Location, and Date). NB; The column date was inserted during the cleaning.  
Date Range: 1908 — 2024.  
Duplicates: None found.  
Missing Values: Very minimal (~2 per column on average).

## Methodology — Data Cleaning & Processing

The data cleaning process was crucial to ensure that the analysis and dashboard were based on accurate and consistent information.  
  
1. Correction and Replacement of Misspellings Using Power Query Editor  
- Corrected spelling errors in Columns most especially Country/Region column.  
- Unified inconsistent names using Replace values and researching right spellings.  
  
2. Filtering Out Incorrect Entries  
- Removed entries that didn’t correspond to the column name (e.g., numeric value ocean name under Country/Region column).  
- Ensured each column held valid, relevant data only.  
  
3. Cleaning and Trimming Extra Spaces  
- Applied TRIM() and CLEAN() functions to most column in the dataset to remove unnecessary spaces in text fields.  
  
4. Date and Format Validation  
- Ensured all Date values were valid and formatted properly.

5. Duplicate Check and Verification  
- Used Remove Duplicates tool (none found) and rechecked data consistency.

6. Merged Day, Month and Year column to get a standard date column.  
  
Result: Dataset reduced from over 5,000 to 4,347 valid records, all consistent and ready for analysis.

## Dashboard & Visualizations

The Excel dashboard includes:  
- Monthly Trend (Clustered Column Chart)  
- Regional Distribution (Stacked Bar Chart)  
- Crashes by Aircraft (Stacked Column Chart)  
- Fatalities over Time (Line Chart) etc.  
Each chart includes slicers for Year, Manufacturer, and Country.

A graph of a number of blue bars

AI-generated content may be incorrect. A blue rectangular object with text

AI-generated content may be incorrect. Monthly Trend Regional distribution(crashes)

A graph of a plane crash

AI-generated content may be incorrect. A graph showing the fall of a company

AI-generated content may be incorrect. Crashes by Aircraft Fatalities over time( yearly)

## Analysis & Findings

Key Findings:  
- Crashes Over Time: Peaks mid-20th century; declines in modern decades.  
- Top Countries: Russia (249), Brazil (164), Colombia (148), France (123), India (98), Indonesia (90), China (80).  
- Top Manufacturers: Douglas (871), Boeing (340), Antonov (247), Lockheed (177), De Havilland Canada (134), Ilyushin (129).  
- Fatalities: High in early decades; reduced in recent years due to safety improvements.  
- Data Quality: Excellent after cleaning, no duplicates, consistent naming, minimal nulls.

## Discussion

Crashes were more frequent during aviation’s early growth and declined due to improved technology, regulation, and pilot training.   
Higher crash counts in regions like Russia or Brazil reflect larger flight volumes and varied reporting quality. Cleaning the data eliminated category duplication and ensured accuracy in visualization.

## Recommendations

1. Improve naming consistency and standardize manufacturer labels.  
2. Add or map Cause data into categories (Human Error, Weather, Mechanical).  
3. Improve maintenance and pilot training focus in high-risk regions.  
4. Update the dataset yearly for ongoing trend monitoring.

## Conclusion

Air crash data from 1908–2024 shows clear improvement in aviation safety. Data cleaning: correcting misspellings, filtering incorrect entries, and trimming spaces ensured reliability.   
The final dashboard presents meaningful insights to guide aviation safety and risk prevention initiatives.

## References

- Dataset: Kaggle – Air Crashes Full Data (1908–2023)  
- Project Brief: EXCEL PROJECT Assignment Document (PDF)

## Appendix — Data Summaries

Original dataset: >5,000 rows  
Cleaned dataset: 4,347 rows  
Duplicate rows: 0  
Top Countries: Russia (249), Brazil (164), Colombia (148), France (123), India (98)  
Top Manufacturers: Douglas (871), Boeing (340), Antonov (247), Lockheed (177), De Havilland Canada (134)