Evaluating the Influence of the International Civil Aviation Organization on Aircraft Accident Rates and Fatalities: A Seven-Decade Historical Data Analysis

Rossi Passarella[1]* , Harumi Veny[2] , Muhammad Fachrurrozi[3] , Samsuryadi Samsuryadi[3] , Marsella Vindriani[1]

[1] Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya, 30139 Palembang, Indonesia

[2] College of Engineering, School of Chemical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

[3] Department of Informatics, Faculty of Computer Science, Universitas Sriwijaya, 30139 Palembang, Indonesia

* Correspondence: Rossi Passarella (passarella.rossi@unsri.ac.id)

Abstract: The advent of air travel, originally proposed by the Wright brothers, has led to a significant surge in aircraft usage for human transportation. In its nascent stages, this mode of transport was linked with a high frequency of accidents and consequent fatalities, placing it in the high-risk category. To counter these risks, the International Civil Aviation Organization (ICAO) was established in 1947 as a collaborative effort among numerous countries with the primary goal of enhancing aviation safety regulations. This study analyzed archival data from the Bureau of Aircraft Accidents Archives (B3A), covering a span of 72 years from 1918, the year of the first commercial airplane crash, until 2020. The objective was to understand the ICAO's impact on altering accident rates, fatalities, and underlying causes. Analytical methodologies encompassed both descriptive statistics—examining data distribution, central tendencies, and category frequencies—and exploratory data analysis (EDA) to identify variable relationships and outlier identification. The results indicated that ICAO's interventions have led to a notable decline in accident rates, with an annual average reduction of 70.9%, and a corresponding decrease in incidents attributed to technical factors. However, an unexpected trend was the increase in fatalities despite the drop in accident numbers, attributable to the introduction of larger aircraft designs carrying more passengers per flight. The findings underscore the ICAO's successful efforts in reducing aircraft accidents, but also suggest a need for further exploration into factors contributing to the rise in fatalities.

Keywords: Commercial aircraft crash data; Data analysis; Descriptive statistics; Exploratory data analysis; ICAO impact

1 Introduction

The aviation industry, particularly in the realm of commercial operations, has been witnessing a consistent upsurge in air traffic volume and frequency over recent decades. This underscores the transformation of airplanes into a predominant means of transportation. As per data released by the International Civil Aviation Organization (ICAO), approximately 4,397 billion passengers were reported to have chosen airplanes as their primary mode of transportation in 2019 [1]. Moreover, projections suggest a yearly growth rate of about 1.8% in air travel demand over the forthcoming two decades [2]. This sustained interest in air travel can be attributed to its superior safety record compared to other forms of transportation [3–5]. Evidence supporting this comes from the Bureau of Transportation Statistics, which documented the lowest accident percentage for air transportation across the 2004–2019 timeframe [6]. This impressive safety record is largely credited to advancements in aircraft design, technology, and safety standards [7] along with the oversight of safety organizations like the ICAO.

Despite aviation's exemplary safety record, it remains crucial to analyze past aircraft accidents to prevent recurrence and enhance overall aviation safety. In order to preclude accident recurrence, the establishment of an

international organization is advocated. This research is primarily focused on assessing the impact of the ICAO, an international organization, in preventing recurring aviation accidents and augmenting aviation safety.

Further improvements in aviation safety and accident prevention may be achieved through an examination of historical accident data. This allows for a comprehensive understanding of statistical data characteristics and the extraction of concealed knowledge [8, 9]. Most existing commercial aircraft accident analyses are geographically limited [10], making them less effective given the transnational and intercontinental nature of commercial flight activities. Previous methods also heavily relied on expert experience, which can sometimes lead to user errors during data analysis [11]. Moreover, the absence of formal accident investigation processes and the disregard for loss of life have resulted in substandard accident analyses [12]. Newer methods, such as machine learning, are constrained by the hidden knowledge and insights in the data [11]. To address these issues, it is proposed to adopt analytical methodologies such as descriptive statistics [13] and exploratory data analysis (EDA) [14].

Considering that the aviation accident data to be analyzed is in tabular form, employing EDA seems most fitting as it scrutinizes each data observation per variable (column). The goal of this study is to unravel variable relationships and analyze shifts in fatalities and commercial aircraft accidents pre- and post-ICAO establishment, thereby gauging its impact on aviation safety. This will help ascertain if the ICAO, as an international civil aviation safety organization, is achieving its main objectives.

This study has been structured as follows: The first section provides the introduction and outlines the study objectives; the second section discusses the methodology, detailing the strategies adopted to fulfill the study objectives, in addition to elaborating on data collection, processing, and EDA implementation procedures; the third section presents the results derived from the methodology's application and the subsequent analysis process; the final section furnishes the conclusions drawn from the study.

