Airplane Crashes Data Visualization

By Zhongyan Tang

Figure 1: Word Cloud of Crash Summaries

Figure 1 highlights the most frequently occurring terms in crash summaries from the dataset. The size of each word represents its frequency, with larger words like "crashed," "pilot," and "engine" emphasizing recurring themes of mechanical failures, pilot error, and adverse weather conditions. This visualization offers an immediate, holistic view of key factors contributing to airplane crashes over the decades.

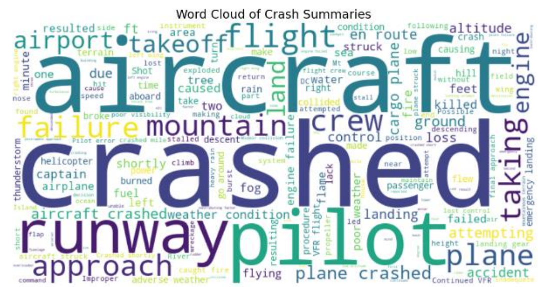


Figure 2: Distribution of Fatalities Per Crash

Figure 2 presents a histogram showing the distribution of fatalities per crash. Most crashes resulted in fewer than 50 fatalities, with a few extreme outliers exceeding 400. This highlights the generally low-fatality nature of aviation accidents but also underscores the devastating impact of large-scale crashes.

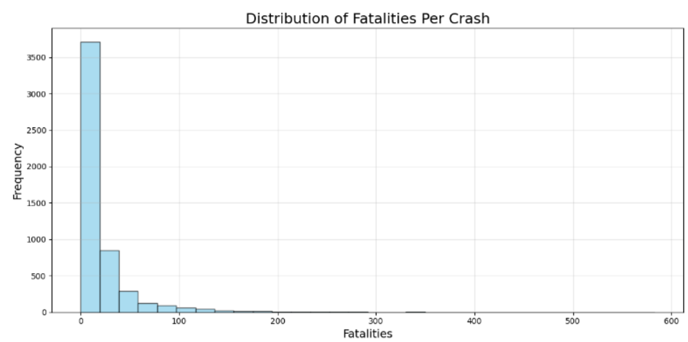


Figure 3: Fatalities vs. Aboard

Figure 3 explores the relationship between the total number of passengers aboard an airplane and the fatalities for each crash. Larger points indicate crashes with significant ground fatalities, while colors represent decades. The scatter plot reveals a general trend where larger aircraft (with more passengers aboard) often result in higher fatalities, emphasizing the heightened stakes of modern aviation.

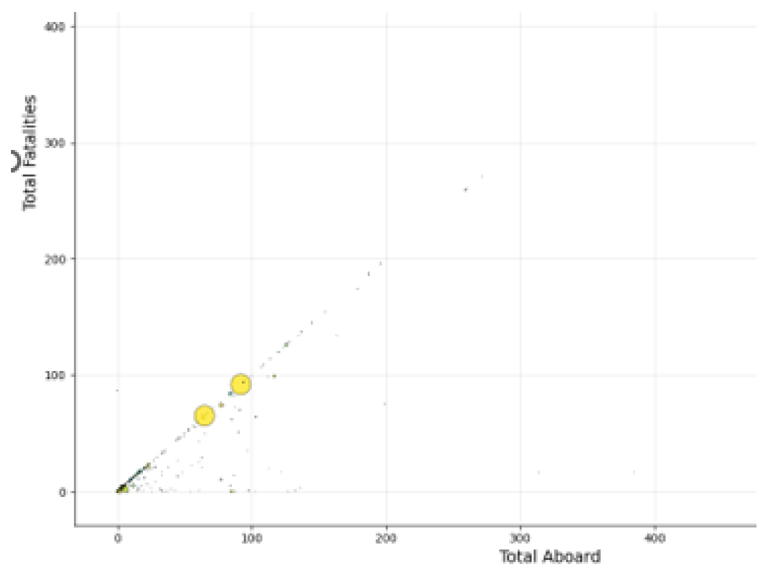


Figure 4: Top 10 Airplane Types in Crashes

Figure 4 is a pie chart displaying the top 10 airplane types involved in crashes. The Douglas DC-3 and DC-47 dominate the chart, reflecting their historical prominence in aviation. This insight highlights the need for continuous improvement in aircraft safety.

