A Data-Driven Study of Aviation Accidents, Fatalities, and Safety Trends

An exploratory data analysis based on historical aviation accident records

Rodrigo Driemeier dos Santos

Independent Data Analysis Project

December 24, 2025

Table of Contents

1. Objectives of the Study

The objective of this study is to conduct a rigorous, data-driven exploration of aviation accidents and fatalities over the period 1919–2023, with the purpose of identifying long-term safety trends, structural changes in accident characteristics, and the historical factors that shaped aviation risk. Rather than focusing on isolated events, the analysis aims to place aviation safety within a broad temporal framework, allowing patterns to emerge across different technological, operational, and regulatory eras.

A key goal of the study is to distinguish between exceptional historical disruptions and underlying structural trends. In particular, the analysis investigates how World War II and the events surrounding 2001 represent fundamentally different types of safety shocks, one driven by sustained military exposure and the other by isolated acts of unlawful interference. These periods are examined not only to confirm their historical relevance but also to understand how they differ from the long-term evolution of aviation safety.

Another central objective is to critically assess the reliability and interpretability of historical aviation data. Given the heterogeneous nature of the dataset, the study places strong emphasis on data cleaning, validation, and consistency checks prior to analysis. This includes evaluating missing data patterns, verifying historical plausibility, and redefining categorical variables where necessary to ensure that analytical conclusions are based on coherent and meaningful indicators.

Beyond accident frequency, the study aims to evaluate safety improvements through multiple complementary dimensions. This includes analyzing fatality trends, deriving a quantitative proxy for accident severity, and examining how the distribution of accident types evolved over time. By integrating absolute counts with relative and structural measures, the study seeks to move beyond simplistic interpretations and provide a more nuanced understanding of aviation safety progress.

Ultimately, the objective is not to produce predictive models or risk forecasts, but to offer a historically grounded, analytically robust perspective on how aviation safety has evolved. The study aims to demonstrate that meaningful safety improvements can be identified only when absolute numbers are interpreted alongside severity, proportional distributions, and long-term structural change.

2. Dataset

2.1. Dataset Presentation and Source

The dataset used in this study is derived from the Aviation Safety Network (ASN) and provides structured information on civil and military aviation occurrences worldwide. Each record corresponds to a single aviation event and combines temporal, geographic, operational, and categorical attributes. The core temporal variable is the year of occurrence, which enables long-term trend analysis, while spatial information is represented by the country field. Operational context is further described through variables such as aircraft type, operator, and registration, allowing the dataset to capture aspects related to fleet composition, technological evolution, and organizational factors. The fatalities variable records the total number of deaths associated with each event, including passengers, crew members, and, when applicable, ground victims, making it a comprehensive outcome-oriented measure.

Accident classification follows the ASN event taxonomy, in which categories are intended to describe the nature of the event rather than its severity. Category A (Accidents) includes unintended events such as crashes or loss of control; O (Operational occurrences) refers to technical or procedural issues that may or may not escalate into accidents; C (Criminal acts) encompasses deliberate acts such as sabotage; H (Unlawful interference) covers hijackings and related security breaches; I (Incidents) denotes minor occurrences with limited operational consequences; and U (Unknown) is used when insufficient information prevents reliable classification. These categories are not ordinal and do not imply increasing or decreasing severity. Their role is to characterize what happened, not how severe the outcome was, a distinction that directly informs the methodological choices adopted in later sections of this report.

The data were obtained from a publicly available aviation accident database compiled from multiple historical and institutional sources and made accessible through the Kaggle platform. The dataset can be consulted online, with full access information provided in the reference section [1]. Given its aggregated and historical nature, the dataset reflects heterogeneous reporting standards, varying levels of data completeness, and evolving classification practices across different time periods. Earlier records, particularly those from the first half of the twentieth century, tend to contain less detailed information and a higher proportion of missing or ambiguous values, whereas more recent entries benefit from improved documentation, standardized reporting procedures, and enhanced data availability.

Each row in the dataset represents a single aviation occurrence and combines quantitative information, such as the number of fatalities, with qualitative descriptors related to event classification and aircraft characteristics. Due to its long temporal span and heterogeneous origins, the dataset is well suited for exploratory and descriptive analysis, provided that appropriate data cleaning and reliability checks are performed. Accordingly, the dataset is treated not as a source for precise accident rate estimation, but as a historical record of reported aviation events. The analyses presented in this study therefore focus on relative trends, temporal patterns, and contextual interpretation rather than absolute statistical inference, and all subsequent preprocessing and validation steps are conducted with these considerations in mind.

2.2 Data Cleaning and Preprocessing

All data cleaning, preprocessing, and validation steps were conducted using the Python programming language, leveraging widely used data analysis libraries. Pandas was employed as the primary tool for data manipulation, filtering, transformation, and consistency checks, while NumPy was used for numerical operations and handling missing values. These libraries provided an efficient and reproducible framework for preparing the dataset prior to analysis. The same computational environment was subsequently used for exploratory analysis and data visualization, with graphical outputs generated using Matplotlib. This unified workflow ensured consistency between data preparation, analysis, and visualization stages throughout the study.

Given the size and complexity of the dataset, selected code excerpts are presented throughout this section to illustrate key preprocessing steps, particularly those related to data cleaning and reliability verification. These excerpts are intended to clarify the methodological decisions adopted prior to analysis. In contrast, the full implementation of the analytical procedures and visualization routines is not reproduced in this document and is instead made available in a public repository for reproducibility and transparency purposes [2].

2.2.1 Duplicate Removal

To begin with, duplicate records were removed only when all columns matched exactly another row. Since several attributes, such as aircraft model, number of fatalities, and year, can plausibly coincide across different events, a record was considered a true duplicate (and therefore subject to removal) only when all variables matched simultaneously, including the accident date.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

db = pd.read_csv("aviation-accident-data-2023-05-16.csv")

db = db.drop_duplicates()
```

2.2.2 Removal of Unnecessary Columns

In order to streamline the dataset and maintain analytical focus, a set of variables deemed non-essential for the objectives of this study was removed during the data cleaning process. The guiding principle behind this decision was not merely to reduce dimensionality, but to ensure consistency between the retained variables and the analyses effectively conducted throughout the report.

As the core of the study relies on long-term temporal patterns, accident categorization, fatalities, and broad spatial context, certain columns, although informative in isolation, were not directly leveraged in subsequent analyses. Variables such as registration, operator, and precise location were therefore excluded, as they were neither required for the statistical evaluations performed nor for the historical and structural interpretations developed later in the report. Retaining such fields would introduce unnecessary complexity without providing additional analytical value.

```python
cols_to_drop = [
    "location",
    "registration",
    "operator"
]

db.drop(columns=cols_to_drop, inplace=True)
```

A similar rationale motivated the removal of the date column. Since the analyses are conducted at a yearly resolution, retaining day and month information would increase data granularity without improving interpretability. Prior to dropping this column, a verification step was carried out to ensure that no record contained valid year information exclusively embedded in the date field. Specifically, the dataset was checked for cases where a valid date existed alongside a missing or unknown year value. As no such inconsistencies were found, the year column was confirmed to be a complete and sufficient temporal reference, allowing the date column to be safely removed.

This selective column removal ensures that the dataset remains concise, internally coherent, and aligned with the analytical scope of the study. By retaining only variables that are actively used and methodologically justified, the preprocessing step reduces noise, improves clarity, and strengthens the interpretability of the results presented in subsequent sections.

By printing the 5 first lines, we can verify that the date column, along with the others, were dropped.

```python
# If all unkown dates corresponds to an unknown year then we can simply use year column and drop the date column
cond = (db["year"].isna()) & (db["date"] != 'date unk.')

if cond.sum() == 0:
    db.drop("date", axis = 1, inplace = True)

print(db.head(5))
```

```
             type fatalities          country cat     year
  Antonov An-12B        NaN  Unknown country  U1  unknown
  Antonov An-12B        NaN  Unknown country  U1  unknown
  Antonov An-12B        NaN  Unknown country  U1  unknown
 Antonov An-12BK        NaN           Russia  A1  unknown
 Antonov An-12BP          0          Eritrea  A1  unknown
```

2.2.3 Standardization of Unknown Values as NA

Native pandas functionality was used to convert unknown or ambiguous values into standardized NA representation. For each variable, the specific formats in which missing or unknown information appeared were explicitly identified and handled according to the logic of the dataset. This approach ensured that semantically equivalent unknown values were consistently treated across different columns.

For the fatalities column, missing values were not explicitly converted at this stage. Instead, the NA representation was implicitly applied later during the conversion of the column to a nullable integer type, as discussed in the following subsection.

```
db.loc[db["year"] == 'unknown', "year"] = pd.NA

# Unknow country standard: NA

db.loc[(db["country"] == '?') | (db["country"] == 'Unknown country'), "country"] = pd.NA
```

2.2.4 Year and Fatalities Conversion to Integer Format

Prior to processing the fatalities variable, the year column was converted to a numeric format. Since the year information was already isolated and consistently represented after the removal of the date column, this conversion required no additional transformation beyond casting the variable to a nullable integer type. This step ensured that the year could be safely used in numerical operations and grouping procedures while preserving missing values through the standardized NA representation.

Finally, the fatalities variable was standardized to ensure numerical consistency while preserving the integrity of the original data. No artificial imputation was applied to missing values, as replacing unknown fatality counts with zeros or statistical estimates could lead to misleading conclusions in subsequent analyses. Instead, only records with explicitly reported fatality information were considered during quantitative evaluations.

In some cases, fatality counts were recorded as composite values (for example, "1+2"), indicating multiple deaths associated with the same event, including delayed fatalities attributable to the accident. These entries were systematically processed by splitting the composite values and summing their components, producing a single integer value that accurately reflects the total number of fatalities for each event.

After this transformation, the fatalities column was converted to Pandas' nullable integer type (Int64). This conversion implicitly standardized missing values to the NA format while ensuring that all valid fatality counts were stored as integers, allowing for consistent numerical analysis in subsequent sections.