2 Methodology

The Bureau of Aircraft Accidents Archive (B3A) database served as the data source for this investigation. A meticulous analysis was performed on the B3A database to ascertain the ICAO's influence on enhancing transportation safety. A flowchart representing the stages of study is depicted in Figure 1.
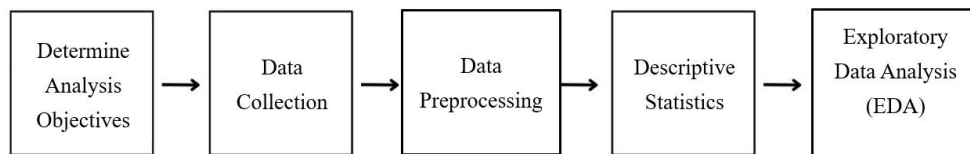
Figure 1. Stages of study

The initial phase involved delineating the analysis' purpose. By defining the objective of the data analysis, appropriate variables can be selected to facilitate conclusion drawing and hypothesis or research question testing [15]. The intention of this study was to scrutinize the ICAO's impact on aviation safety, accomplished by examining fluctuations in accident count, fatalities, and material losses over time.

The subsequent stage encompassed data collection. This phase involves gathering and measuring variable-related information to analyze data procured from diverse sources [16]. Next came data preprocessing, a phase intended to rectify inconsistencies and anomalies present in raw data, such as outliers, null values, etc. [17]. Given its significance in problem rectification, data preprocessing is considered an essential precursor to the data analysis process [18]. In this investigation, three data preprocessing techniques were employed: data cleaning, data transformation, and data reduction [19].

The data cleaning process was implemented via the Python programming language's Pandas library to identify missing values. Given that accident data represents factual information and cannot be estimated, any observation row data exhibiting a null value was eliminated. The data transformation phase involved converting raw data from character (object) to numeric format, significantly simplifying the EDA analysis process. Data reduction was the final preprocessing step, where insignificant raw data variables (from 15 variables) were discarded to streamline the analysis process. This process was conducted using a statistical approach, specifically examining the correlation value between variables.

Post-data preprocessing, the analysis was carried out employing two distinct methods: Descriptive Statistics and Exploratory Data Analysis (EDA). Descriptive Statistics involves simplifying and summarizing data with statistical calculations, presenting it in an attractive, helpful, and easily comprehensible manner [20]. Data description techniques vary depending on the data type. Quantitative data types in variables utilize measures of frequency [21], central tendency [22], dispersion [23], and kurtosis and skewness [24]. Qualitative data types in variables, on the other hand, employ the technique of calculating the frequency of occurrence and the proportion or percentage of each value category [22].

EDA aims to comprehend data structure and patterns, serving multiple analytical purposes, including identifying outliers, detecting data deviations, determining crucial variables, suggesting hypotheses, and uncovering hidden patterns [25]. Statistical analyses or graphic representations can be deployed to address hypotheses built with EDA [26]. Specific techniques, such as measures of frequency, central tendency, and dispersion, were applied to the dataset for quantitative data, with the calculation of frequency and proportion applied for qualitative data. Graphical representations were also generated. All these stages were conducted using the Python programming language [27].

3 Results and Discussion

3.1 Determine Analysis Objectives

The analysis aimed to address the research question by characterizing commercial aircraft accident data through statistical calculations and Exploratory Data Analysis (EDA) to assess changes in commercial aircraft accidents and fatalities pre and post the International Civil Aviation Organization (ICAO) establishment. The initial hypothesis posited a decline in the number of accidents and fatalities involving commercial aircraft subsequent to the ICAO's establishment, due to the standardization of international civil aviation rules and conditions by the ICAO [28].

3.2 Data Collection

Data, observed from 1918 to 2020, were collected in December 2020 from the website https://www.baaa-acro.com/ and preserved in CSV format.

3.3 Data Preprocessing

The aircraft accident data underwent several preprocessing steps to prepare it for subsequent analysis. The data was loaded into Jupyter notebook software and stored in a data frame consisting of 592 rows and 15 columns. The variable columns were renamed to facilitate code readability, with each variable character changed to lowercase and the space between words replaced with an underscore. Missing values were checked and found to be absent in the data frame, obviating the need for deletion. Non-commercial flight accident data was removed by eliminating rows with flight types other than scheduled revenue flights, air/taxi, cargo, and postal. This step reduced the number of rows to 110. Rows with unknown causes of accidents were also removed, resulting in 91 rows of data. Finally, variables focusing on answering research questions were selected, namely the year, number of deaths, and possible causes. This process reduced the number of columns to three.

As a result of the preprocessing stage, the data frame was reduced to 91 commercial aircraft crash data lines, which we refer to as the dataset. The next step is to implement descriptive statistical and EDA methods.

The dataset underwent several preprocessing steps to prepare it for analysis. The data was loaded into Jupyter notebook software and stored in a data frame consisting of 592 rows and 15 columns. Variable columns were renamed to enhance code readability, with each variable character converted to lowercase and the space between words replaced with an underscore. Missing values were checked and found to be absent in the data frame, rendering deletion unnecessary. Non-commercial flight accident data was eliminated by removing rows with flight types other than scheduled revenue flights, air/taxi, cargo, and postal, resulting in a reduction of rows to 110. Rows with unknown causes of accidents were also removed, leading to 91 rows of data. Finally, we selected variables that focused on answering research questions, namely the year, number of deaths, and possible causes, reducing the number of columns to three.