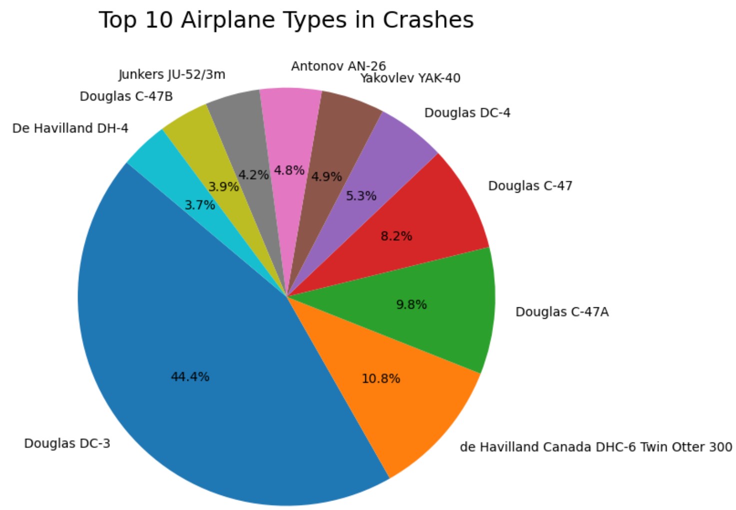
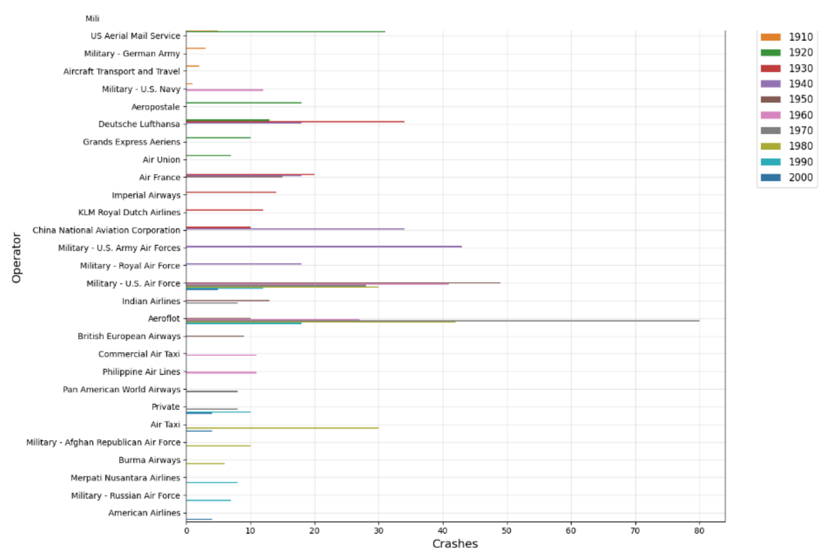


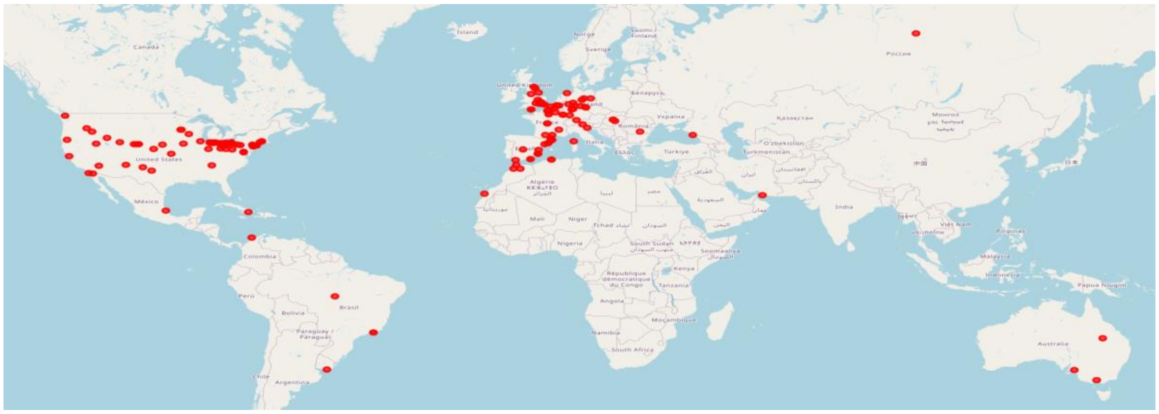
Figure 5: Top Operators by Crashes Per Decade