```
# Year must be an int

db["year"] = db["year"].astype(dtype = "Int64")

# For fatalitites, we're not artificially filling NaN with 0 or with the mean, since it will lead to unprecise final statement
# (when we start doing the analysis, we're gonna consider only fatality data different from NaN)
# Fatalitites must be an int (here, if we have 1 + 2, we will sum it and make it 3 - considering late deaths caused by the accident are indeed an accident fatality)

db["fatalities"] = db["fatalities"].str.split("+")

for idx, fatality in db["fatalities"].items():
    if isinstance(fatality, list):
        if pd.notna(fatality).all():
            db.at[idx, "fatalities"] = sum(int(i) for i in fatality)


db["fatalities"] = db["fatalities"].astype("Int64")
```

2.3 Data Conversion and Reliability Check

Before proceeding with the exploratory analysis, a set of data conversion and reliability checks was performed to ensure the internal consistency and interpretability of the dataset. These steps aimed to verify that key variables were correctly formatted, semantically coherent, and suitable for quantitative and historical analysis. Particular attention was given to identifying potential information loss during preprocessing, assessing missing data patterns, and validating the coherence of categorical classifications within their historical context. This preliminary validation stage was essential to define the scope and limitations of the dataset and to ensure that subsequent analyses were based on reliable and well-understood data.

2.3.1 Consistency Analysis of Accident Categories

As part of the data reliability assessment, the accident category variable (cat) was evaluated to determine whether it could be reliably interpreted as an indicator of accident severity. This analysis initially compared accident categories against recorded fatality counts by examining both the most fatal events in the dataset and events with zero recorded fatalities.

```
# Looking for inconsistencies between accident's category and it's number of fatalities
test_cat_fat_top = db.sort_values(by = "fatalities", ascending =  False).head(5)
test_cat_fat_low = db[db["fatalities"] == 0]
print(test_cat_fat_top.loc[:, ["fatalities", "cat"]])
print(test_cat_fat_low.loc[:, ["fatalities", "cat"]])
```

The results revealed clear inconsistencies: some of the most fatal accidents were classified under categories such as H1, while several non-fatal events appeared under categories typically associated with severe accidents, such as A1. This lack of alignment demonstrates that the accident category variable does not consistently encode accident severity or lethality.

fatalities	cat
1692	H1
965	H1
520	A1
346	A1
335	A1

fatalities	cat
0	A1
0	A2
0	A1
0	O1
0	A1
...	..
0	A2
0	A2
0	A2
0	A1
0	A2

Given these inconsistencies, the analysis was extended to assess whether the category variable could still be considered reliable as a qualitative descriptor of event type, rather than a proxy for severity. In particular, events classified as H1 and H2 were isolated and examined independently of their fatality counts. These events were analyzed in relation to aircraft type, year, and contextual information, and selected cases were further investigated using external aviation safety records.

```
# We investigate, even though 'cat' isn't coherent with the number of fatalities, if it is still coherent with the event itself
test_cat_valid_for_event = db[(db["cat"] == 'H1') | (db["cat"] == 'H2')]
print(test_cat_valid_for_event.loc[:, ["year", "type"]])
```

The observed distribution spans both early aviation periods, featuring aircraft such as the Ford Tri-Motor, Douglas DC-3, and C-47, and modern civil aviation, including aircraft such as the Boeing 737, Cessna 208B Grand Caravan, and Embraer models. This temporal and technological diversity indicates that the H1/H2 classification is not confined to a specific era or aircraft class, but is consistently applied to passenger-capable aircraft across different historical contexts. The coherence between event category, aircraft type, and time period supports the interpretation of the category variable as a qualitative indicator of intentional human interference, rather than as an artifact of data entry errors or period-specific misclassification.

[10705 rows x 2 columns]

	year	type
167	1931	Ford Tri-Motor
5910	1948	Douglas DC-3
5931	1948	Siebel Si 204
5957	1948	Douglas C-47 (DC-3)
5964	1948	Douglas C-47 (DC-3)
...	...	...
23136	2019	Boeing 737-8E9 (WL)
23332	2019	Cessna 208B Grand Caravan
23592	2021	Fokker 100
23602	2021	Embraer
23649	2021	Cessna 208B Grand Caravan

By specifically consulting some accidents shown in the table plotted, it became clear that the category is in fact coherent with its events. For instance, an incident documented on 25 March 2021 involved an Embraer aircraft operated by Mauritania Airlines International at Nouakchott–Oumtounsy International Airport (Mauritania)**.** In this case, an individual unlawfully boarded the aircraft and threatened to set it on fire, but was restrained before any fatalities occurred. The event was classified as unlawful interference rather than a conventional hijacking, illustrating the broader application of the H1/H2 classification framework [3].

Similarly, another documented incident on 7 July 2021 involved a Cessna 208B Grand Caravan operating on a domestic flight in Alaska, United States. During the flight, a passenger attempted to take control of the aircraft and deliberately crash it. The attempt was unsuccessful and no fatalities were recorded; however, the event was categorized as unlawful interference due to the intentional nature of the act. This case further demonstrates that H1/H2 classifications are applied to attempted or incomplete acts of interference, rather than exclusively to successful hijackings or events defined by high fatality counts [4].

Taken together, these findings indicate that the accident category variable should not be interpreted as a reliable measure of accident severity. However, when used as a flexible qualitative indicator of event nature, particularly to distinguish events involving unlawful or intentional human interference from operational or technical accidents, the variable remains coherent and informative. This distinction is critical for subsequent analyses, as it defines the appropriate scope within which accident categories can be meaningfully used in this study.

While this validation focused explicitly on categories H1 and H2, this choice was deliberate rather than restrictive. These categories present a specific ambiguity, as their labels may suggest severity (e.g., terrorism or hijacking) while, in practice, they encompass a broader range of unlawful interference events with highly variable outcomes. No comparable ambiguity was identified for the remaining categories. Accident (A), Operational (O), Criminal (C), and Incident (I) classifications are semantically stable and consistently describe the nature of the event itself rather than its consequences. Therefore, these categories can be reliably employed in subsequent temporal and structural safety analyses without requiring additional event-level validation.

2.3.2 Detection of Years with Abnormally Low or Missing Data

As part of the dataset reliability assessment, the temporal distribution of records was examined to identify potential years with insufficient or missing information. This analysis was conducted in two stages: first, by verifying whether any year contained no recorded accidents at all, and second, by evaluating the distribution of missing values in the fatalities variable across years.

```
# Analyzing if there are years with no accidents
accidents_per_year = db.groupby("year").size().reset_index(name="accidents_per_year")
cout = 0
for i in accidents_per_year["year"]:
    if i == 0:
        cout+= 1
print(f"We found {cout} years with no accidents")

# Analyzing if there is any year with an outlier number of NAs in 'fatalities': may indicate we have few data for that year...
nas_fat_per_year = db["fatalities"].isna().groupby(db["year"]).sum().reset_index(name = "quantity_of_NAs").sort_values(by = "quantity_of_NAs").tail(10)
print(nas_fat_per_year)
```

The analysis of accident counts per year showed that no year in the dataset contains zero recorded accidents. This result indicates that, at a yearly resolution, the dataset provides continuous coverage throughout the entire temporal range considered. From a reliability perspective, this finding suggests that there are no completely unrepresented years that could introduce artificial discontinuities or gaps in long-term trend analyses.

```
We found 0 years with no accidents
```

Subsequently, the number of missing values in the fatalities column was computed for each year in order to detect years with an unusually high proportion of unavailable fatality data. The results reveal a clear concentration of missing fatality records during the early and wartime periods, particularly between 1940 and 1945, with the highest counts observed in 1944 and 1945. Earlier years such as 1940–1943 also exhibit elevated levels of missing data when compared to later decades.

year	quantity_of_NAs
1951	41
1969	45
1941	53
1947	54
1946	84
1940	96
1942	136
1943	342
1944	830
1945	913

This pattern is consistent with historical expectations. During World War II, aviation activity increased substantially, particularly in military contexts, while systematic record-keeping and standardized accident reporting were limited or secondary to operational priorities. As a result, while the occurrence of aviation events is well represented in the dataset, detailed fatality information is often incomplete or unavailable for this period.

Importantly, this observation does not indicate a structural flaw in the dataset, but rather reflects historical reporting limitations inherent to early and wartime aviation data. Recognizing this limitation is essential for interpreting subsequent analyses. In particular, quantitative comparisons of fatality counts across years must account for the higher proportion of missing data during these periods, and conclusions regarding accident severity in early decades should be framed cautiously.

Overall, this reliability check confirms that the dataset provides consistent yearly coverage, while also highlighting specific historical periods, most notably World War II, where fatality data completeness is reduced. These findings inform the analytical scope of the study and guide the interpretation of temporal trends presented in later sections.

2.3.3 Outlier Analysis in Fatality Counts

To assess the presence of potential data entry errors or implausible values, an outlier analysis was conducted on the fatalities variable by examining the highest fatality counts recorded for individual aviation events. The objective of this analysis was to verify whether extreme values reflected data inconsistencies or corresponded to historically and physically plausible events.