Following the preprocessing stage, the dataset comprised 91 commercial aircraft crash data lines, which we refer to as the dataset. Descriptive statistical and EDA methods were then implemented.

3.3.1 Variables with a quantitative data type

In this segment, the exploration of two quantitative variables within the dataset - 'year' and 'fatalities' - was performed using a descriptive statistical analysis approach, encompassing frequency measurement, measures of central tendency, dispersion, and assessments of skewness and kurtosis [29].

The 'year' variable, indicating the year of the accident, underwent frequency measurement to ascertain the prevalence of particular years in the dataset. The analysis revealed fifty distinct accident years within the dataset, spanning from 1918 to 2020. The year marked by the highest frequency of incidents was 1943, with five recorded accidents. This was followed by years 1944, 1936, 1938, 1925, and 1948, each accounting for four accidents. Further detail regarding the frequency measurement of the 'year' variable is depicted in Figure 2.

Moreover, the 'total fatalities' variable was assessed for the frequency of fatalities per accident, per year, as illustrated in Figure 3. The data revealed 34 unique categories of fatalities.

Upon completion of frequency measurements for both variables, the central location of data distribution was determined. A few parameters, such as the mean, median, and mode, were considered for this task [22]. The choice of appropriate parameters required understanding of the data distribution type for each variable, with three possibilities: symmetrical, positively skewed, and negatively skewed [30].

Symmetrical distribution can be identified when mean, median, and mode values are equal. Positively skewed distribution exhibits a median value less than the mean value, an asymmetric distribution, and a tail curving towards the right. Negatively skewed distribution, conversely, is characterized by a median value exceeding the mean value, an asymmetric distribution, and the tail of the curve leaning leftwards.
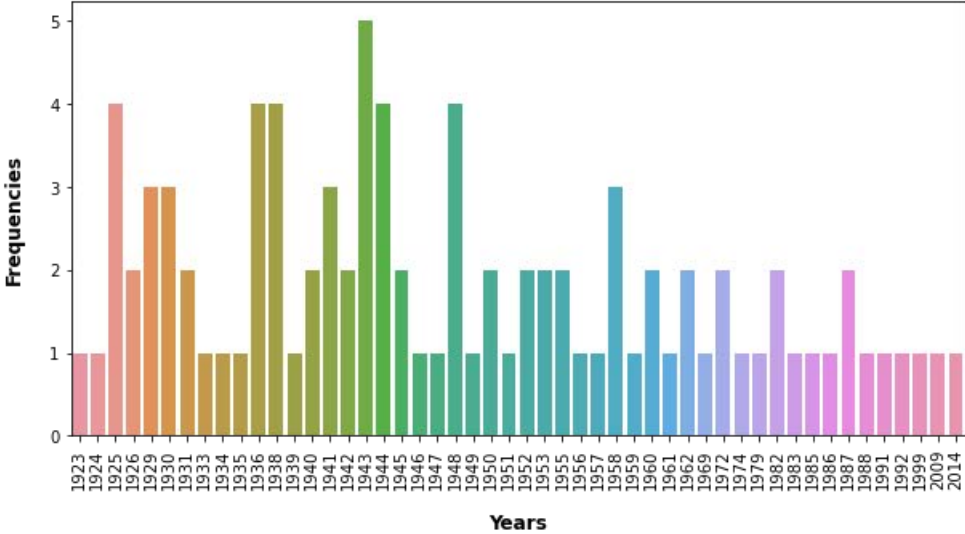

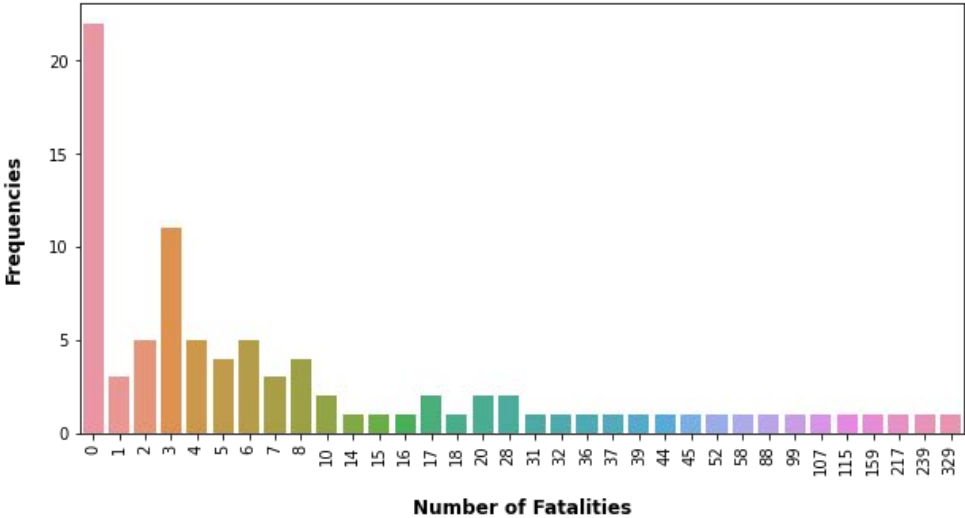**Figure 2.** Measures of frequency on variable years

**Figure 3.** Measures of frequency on variable total fatalities

This analysis was performed using the Python programming language, more specifically, the Pandas library with the "*df.describe()*" command. The "df" stands for the data frame, which represents the dataset, and the *describe()* method is employed to summarize central tendency, dispersion, and the shape of the data distribution.