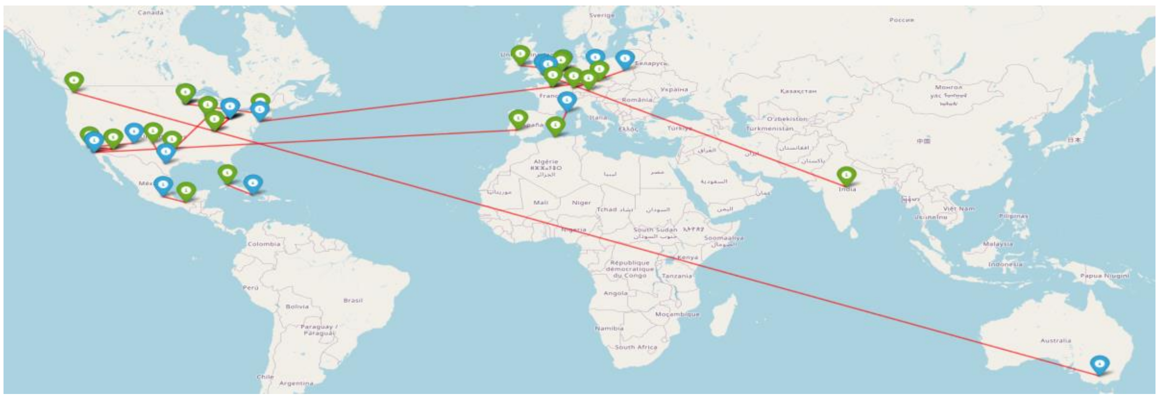
Figure 5 visualizes the top airline operators with the most crashes in each decade. This grouped bar chart shows a concentration of incidents among certain operators, such as military fleets and national carriers, emphasizing the risks associated with early aviation and wartime operations.



Combined Analysis of Global Distribution and Flight Routes

Figures 8 and 9 provide complementary perspectives on airplane crashes, focusing on their global distribution and flight routes. Figure 8 displays the worldwide locations of crashes, with dense clusters in North America and Europe, which historically served as major aviation hubs. Sparse data points in Africa and South America suggest lower aviation activity or potential underreporting in these regions. In contrast, Figure 9 visualizes the routes of flights involved in crashes, connecting departure and destination points. The red lines highlight long-haul transcontinental routes as frequent sites of incidents, likely reflecting the higher risks associated with extended flight durations and challenging international airspace management. Together, these visualizations underscore the significance of geographical and operational factors in aviation safety, with Figure 8 highlighting static crash locations and Figure 9 offering dynamic insights into flight trajectories and their associated risks.





Data and Methods

The dataset, sourced from public airplane crash records, contains information on crashes, fatalities, aircraft types, operators, and locations. Python libraries such as pandas, matplotlib, seaborn, folium, and wordcloud were used to process and visualize the data. Geocoding was performed using the OpenStreetMap API to map crash locations and flight routes.

Importance

This report provides critical insights into aviation safety by analyzing airplane crashes through diverse visualizations. Key factors like mechanical failures, pilot error, and adverse weather emerge as recurring themes. Quantitative analyses highlight the impact of large-scale crashes and temporal trends, showing advancements in safety over decades. Geospatial maps reveal clusters in major aviation hubs and risks associated with long-haul routes, offering a comprehensive understanding of operational and geographical complexities. By integrating these findings, the report serves as a valuable resource for enhancing aviation safety standards and guiding risk mitigation strategies.

Dataset:https://www.kaggle.com/datasets/imtkaggleteam/airplane-crashes?resource=download

Github:https://github.com/DavidTang1012/Plane-Crash-Data-visualization