```
# Verifying max value for fatility in one accident, in order to be sure there is no absurd value
print(db.dropna(subset = ["fatalities"]).sort_values(by = "fatalities").tail(10).loc[:, ["year", "type", "fatalities"]])
```

The inspection of the ten most fatal accidents shows that elevated fatality counts are strongly associated with large-capacity aircraft operating in high-density contexts. Accidents involving aircraft such as the Douglas DC-10, Boeing 727, and Boeing 747 exhibit fatality counts in the range of several hundred, which is consistent with their seating capacities and operational roles. For example, the 1985 Boeing 747SR-46 accident, with over five hundred fatalities, corresponds to a well-documented catastrophic event involving a high-density aircraft configuration and therefore does not constitute an implausible outlier.

The most extreme fatality values appear in 2001, involving two accidents with Boeing 767 aircraft, with recorded fatality counts of 965 and 1,692, respectively. These values significantly exceed typical aviation accident fatality levels; however, they are not indicative of data errors. In these cases, the fatality counts include not only passengers and crew onboard the aircraft, but also victims on the ground, resulting from impacts with occupied buildings. When ground casualties are considered, the magnitude of these values becomes historically and contextually consistent with the events they represent.

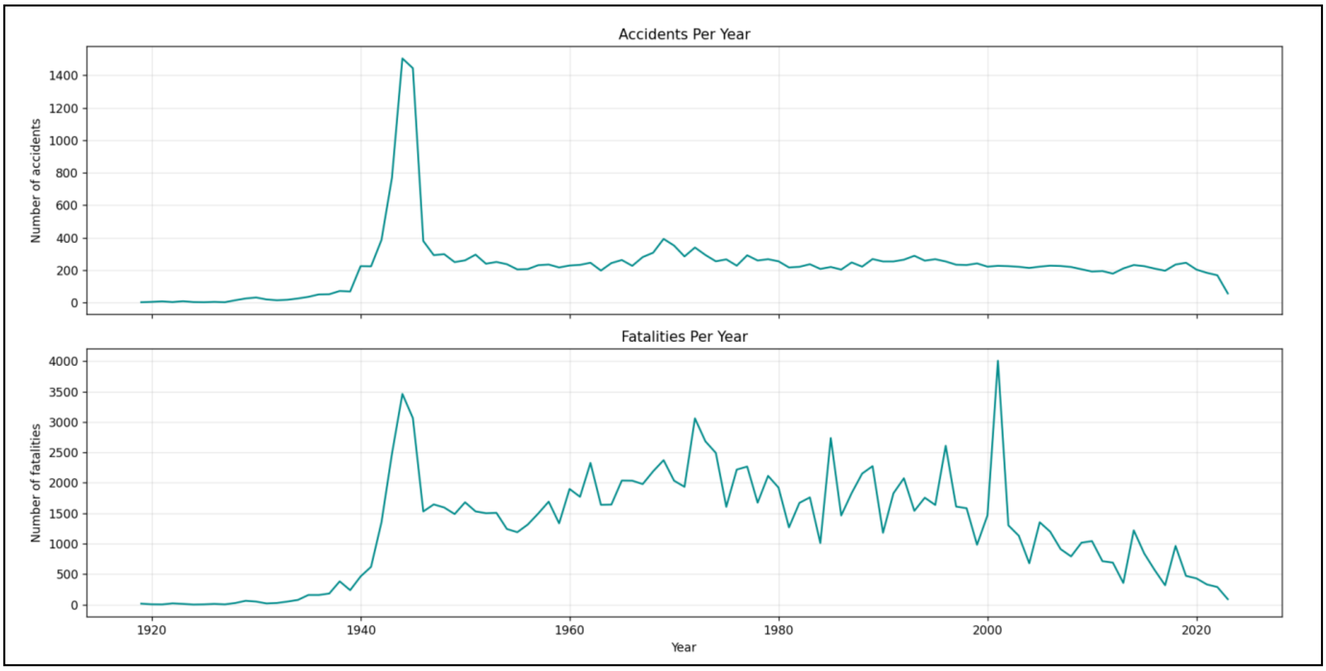
Importantly, none of the observed fatality counts exceed what would be physically plausible when accounting for aircraft capacity, operational context, and external casualties. The progression of fatality magnitudes aligns with known characteristics of the aircraft involved and with documented historical events. As a result, no records were identified as erroneous or requiring removal based on extreme fatality values.

year	type	fatalities
1988	Airbus A300B2-203	290
2014	Boeing 777-2H6ER	298
1980	Lockheed L-1011 TriStar 200	301
1996	Boeing 747-168B	312
1985	Boeing 747-237B	329
1977	Boeing 747-121	335
1974	DC-10-10	346
1985	Boeing 747SR-46	520
2001	Boeing 767-222	965
2001	Boeing 767-223ER	1692

This analysis confirms that the upper tail of the fatality distribution reflects genuine historical events, including rare but exceptional circumstances, rather than data quality issues or inflated values. Consequently, these observations were retained in the dataset and are interpreted as meaningful anomalies that provide critical insight into aviation risk under extreme conditions, rather than as statistical noise.

2.3.4 Basic Historical Consistency Check

As a final reliability check, the temporal evolution of both accident frequency and fatality counts was analyzed to assess the overall historical coherence of the dataset. The number of accidents per year and the total number of fatalities per year were plotted over the full time span covered by the data in order to verify whether the observed trends align with well-established historical patterns in aviation history.



The results show a pronounced and coherent peak in both accident frequency and fatalities during the early 1940s, corresponding to World War II. This behavior is consistent with the sharp increase in aviation activity during wartime, particularly in military operations, combined with limited safety standards and incomplete reporting practices. The alignment of both curves during this period reinforces the historical plausibility of the data.

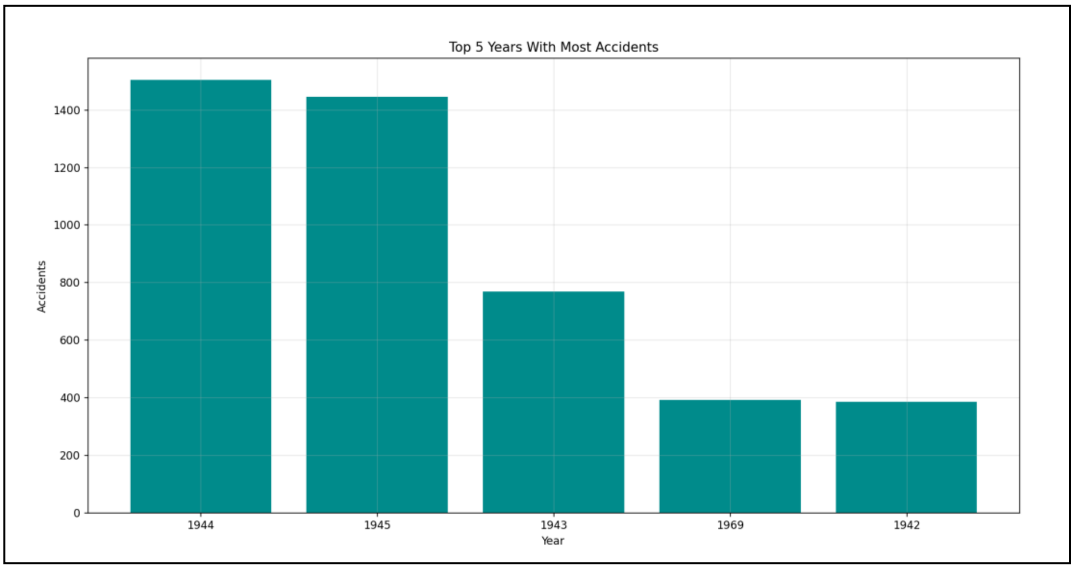
Following the post-war period, the dataset exhibits a stabilization in the number of recorded accidents per year, alongside a long-term decline in fatality counts, despite sustained aviation activity. This divergence between accident frequency and fatalities is consistent with the progressive improvement of aircraft technology, operational procedures, and regulatory frameworks, which led to enhanced survivability and overall safety.

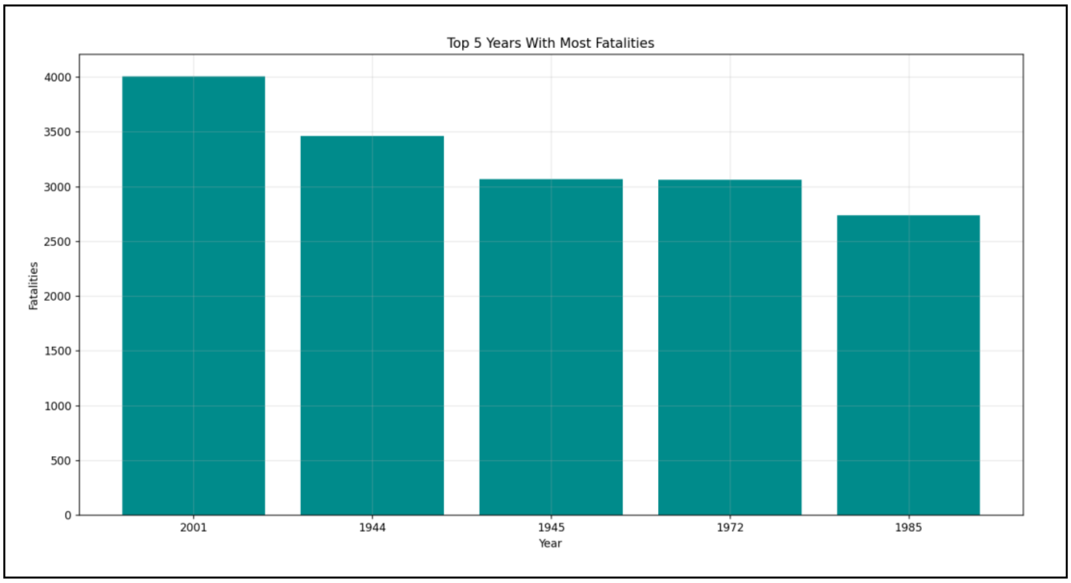
A distinct spike in fatalities is observed around 2001, while the number of accidents does not increase proportionally. This pattern reflects exceptional historical events involving intentional human interference rather than a general deterioration of aviation safety, further supporting the internal consistency of the dataset.

Overall, the temporal evolution of accidents and fatalities aligns with known historical events and long-term safety trends in aviation. This consistency provides additional confidence that the dataset captures meaningful historical dynamics rather than artifacts of data collection or processing, thereby supporting its use for subsequent analytical sections.

3. Investigation of Historical Periods

Building upon the historical consistency analysis, the identification of years with the highest number of accidents and fatalities provides a clear basis for selecting the most relevant periods for deeper investigation. The bar charts highlighting the top five years by accident count and by fatality count reveal a strong concentration of extreme values during World War II (1943–1945) and in 2001. These peaks confirm that these periods represent exceptional disruptions in the historical evolution of aviation safety, driven by fundamentally different mechanisms, large-scale military operations in the former case and intentional unlawful interference in the latter. Given their prominence in both accident frequency and fatality magnitude, these periods are examined separately in the following sections in order to better understand the specific characteristics, contributing factors, and safety implications associated with each context.

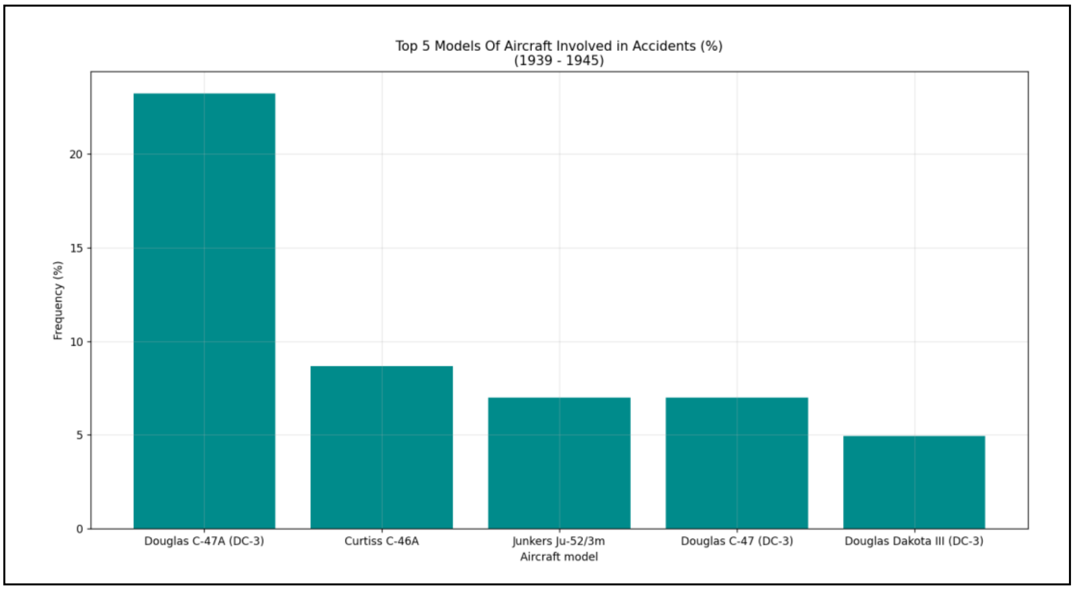




3.1 World War II (1939–1945)

3.1.1 Aircraft Types Most Frequently Involved in Accidents

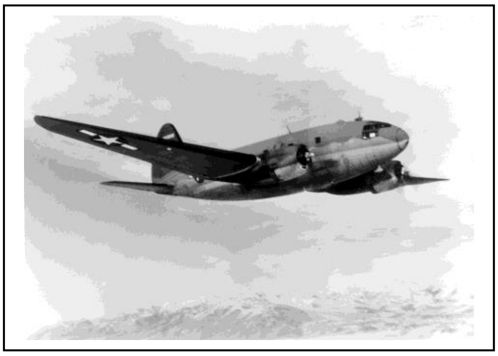