Variables with symmetrical data distribution could utilize either mean, median, or mode parameters. However, variables exhibiting skewed data distribution are best represented by the median parameter, given its resistance to outliers' influence [30]. The findings from these calculations are presented in Table 1.

Figure 4 displays the data distribution for the variables 'year' and 'fatalities' with their respective mean, median, and mode values. In Figure 4(A), the 'year' variable exhibits a positively skewed data distribution, thus the median, which is 1945, serves as the central location [31]. The graph for the 'fatalities' variable (Figure 4(B)) also demonstrates a positively skewed data distribution. The mean value is positioned to the right of the median value, while the mode value lies to the left of the median. Hence, the median, with a value of 4, signifies the central location of the fatalities distribution.

Table 1. The results of central tendency measurement parameters calculation on years and fatalities variables

Parameters	Years	Fatalities
Mean	1951.09	23.17
Median	1945	4.0
Mode	1943	0
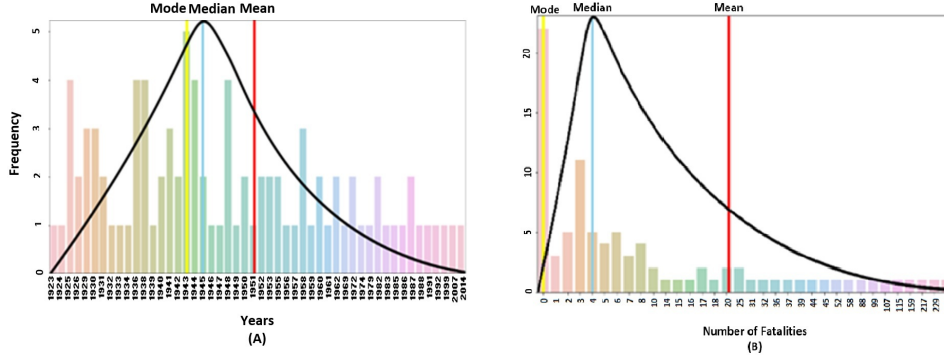
Figure 4. (A) Measures of central tendency visualization on years variable; (B) Measures of central tendency visualization on total fatalities variable

Upon establishing the central location, the dispersion of data around this central point was assessed [32]. Low dispersion implies a concentration of data points around the central location, denoting consistency, while high dispersion indicates inconsistency and a broad spread of data points from the central location [23]. Several parameters, such as range, interquartile range (IQR), sample variance, and standard deviation, were considered for dispersion measurement. Similarly to previous measurements, the choice of parameter hinged upon the type of data distribution for each variable. For variables exhibiting symmetrical distribution, a combination of mean, variance, and standard deviation sufficed. However, for variables with skewed distribution, the IQR was chosen due to its resistance to outliers [23]. The results of dispersion measurement for 'year' and 'fatalities' variables are presented in Table 2. As previously determined, the 'year' variable showcases a positively skewed distribution, indicating the presence of outliers [31]. Therefore, the IQR emerges as the appropriate measurement for data dispersion as it is less impacted by outliers. Figure 5(A) visualizes the data dispersion for the 'year' variable using IQR. The IQR, representing the range between the first (Q1) and third quartiles (Q3), yielded a value of 22.5, with Q1 and Q3 values of 1937 and 1959.5 respectively. Certain years such as 1999, 2009, and 2014 were identified as outliers as they fall outside the minimum and maximum distribution limits. In contrast, the 'fatalities' variable exhibits an IQR of 16, within a range of 329, as demonstrated in Figure 5(B). However, a substantial number of outliers, ranging from 42 to 329, are observed, indicating a considerable spread from the central location.

Table 2. The results of dispersion measurement parameters calculation on years and fatalities variables

Dispersion Parameters	Years	Fatalities
Range	91	329
Interquartile Range (IQR)	22.5	16
Variance	423.60	2772.58
Standard Deviation	20.581	52.65

The final step in the quantitative analysis involved the evaluation of skewness and kurtosis. Though skewness could be inferred from Figure 4, numerical values for skewness and kurtosis were obtained to verify the findings from central tendency calculations. The skewness value indicates how the data distribution deviates from a normal distribution. If skewness equals zero, the data distribution is symmetrical. However, if skewness is either less than -1 or greater than 1, the distribution is highly skewed. If skewness ranges between -1 and -0.5 or 0.5 and 1, the distribution is moderately skewed. Lastly, if skewness falls between -0.5 and 0.5, the data distribution is approximately symmetrical [33]. A positive skewness value indicates a rightward tilt of the data distribution curve, whereas a negative value suggests a leftward tilt [34].