The distribution of aircraft models involved in accidents during the World War II period (1939–1945) provides important contextual insight into the nature of aviation activity at the time. By restricting the analysis to this period and examining the relative frequency of aircraft types involved in accidents, it becomes evident that the dataset is overwhelmingly dominated by military and military-derived aircraft.



The results show that models such as the Douglas C-47 / DC-3, Curtiss C-46A, and Junkers Ju-52/3m account for a substantial share of recorded accidents during this period. These aircraft were extensively used for troop transport, cargo missions, and logistical operations throughout the war, often operating under adverse conditions, high operational tempo, and limited safety margins. Their prominence in the accident records therefore reflects exposure and usage intensity rather than inherent design deficiencies.

The strong concentration of accidents among a small set of aircraft models further reinforces the historical coherence of the dataset. Civil aviation played a comparatively minor role during this period, and the dominance of military transport aircraft in the accident distribution is consistent with known patterns of wartime aviation activity. As such, this result supports the interpretation that the observed accident patterns during World War II are driven primarily by operational context and mission profile, rather than by random variability or data artifacts.

Douglas C-47 (DC-3)



Curtiss C-46A

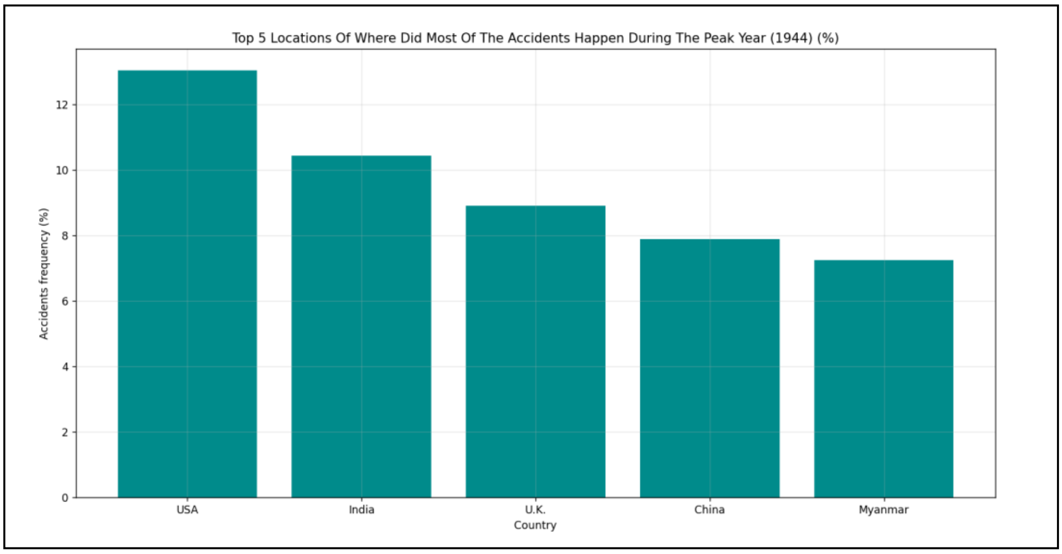
Junkers Ju-52/3m

Douglas C-47A (DC-3)

Douglas Dakota III (DC-3)

3.1.2 Locations with the Highest Accident Occurrence

To further characterize the aviation accident landscape during World War II, the geographical distribution of accidents was examined by focusing on 1944, the year with the highest number of recorded accidents and one of the peaks in fatalities. By analyzing the relative frequency of accidents by country during this peak year, it is possible to identify the regions most exposed to aviation activity and operational risk at the height of the conflict.



The results show that accidents were predominantly concentrated in the United States, followed by India, the United Kingdom, China, and Myanmar. This distribution closely mirrors the geographical scope of large-scale military operations, logistics networks, and training activities during the later stages of the war. The prominence of the United States reflects its extensive role as a production hub, training ground, and logistical base for Allied air operations, involving a high volume of domestic and overseas flights.

Similarly, the presence of India, the United Kingdom, China, and Myanmar can be explained by their strategic importance as operational theaters, transit corridors, and supply routes, particularly within the Asia-Pacific and European fronts. These regions hosted intense military air activity, often under challenging environmental and operational conditions, which naturally increased accident exposure.

Importantly, these results should not be interpreted as indicators of country-specific safety performance. Instead, they primarily reflect exposure-driven risk, where accident frequency scales with the intensity of aviation operations rather than with inherent deficiencies in infrastructure or procedures. The geographical accident distribution observed for 1944 is therefore consistent with known historical patterns of wartime aviation activity and further supports the historical coherence of the dataset.

3.2 The Era of Terrorism and the 2001 Attacks

3.2.1 Interpretation of H1 and H2 as Event-Type Categories

Before conducting any temporal or impact-based analysis of aviation-related terrorism, it is essential to clearly define the meaning and analytical scope of accident categories H1 and H2 within the dataset. As previously discussed in the reliability assessment phase, these categories do not provide a consistent or reliable measure of accident severity. Instead, they function primarily as indicators of the nature of the event.

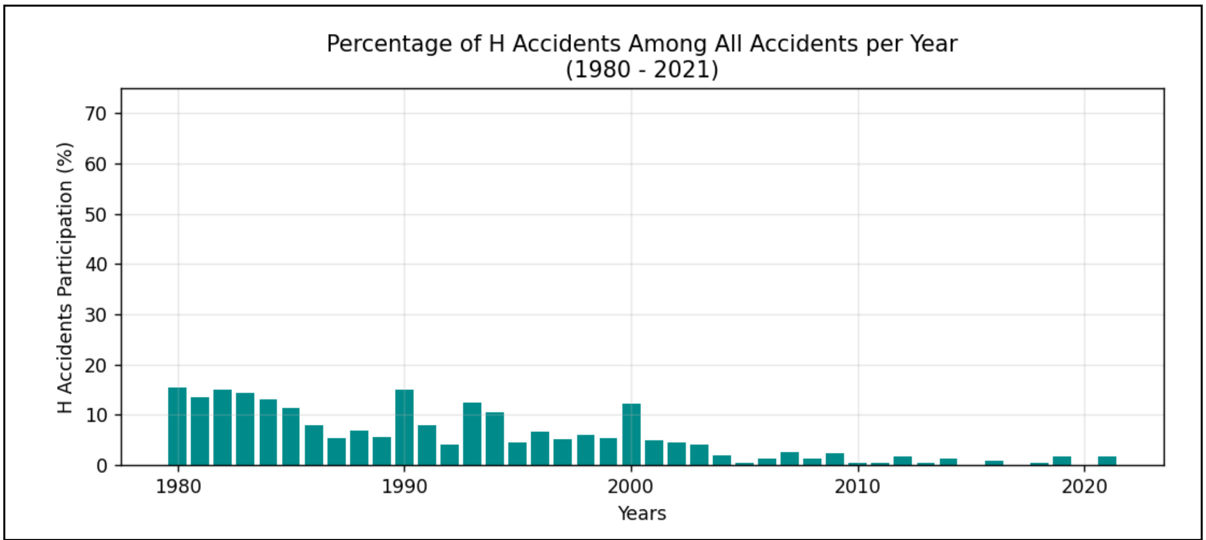
Consequently, within the scope of this study, categories H1 and H2 are reinterpreted as a single aggregated class, denoted as H, representing unlawful interference events. This aggregation removes unnecessary granularity while preserving the core semantic meaning of the classification. The distinction between H1 and H2, originally intended to reflect degrees of severity, is not analytically useful here, given the dataset's demonstrated limitations in encoding severity reliably.

This methodological choice allows the analysis to focus on event characterization rather than outcome magnitude. In other words, category H is used to answer questions related to when and how often unlawful interference events occurred, rather than how deadly they were on average. Severity-related conclusions are instead derived directly from fatality counts, which offer a more objective quantitative measure.

3.2.2 Temporal Evolution of Unlawful Interference Events (1980–2021)

With H1 and H2 redefined as a single category representing unlawful interference events, the analysis now focuses on their temporal behavior within the context of modern aviation. To ensure relevance and data consistency, the period from 1980 onward was selected, corresponding to an era of increased commercial aviation activity and more systematic accident reporting.

Rather than analyzing raw counts, the evolution of unlawful interference events was examined through their annual proportion relative to the total number of aviation accidents. This normalization is essential, as it accounts for the substantial growth in global air traffic over time and prevents misleading conclusions driven solely by volume effects.



The resulting trend reveals that unlawful interference events consistently represent a small fraction of total aviation accidents throughout the entire period. While moderate fluctuations are observed during the 1980s and early 1990s, there is no evidence of a sustained upward trajectory. On the contrary, the proportion of such events declines progressively after the mid-1990s and remains marginal in the decades that follow.

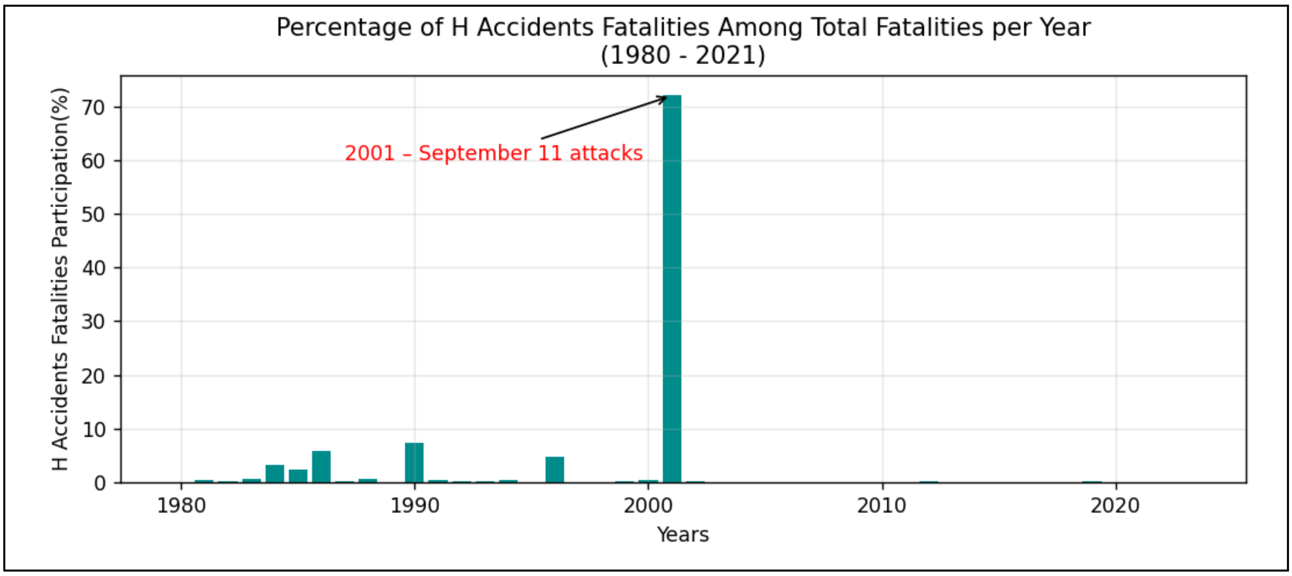
Crucially, the year 2001 does not correspond to a peak in the frequency of unlawful interference events. Despite its undeniable historical significance, the data show that the relative occurrence of category H accidents in that year remains comparable to surrounding periods. This indicates that the 2001 attacks did not reflect a structural increase in the prevalence of aviation-related unlawful interference, but rather an isolated and exceptional episode.

Overall, the temporal analysis demonstrates that unlawful interference events have remained rare and non-persistent within the broader accident landscape. Their historical relevance lies not in their frequency, but in their potential consequences, an aspect explored in the following section through an impact-based analysis of associated fatalities.

3.2.3 Fatality Impact and the 2001 Outlier

While unlawful interference events remain rare in terms of frequency, their relevance becomes evident when the analysis shifts from occurrence to fatality impact. To assess this dimension, the contribution of category H events to the total number of aviation-related fatalities per year was examined for the same period (1980–2021).

To ensure that this impact-based analysis accurately reflects the relative contribution of unlawful interference events, particular attention was given to the treatment of missing data. Years in which no fatalities were associated with category H events naturally result in undefined proportions when compared to total annual fatalities. Rather than excluding these years or artificially inflating their relevance, such cases were explicitly treated as zero-impact years, preserving their informational value within the temporal context. This approach allows the analysis to distinguish between the absence of impact and the absence of data, ensuring continuity across the time series and preventing misleading discontinuities in the visualization. By doing so, the resulting proportions faithfully represent the true contribution of unlawful interference events to annual fatality counts, while maintaining methodological consistency with the frequency-based analysis presented earlier.



Unlike the relatively stable and low proportions observed in the frequency-based analysis, the fatality-based perspective reveals a markedly different pattern. For most years, unlawful interference events account for a negligible share of total fatalities, often close to zero. This reinforces the notion that, under normal circumstances, such events do not significantly shape overall aviation fatality statistics.

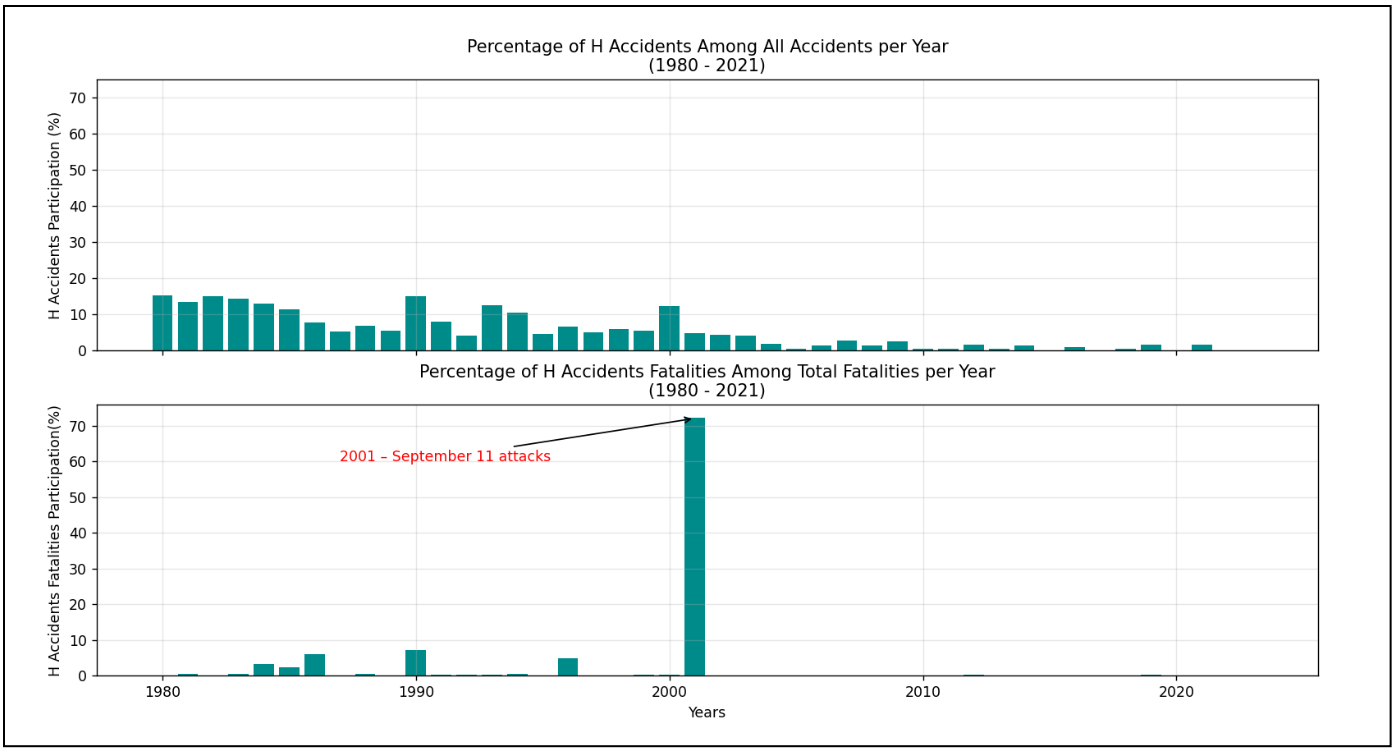
However, this pattern is abruptly disrupted in 2001. In that year, fatalities associated with unlawful interference represent an overwhelming proportion of total aviation deaths, dwarfing all other years in the dataset. This extreme concentration clearly identifies 2001 as a statistical outlier, both in magnitude and in nature.

It is important to emphasize that this spike does not reflect a generalized increase in the lethality of aviation accidents, nor does it indicate a systemic failure in aviation safety. Rather, it results from a small number of extraordinary events whose consequences extended well beyond the aircraft itself, including a large number of ground casualties. As such, the fatality figures for 2001 cannot be interpreted solely as an indicator of aviation operational risk.

The absence of comparable spikes in subsequent years further supports this interpretation. Following 2001, the proportional contribution of unlawful interference events to annual fatalities returns to consistently low levels, suggesting that the extreme impact observed was neither recurrent nor indicative of a long-term trend.

In summary, the fatality analysis confirms a critical asymmetry: unlawful interference events are statistically rare, yet capable of producing disproportionate human losses under exceptional circumstances. This distinction underscores the importance of separating event frequency from event impact when interpreting aviation accident data and provides a coherent transition toward broader discussions on safety, risk perception, and systemic resilience in the following sections.

3.2.4 Final Considerations



By juxtaposing the frequency-based and fatality-based analyses, a clear structural distinction emerges between the occurrence and the consequences of unlawful interference events in aviation history. When observed simultaneously, the two dimensions highlight that the societal impact of such events is not driven by their recurrence, but by their capacity to concentrate extreme outcomes into isolated points in time. This contrast would be far less evident if either metric were analyzed independently.

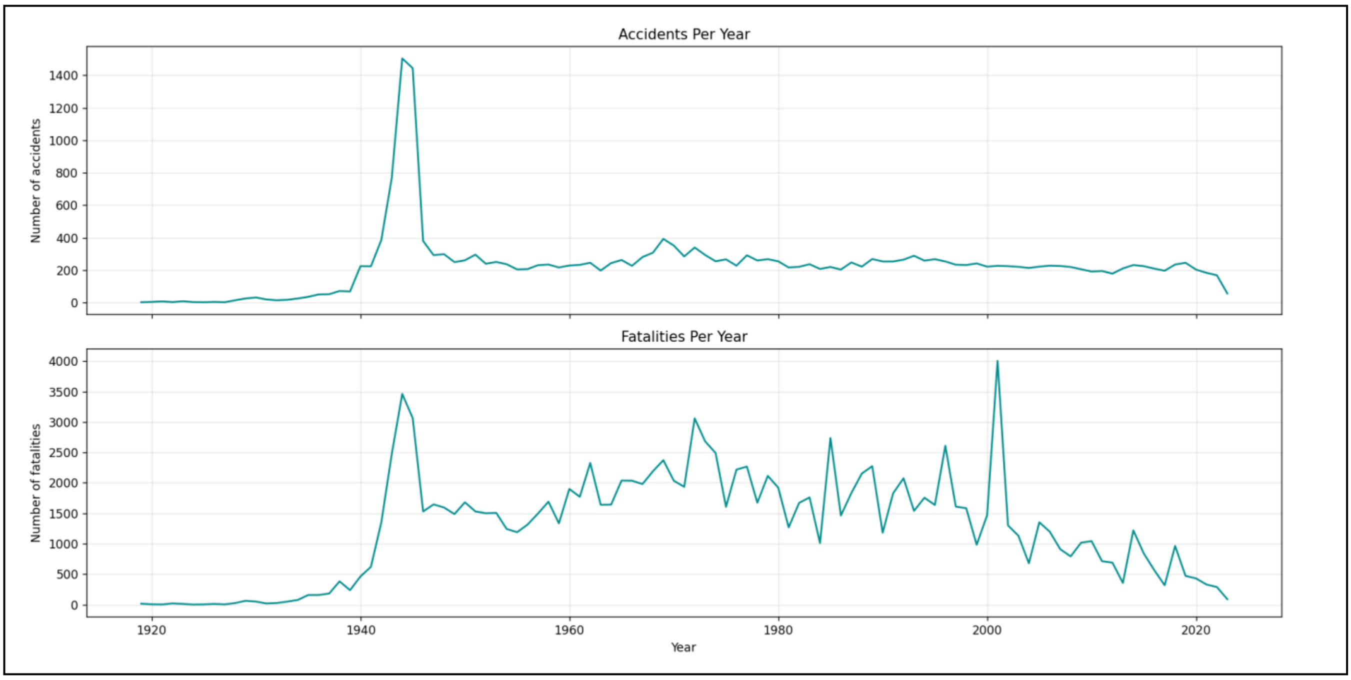
The empirical evidence further suggests that aviation-related unlawful interference should be interpreted as an exogenous risk rather than an intrinsic operational vulnerability of the aviation system. Unlike technical failures or human-error-related accidents, these events do not follow gradual trends, nor do they scale with traffic volume or technological complexity. Their effects are abrupt, discontinuous, and heavily context-dependent, reinforcing the importance of treating them separately from conventional safety indicators.

Finally, the absence of recurring high-impact events in the post-2001 period provides indirect evidence of systemic resilience. While the dataset does not encode security measures explicitly, the sustained decline in both frequency and impact of unlawful interference events suggests that structural changes introduced after major historical shocks were effective in limiting recurrence. This observation provides a natural transition toward a broader discussion on long-term aviation safety trends and the mechanisms through which risk is mitigated over time.

4. Aeronautical Safety: Temporal and Structural Evolution

4.1. Long-Term Evolution of Accident and Fatality Counts