The subsequent parameter, kurtosis, indicates the peak value of the data distribution curve and the thickness of its tail [35, 36]. Based on kurtosis values, data distributions are classified as mesokurtic (normal distribution, kurtosis of 3), platykurtic (negative kurtosis, kurtosis less than 3, light-tailed curve with fewer outliers), or leptokurtic

(kurtosis greater than 3, heavy-tailed curve with more outliers) [37].
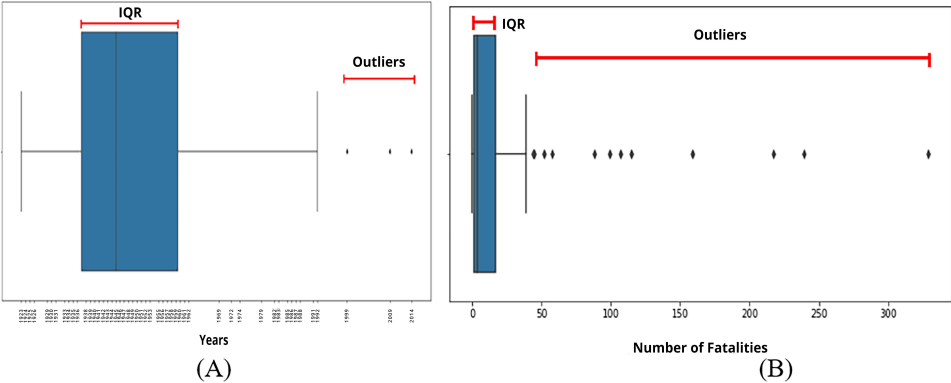
Figure 5. (A) Measures of dispersion visualization on variable years; (B) Measures of dispersion visualization on total fatalities

The computation of skewness and kurtosis for the dataset yielded a positive skewness (0.9979) for the 'year' variable, closely approximating normal distribution. This implies a rightward tilt of the distribution curve [38]. The corresponding kurtosis was 0.42503, suggesting a platykurtic kurtosis type, hence fewer outliers. On the other hand, the 'fatalities' variable revealed a skewness of 3.78 and kurtosis of 15.37, indicating a positively skewed and leptokurtic distribution. This suggests a considerable number of outliers in the fatalities data. These findings corroborate the graphical representations in Figures 4 and 5. The respective skewness and kurtosis values are summarized in Table 3.

Table 3. The results of skewness and kurtosis parameters

Curve Parameters	Years	Fatalities
Skewness	0.9979	3.76
Kurtosis	0.42503	15.37

3.3.2 Categorization of qualitative data variables

The third variable of consideration, denoted as 'probable causes', falls under the qualitative data type. This variable encompasses various categories of causative factors for commercial aircraft accidents, spanning from 1918 to 2020. A computation of the frequency and proportion of each category, based on their association with the accidents, is carried out.
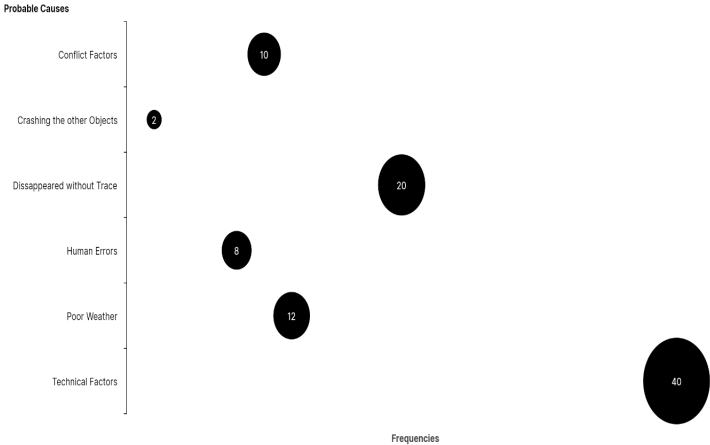
Figure 6. Visualization of the frequency of occurrence of a category of causes of accidents in the probable causes' variable

The dataset categorizes probable causes into six variables: conflict factors, collision with other objects, disappearance without trace, human error, adverse weather, and technical factors. The frequency of occurrence of these

probable causes is visualized in Figure 6.