4.1.1 Global Historical Evolution (1919–2023)

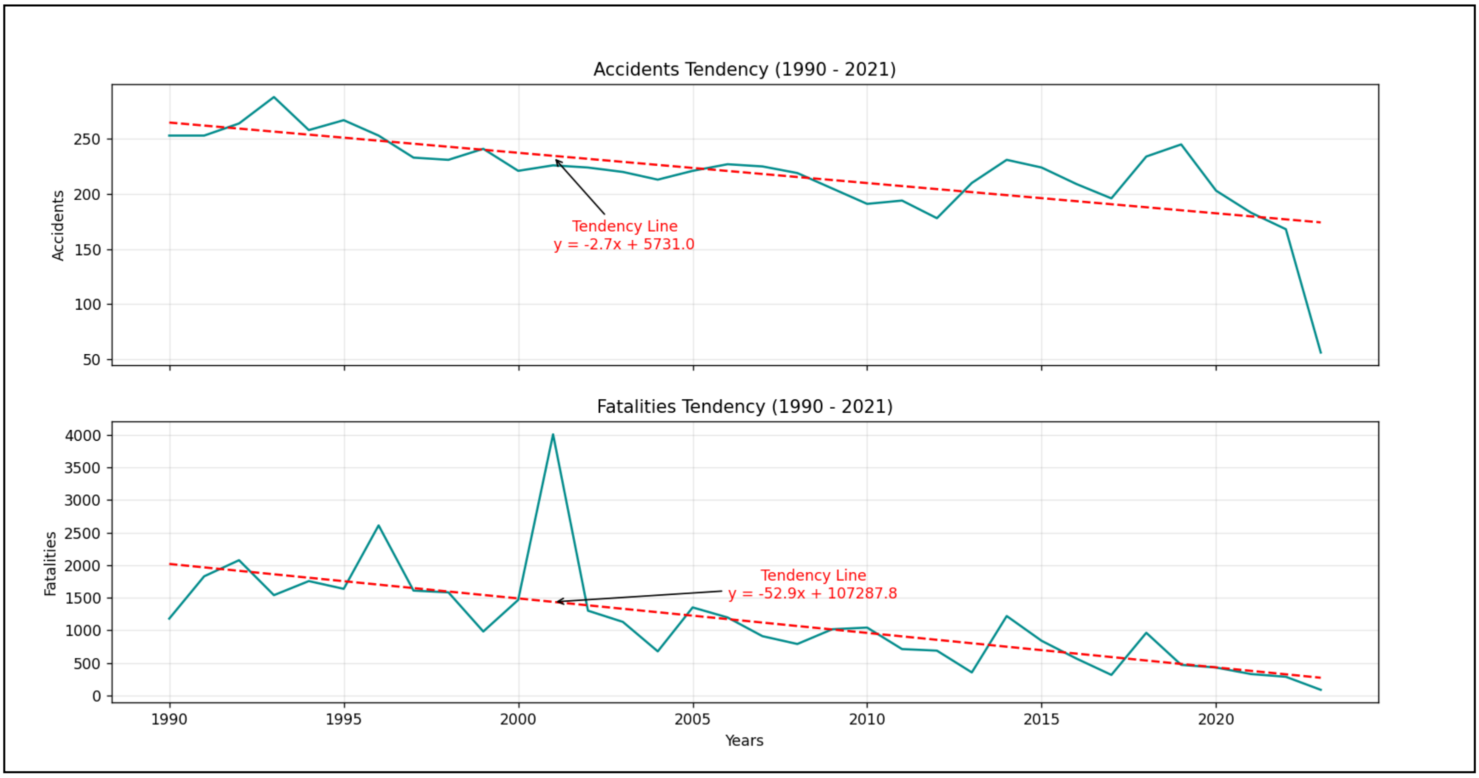


The time series of annual accident and fatality counts, previously introduced in the broader historical analysis, provides a consolidated overview of aviation safety evolution over more than a century. A gradual increase is observed during the early decades of aviation, followed by exceptionally pronounced peaks during the Second World War, reflecting the massive operational use of aircraft and the consequent increase in exposure to risk. After this period, the data enters a phase of relative stabilization, with fluctuations mainly associated with isolated high-impact events and transitional phases in global aviation, such as the expansion of commercial air transport and the terrorist attacks of 2001.

While this global perspective is essential for contextual understanding, it also highlights an important analytical limitation: interpretations based solely on absolute counts can be misleading. The continuous growth of worldwide air traffic implies that raw totals do not directly translate into changes in safety levels. This observation motivates a more focused analysis on modern periods and the use of quantitative tools capable of capturing structural, long-term safety trends.

4.1.2 Modern Aviation Era and Long-Term Safety Trends (1990–2021)

Restricting the analysis to the post-1990 period, representative of modern commercial aviation, allows safety trends to be examined under relatively stable technological, regulatory, and operational conditions. Linear trend lines fitted to both accident and fatality time series reveal a consistent pattern: despite year-to-year variability and the presence of notable outliers, both metrics exhibit a clear downward trend over time.



From a technical standpoint, the decreasing trend in accident counts suggests a progressive reduction in the frequency of adverse events, while the more pronounced decline in fatalities points to substantial improvements in risk mitigation. These improvements likely stem from advances in aircraft design, enhanced safety regulations, improved crew training, and more effective emergency response procedures. The simultaneous decline of both indicators strengthens the interpretation that observed improvements are structural rather than incidental.

Importantly, these negative trends emerge even without explicit normalization by air traffic volume, which increased significantly over the same period. Consequently, the reduction in absolute accident and fatality counts constitutes strong evidence of genuine improvements in aviation safety, providing a solid foundation for subsequent analyses focused on accident severity and changes in event characteristics over time.

4.2 Evolution of Accident Severity

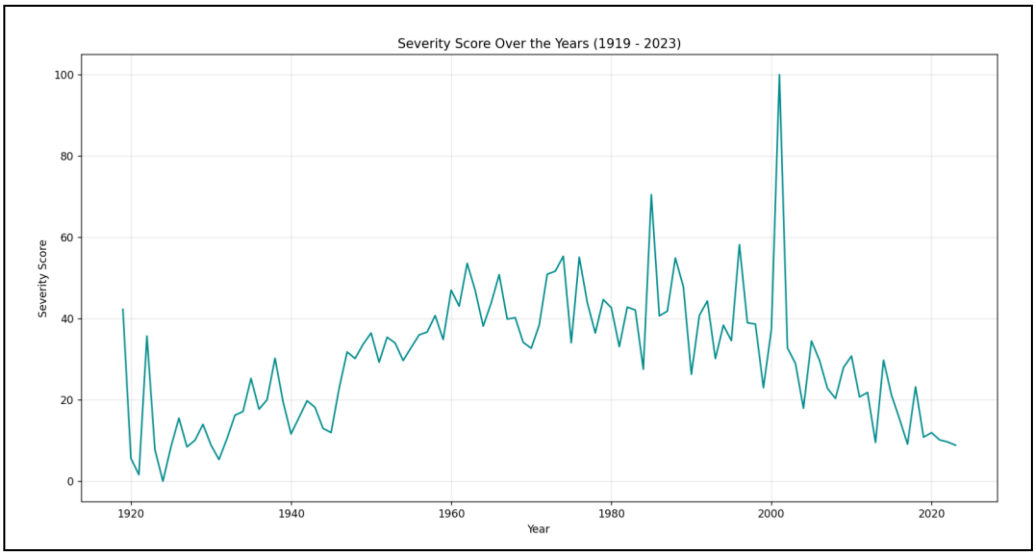
As discussed earlier, the accident category variable provided in the dataset does not reliably represent accident severity, as it primarily encodes the nature of the event rather than its human impact. For this reason, severity is analyzed in this section using a quantitative proxy derived directly from fatality data.

The severity metric adopted here is defined as the average number of fatalities per accident for each year, computed as the ratio between total fatalities and total accidents in that year. To facilitate interpretation and enable consistent comparison across the full historical timeline, this ratio is subsequently normalized by the maximum observed value in the dataset. This normalization scales the severity score between 0 and 100, where the peak value corresponds to the most severe year in relative terms.

Notably, the maximum severity score occurs in 2001, reflecting an exceptional and well-documented outlier rather than a representative operational safety level. Normalizing by this peak allows all other years to be interpreted in relative terms, highlighting long-term structural trends while preserving the visibility of extreme events.

This normalized severity score therefore serves as a comparative indicator of fatality density per accident, rather than an absolute measure of risk, and is particularly well-suited for studying long-term safety evolution.

4.2.1 Annual Fatalities per Accident: Global Severity Patterns



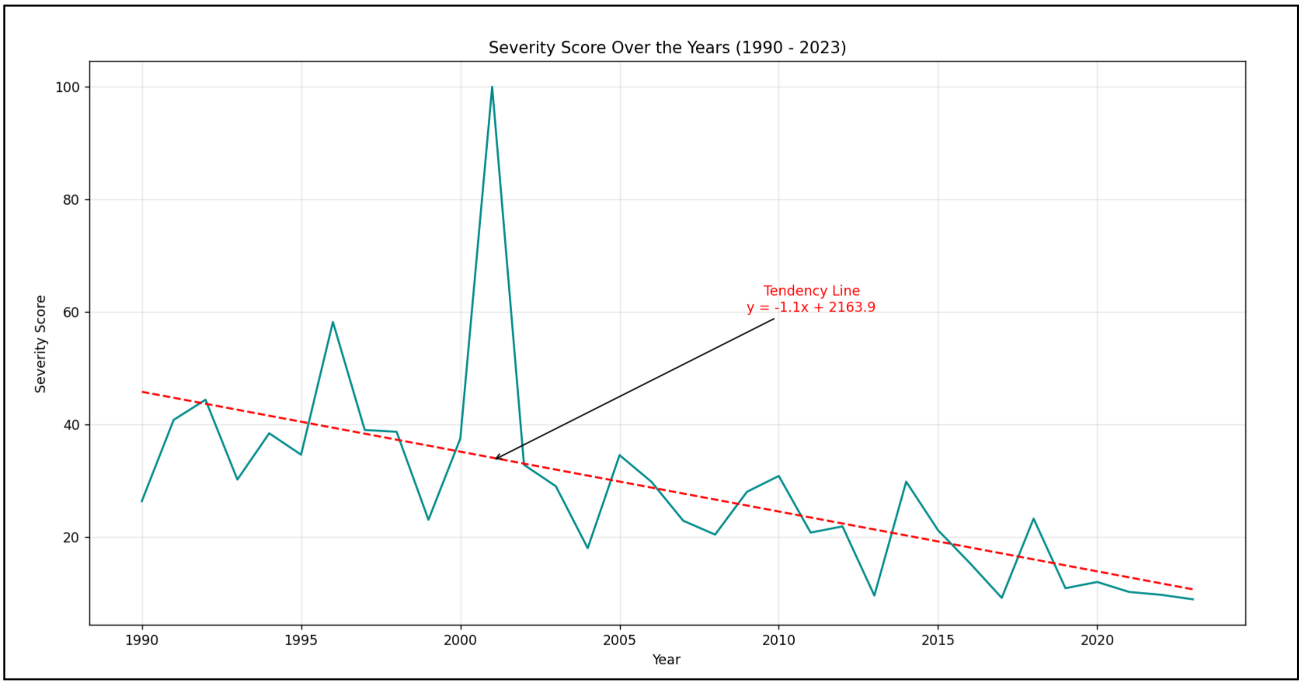
The long-term severity score reveals a clear historical structure. During the early decades of aviation, severity values are highly volatile, reflecting both limited safety standards and the relatively small number of accidents per year, where single high-fatality events strongly influenced annual averages. A pronounced increase in severity is observed during the 1940s, consistent with the Second World War period previously analyzed, when military aviation dominated operations and accidents frequently involved fully loaded aircraft under hostile or high-risk conditions.

In the post-war decades, severity remains elevated but progressively stabilizes. This period corresponds to the rapid expansion of commercial aviation, with increasing aircraft capacity and traffic volume. While total fatalities remain substantial during these decades, the severity score suggests that accidents, on average, became less catastrophic relative to the number of occurrences. Isolated spikes remain visible and are typically associated with singular, high-impact events rather than systemic safety degradation.

Toward the late 20th and early 21st centuries, a structural downward shift becomes evident. Although individual outliers persist, the overall severity score declines markedly, indicating a reduction in fatalities per accident despite the growing scale and capacity of global air transport.

4.2.2 Long-Term Evolution of Severity in the Modern Era (1990–2021)

Focusing on the modern aviation era provides a clearer view of underlying safety dynamics.



From 1990 onward, the severity score exhibits a consistent downward trend, which is reinforced by the fitted linear trend line. This negative slope indicates that, on average, aviation accidents have become progressively less lethal over time.