Analysis of Figure 6 reveals that accidents attributed to technical factors exhibited the highest frequency, appearing 40 times. 'Disappearance without trace' emerged as the second most frequent cause, recorded 19 times. Conversely, collisions with other objects were the least frequent, occurring only twice. The proportional representation of these causes, illustrated in Figure 7, demonstrates that out of 91 commercial aircraft accidents recorded between 1918-2020, 44% were caused by technical factors and 20.9% were due to aircrafts disappearing without trace. Adverse weather was a factor in 13.2% of the accidents, while geopolitical conflicts after the Second World War accounted for 11%. Human error and collision with other objects comprised 8.8% and 2.2% of accidents, respectively.
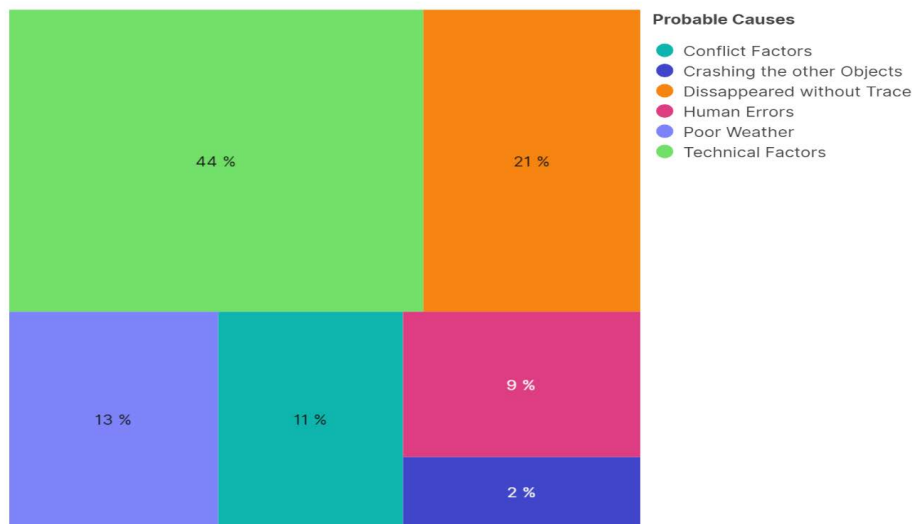

Figure 7. The percentage of each category of causes of accidents in the probable causes' variable

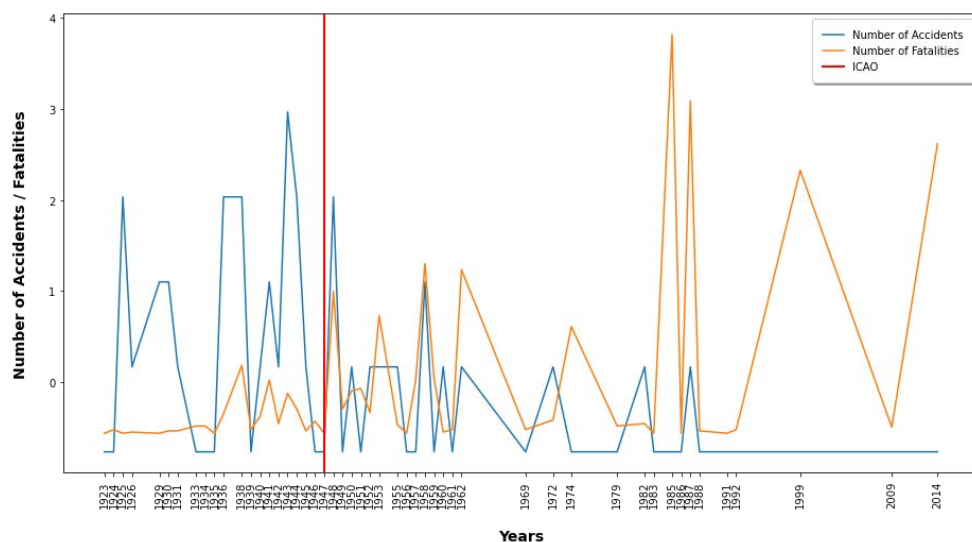## 3.4 Exploratory Data Analysis (EDA)

Figure 8. The graph of the number of B3A commercial aircraft accidents and fatalities per year from 1918 to 2020

Following the descriptive statistical analysis, the inherent characteristics of each variable, such as the frequency of occurrence of each data category, data distribution, central location of data, etc., were identified. Thereafter, the Exploratory Data Analysis (EDA) method was employed to unearth relationships between accidents, fatalities, and probable causes within the timeframe of 1918–2020. Further attention was given to the impact of data in the post-ICAO era (after 1948), illustrated in diagrammatic visualizations.

Initially, EDA was applied to the 'year' and 'fatalities' variables to ascertain the number of accidents. This exploration is depicted in Figure 8, where a timeline barrier denotes the formation year of the ICAO (1947). This visualization demonstrates a significant reduction in the number of aircraft accidents per annum post the ICAO establishment. Additionally, it portrays a decline in the density of accidents in specific periods. In the pre-ICAO era, 48 accidents led to 226 fatalities, while in the post-ICAO era, despite a decrease in the number of accidents to 43, fatalities dramatically increased to 1883.
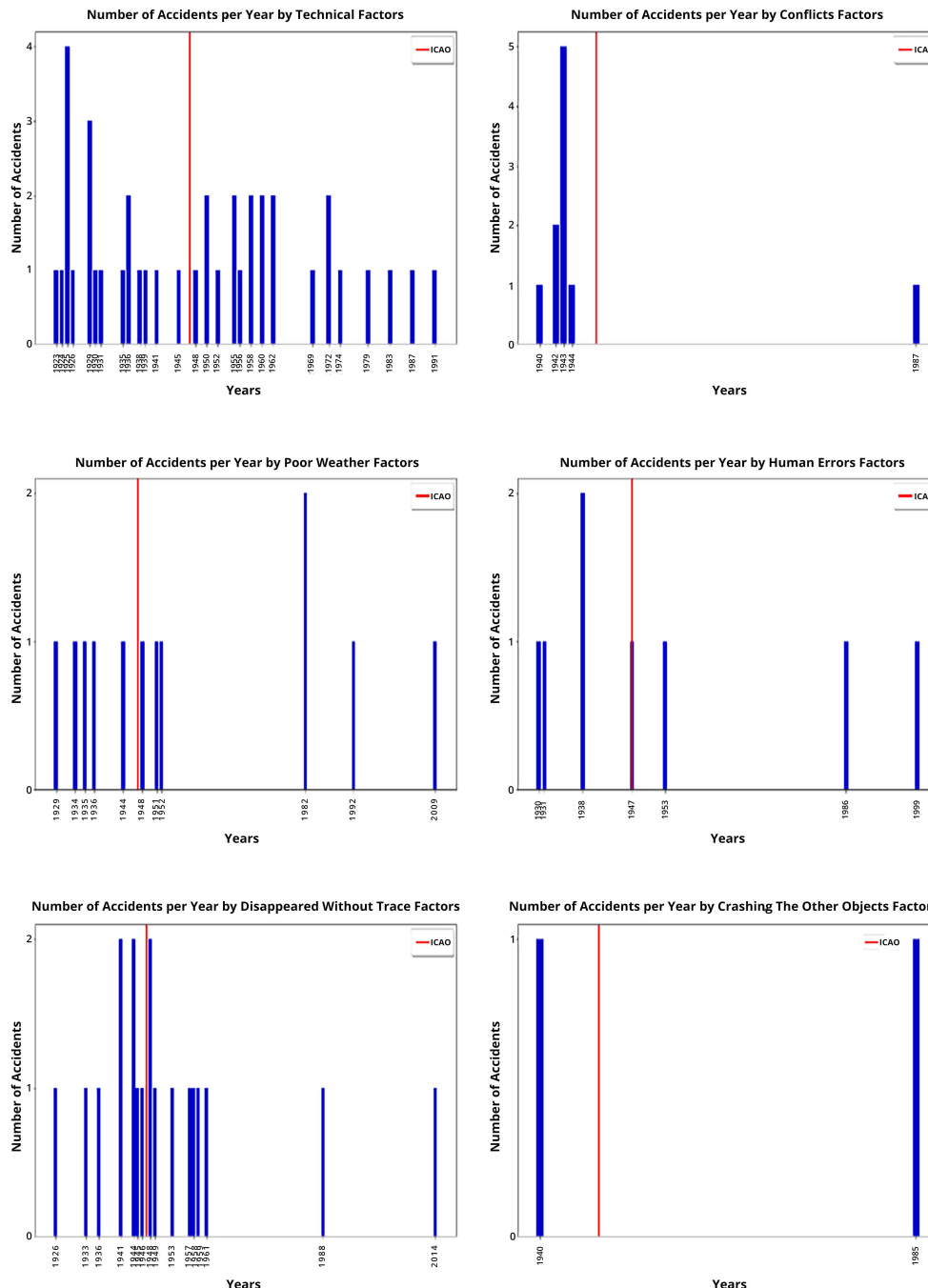
Figure 9. The graph of the number of commercial aircraft accidents per year based on each cause of accidents

The data visualization in Figure 8 underpins a reduction in the annual rate of commercial aircraft accidents post the establishment of the ICAO. However, when the causative factors are examined (Figure 7), technical factors before ICAO formation led to 19 accidents over 19 years (1918-1947), which translates to an average of 1 accident per year. In contrast, in the post-ICAO period spanning 72 years (1948-2020), 21 cases were recorded, reducing the average accident rate to 0.291 per year. This reduction suggests a significant decline of almost 70.9% in the number of accidents attributed to technical factors.

Similar patterns are observed for other factors. Conflict factors were responsible for 9 accidents, averaging 0.47 per year in the pre-ICAO period, but post-ICAO, there was only one incident in 1987, bringing the average down to 0.013 per year. Furthermore, human error led to 5 accidents per year, on average, before the formation of ICAO, which reduced to an average of 0.041 per year, based on the 3 incidents recorded post-ICAO.

Weather-related incidents followed a similar trend. Prior to the ICAO establishment, 5 cases were reported, averaging 0.263 per year. However, in the post-ICAO period, the number of cases rose to 7, but given the longer duration of 72 years, the average rate fell to 0.097 per year.

In contrast, the category 'disappearance without trace' saw an increase in the number of incidents from 9 in the pre-ICAO period, with an average rate of 0.47 per year, to 10 in the post-ICAO period. Yet, due to the longer duration of 72 years, the average yearly rate still decreased to 0.138. Lastly, the 'collision with other objects' category recorded one case each in the pre- and post-ICAO periods, with average accident rates of 0.05 and 0.013 per year, respectively. The reduction in these averages is based on the findings from the EDA using Python programming and is visualized in Figure 9.

It must be noted that the global aviation industry has seen significant evolution to meet increasing demands, especially concerning aircraft size [39]. The decline in accident frequency post the establishment of the ICAO can be attributed to advancements in safety systems, technology, and aircraft design [7], in compliance with ICAO-standardized safety rules and regulations [1]. These measures have led to safer civil aviation operations globally. The late 20th century saw a surge in international travel and a subsequent increase in passenger volume, prompting aircraft manufacturers to scale up aircraft size to cater to these demands. A landmark development was Boeing's introduction of the jumbo jet, Boeing 747, in 1969. While the larger aircraft could carry more passengers, it also meant that any accident would lead to a larger number of casualties. Evidence of this is seen in 1985, where despite only one recorded accident, the number of fatalities was significantly higher compared to other years.