Importantly, this decline persists even when accounting for extreme outliers, such as the sharp spike observed in 2001. As discussed earlier, this anomaly is largely driven by exceptional circumstances rather than operational aviation risk alone. Once such events are contextualized, the broader trend remains robust: modern aviation accidents tend to involve fewer fatalities per occurrence.

This evolution reflects cumulative improvements in aircraft design, safety systems, emergency response protocols, regulatory oversight, and operational training. Larger aircraft capacities in recent decades have not translated into proportionally higher fatality counts per accident, reinforcing the interpretation that safety gains have outpaced increases in exposure.

Taken together, the severity analysis complements earlier findings based on absolute accident and fatality counts. While raw totals alone can obscure progress due to traffic growth, the severity score provides strong quantitative evidence of a structural improvement in aviation safety, particularly in the modern era.

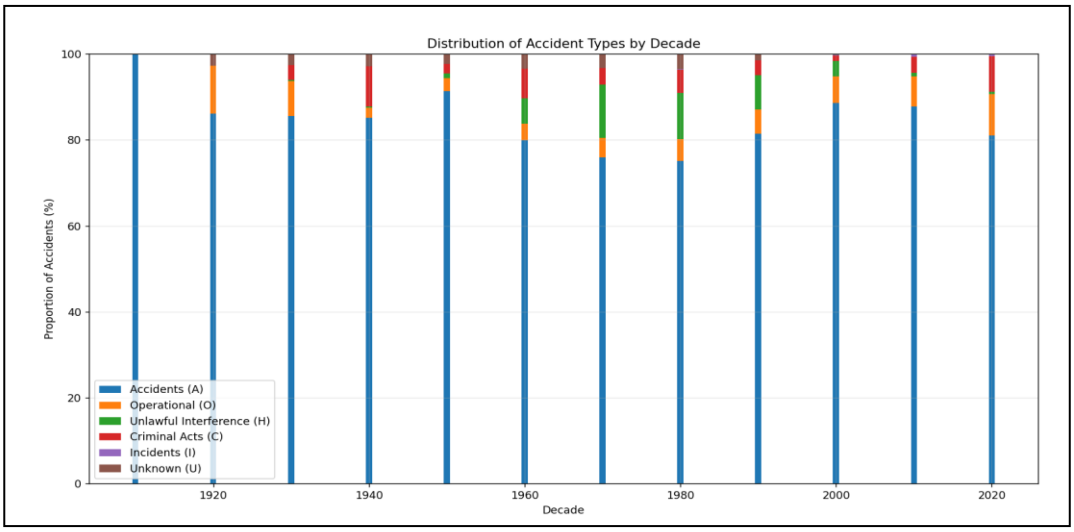
4.3 Changes in Accident Type Distribution Over Time

4.3.1 Reclassification of Accident Categories for Event-Based Analysis

To enable a clearer and more interpretable long-term analysis of accident types, the original categorical structure of the dataset was simplified by aggregating closely related subcategories. Given that several categories in the original database primarily differ by severity levels rather than by the underlying nature of the event, subtypes such as A1/A2, O1/O2, H1/H2, C1/C2, and I1/I2 were merged into unified categories (A, O, H, C, and I, respectively). This consolidation was motivated by the analytical focus on event typology rather than on the database's internal severity labeling, which has already been shown to be inconsistent for comparative severity analysis.

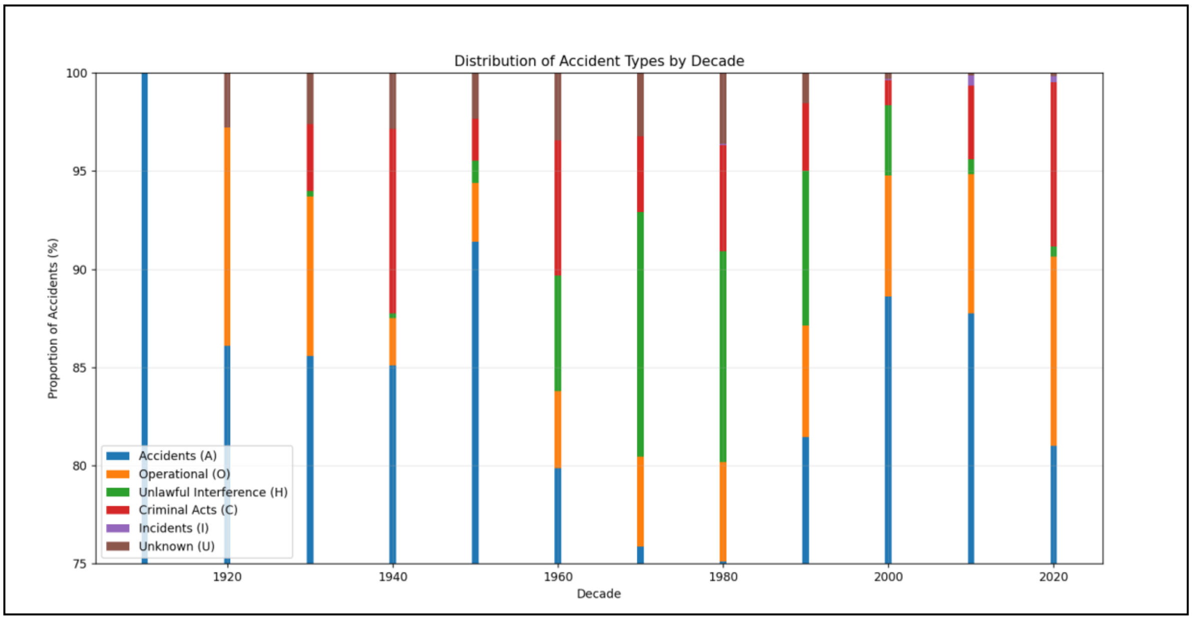
This reclassification preserves the semantic meaning of each event type while significantly reducing categorical fragmentation, making temporal comparisons more robust and avoiding misleading conclusions driven by artificial distinctions. The resulting categories represent broad classes of events: accidents related to operational or technical failures (A), operational or procedural issues without clear accident causality (O), unlawful interference events (H), criminal acts (C), incidents of lower operational impact (I), and a residual category for insufficiently documented cases (U).

4.3.2 Decadal Evolution of Accident Type Proportion



The distribution of accident types by decade is analyzed using stacked bar charts, where each bar represents 100% of recorded events within a given decade. This decade-based aggregation serves two purposes: it smooths year-to-year volatility and allows structural changes in accident typology to be clearly identified over long time horizons.

A first and immediately visible result is the overwhelming dominance of category A (Accidents) across all decades. In every period, this category constitutes the majority of recorded events, confirming that most occurrences in the database correspond to conventional aviation accidents rather than exceptional or external interference events. Because this dominance visually compresses the relative contributions of other categories, a secondary, zoomed-in visualization was produced to better inspect the evolution of non-A categories.



From a historical perspective, the earliest decades, particularly the first decade shown, must be interpreted with caution. The unusually high concentration of a single category and the near absence of others reflect data sparsity and limited documentation rather than a reliable structural pattern. This period constitutes a clear outlier driven by low sample size and incomplete reporting practices, and it should not be overinterpreted in comparative analyses.

As documentation quality improves over time, the Unknown (U) category rapidly diminishes, becoming nearly negligible in the last two to three decades. This trend strongly suggests that the reduction in unknown classifications is not a safety phenomenon but a documentation effect, reflecting advances in investigation standards, reporting protocols, and international data sharing.

More substantively, the relative proportion of Accidents (A) shows a gradual declining trend in recent decades, while categories such as Operational (O) and Criminal (C) display a modest but consistent increase in proportional representation. This shift does not necessarily indicate a rise in absolute occurrences of these events but rather a structural rebalancing of recorded accident causes. As conventional accident rates decrease, non-accident-related events naturally account for a larger share of the remaining occurrences.

The Unlawful Interference (H) category remains relatively contained across decades, with no sustained long-term dominance, reinforcing earlier findings that such events, while impactful, do not define the overall accident landscape in aviation history. Similarly, Incidents (I) maintain a low and stable proportion, consistent with their definition as lower-severity operational events.

Overall, this decade-based distribution analysis reveals a clear structural evolution: improved documentation, declining dominance of traditional accident mechanisms, and a relative increase in operationally and externally driven event categories. These patterns provide strong contextual support for the broader conclusion that aviation safety improvements must be interpreted beyond absolute accident counts, accounting instead for changing operational complexity, reporting standards, and the evolving nature of aviation risks.

4.4 Interpreting Safety Improvements Beyond Absolute Counts

While absolute counts of accidents and fatalities provide an initial overview of aviation safety, they are insufficient on their own to support robust conclusions about long-term risk. Aviation activity has expanded dramatically over the last century, with substantial growth in the number of aircraft, flights, passengers, and operational complexity. In this context, comparing raw accident or fatality totals across distant periods can be misleading, as similar absolute numbers may correspond to fundamentally different risk levels.

The analyses conducted in the previous sections demonstrate that safety improvements become more evident once structural and relative indicators are considered. The long-term decline in accident and fatality trends observed in the modern era gains additional significance when interpreted alongside the reduction in severity metrics. Even in years where accidents still occur, the average lethality per event has decreased, suggesting improvements not only in accident prevention but also in aircraft survivability, emergency response, and operational resilience.

Furthermore, the historical shift in accident type distribution reinforces this interpretation. Traditional technical accident categories, which dominated earlier decades, have gradually declined in relative importance. In contrast, a larger share of recent events falls under operational, organizational, or externally driven categories. This transition indicates that many classical engineering risks have been mitigated to a significant extent, while residual risks increasingly arise from human, procedural, or systemic factors. Such a pattern is consistent with a mature safety system in which failures become rarer and more complex rather than frequent and purely mechanical.

Taken together, these results suggest that modern aviation safety progress should be evaluated through a multidimensional lens. Absolute accident counts alone obscure meaningful improvements that emerge when severity, proportional distributions, and long-term structural changes are considered. The evidence points toward a sustained reduction in both the frequency and impact of aviation accidents relative to operational scale, supporting the conclusion that aviation safety has improved substantially over time, even as the system itself has grown larger and more complex.

5. Conclusion

This study set out to examine the historical evolution of aviation accidents and fatalities through a comprehensive, data-driven analysis spanning more than a century. By systematically cleaning, validating, and contextualizing the dataset prior to interpretation, the analysis established a reliable foundation for exploring long-term safety dynamics while explicitly acknowledging the limitations inherent to historical aviation records.

The results demonstrate that aviation safety cannot be adequately assessed through absolute accident or fatality counts alone. While raw numbers reveal important historical disruptions, most notably World War II and the events of 2001, they also obscure deeper structural trends. When the analysis is extended to relative measures, such as severity indicators and proportional distributions, a consistent pattern of long-term improvement emerges, particularly in the modern aviation era.

The investigation of historical periods highlights a crucial distinction between exposure-driven risk and systemic safety performance. World War II represents a period of sustained, high-exposure aviation activity under wartime conditions, explaining the extreme accident and fatality levels observed. In contrast, the 2001 spike in fatalities is shown to be an isolated outlier driven by unlawful interference, rather than a deterioration of aviation safety mechanisms. The absence of similar patterns in subsequent years suggests that such events, while catastrophic, do not define long-term aviation risk.

From a structural perspective, the analysis of accident severity provides strong evidence that modern aviation accidents are, on average, less lethal than those of earlier decades. This decline persists even when accounting for exceptional outliers and occurs alongside substantial growth in aircraft capacity and traffic volume. Such findings point to cumulative safety gains driven by advances in aircraft design, regulatory oversight, operational procedures, and emergency response capabilities.

Additionally, the evolution of accident type distribution reveals a shift away from traditional technical accident dominance toward a more complex mix of operational, organizational, and externally driven events. This transition reflects both improved safety in core engineering domains and enhanced investigative and reporting standards. The near disappearance of poorly documented events in recent decades further supports the conclusion that modern aviation benefits not only from safer operations but also from greater transparency and data quality.

Taken together, the evidence supports a clear conclusion: aviation safety has improved substantially over time, not merely in terms of reduced accident frequency, but through a broad structural transformation that reduced severity, mitigated traditional risks, and increased systemic resilience. These improvements become visible only when safety is evaluated through a multidimensional lens that goes beyond absolute counts.

In this sense, the study reinforces a broader methodological insight: meaningful safety analysis in complex systems requires historical context, careful data interpretation, and an explicit separation between rare, high-impact anomalies and long-term structural trends. When approached in this way, the historical record of aviation accidents provides compelling evidence of sustained and significant safety progress.

6. References

[1] Aviation Accident Dataset (1919–2023). Available at: https://www.kaggle.com/code/thabresh/insights-aviation-accidents-1919-2023-eda

[2] GitHub Project (codes): https://github.com/rodrigodriemeier/A-Data-Driven-Study-of-Aviation-Accidents-Fatalities-and-Safety-Trends

[3] Aviation Accident March 2021 – Mauritania Airlines - Mauritania. Available at: https://aviation-safety.net/wikibase/319021

[4] Aviation Accident July 2021 – Ryan Air - USA. Available at: https://aviation-safety.net/wikibase/265316