It is important to acknowledge the limitations of this study. The data, sourced from the official B3A website, comprised 592 observations spanning 72 years of accident data. However, due to the specificity of the research objectives, the dataset was reduced to 91 observations, representing a confidence level of 15.37% of the historical accident data from 1908–2020.

4 Conclusions

Through the analytical framework employed in this study, an evaluation was conducted regarding the impact of the International Civil Aviation Organization's (ICAO) establishment. The analysis incorporated an extensive review of aviation accident data, as collected by the Bureau of Aircraft Accidents Archives (B3A), utilizing descriptive statistical methods and Exploratory Data Analysis (EDA).

Complications arose due to the dual presence of qualitative and quantitative data within the dataset, each exhibiting unique characteristics. Descriptive statistical techniques allowed for the extraction of the data's inherent character. On the other hand, to glean insights and discern patterns, the EDA methodology was implemented.

Descriptive statistical analysis yielded a robust dataset, particularly in terms of the 'year' and 'probable causes' variables. However, the 'fatalities' variable posed a challenge due to inconsistencies and a notable presence of outliers. Despite the disparate quality of the data across variables, EDA successfully elucidated the interrelation between the three.

Through the EDA analysis, the pivotal role of ICAO in reducing aviation accidents was highlighted. Notably, a significant decline in accident occurrences was observed following the establishment of the ICAO. More specifically, a substantial decrease of 70.9% per year on average was recorded for accidents caused by technical issues.

Interestingly, while the number of accidents decreased, the data revealed a considerable increase in the number of fatalities. This paradoxical trend can be attributed to the evolution of commercial aircraft design, which has been driven by market demand for larger passenger capacities. Therefore, while the number of accidents has been curtailed, those that do occur have the potential to result in a higher number of fatalities due to the larger aircraft sizes.

In summary, the analysis supports the hypothesis that the establishment of the ICAO had a profound impact on commercial aviation safety. It fostered an enhanced sense of trust in air travel and facilitated the growth of the aviation industry. Future research endeavors might benefit from incorporating analysis related to the rules issued by ICAO (Annexes), aligning with the year of issuance. Such an extension would provide a more holistic understanding of the ICAO's impact over time.

Author Contributions

Rossi Passarella: Writing- Original draft preparation;
Harumi Veny: Data Curation, Reviewing and Editing, Formal analysis;
Muhammad Fachrurrozi: Conceptualization, Methodology, Supervision, Reviewing Data curation;
Samsuryadi Samsuryadi: Formal analysis;

Marsella Vindriani: Investigation and Editing.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] J. Huang, "Aviation safety, ICAO and obligations erga omnes," *Chinese J. Int. Law*, vol. 8, no. 1, pp. 63–79, 2009. https://doi.org/10.1093/chinesejil/jmn039

[2] T. G. Puranik, N. Rodriguez, and D. N. Mavris, "Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning," *Transp. Res. Part C Emerging Technol.*, vol. 120, p. 102819, 2020. https://doi.org/10.1016/j.trc.2020.102819

[3] R. Passarella, G. Oktariani, D. Kurniawan, and P. Sari, "Using the agglomerative hierarchical clustering method to examine human factors in Indonesian aviation accidents," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, 2022. https://doi.org/10.14569/ijacsa.2022.0130938

[4] G. D. Gosling, "Use of air passenger survey data in forecasting air travel demand," *Transp. Res. Rec.*, vol. 2449, no. 1, pp. 79–87, 2014. https://doi.org/10.3141/2449-09

[5] V. E. Zhukov, "Demand analysis models for passenger air transportation," *World Transp. Transp.*, vol. 18, no. 1, pp. 134–144, 2020. https://doi.org/10.30932/1992-3252-2020-18-134-144

[6] Transportation Fatalities by Mode, *Bureau of transportation statistics*. https://www.bts.gov/content/transportation-fatalities-mode

[7] D. Kelly and M. Efthymiou, "An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017," *J. Saf. Res.*, vol. 69, pp. 155–165, 2019. https://doi.org/10.1016/j.jsr.2019.03.009

[8] C. C. Rodrigues, "Aviation safety: Commercial airlines," *Inter. Encycl. Transp.*, vol. 2021, pp. 90–97, 2021. https://doi.org/10.1016/b978-0-08-102671-7.10113-7

[9] A. Barnett, "Aviation safety: A whole new world?" *Transp. Sci.*, vol. 54, no. 1, pp. 84–96, 2020. https://doi.org/10.1287/trsc.2019.0937

[10] P. Illankoon, P. Tretten, and U. Kumar, "A prospective study of maintenance deviations using HFACS-ME," *Inter. J. Ind. Ergon.*, vol. 74, p. 102852, 2019. https://doi.org/10.1016/j.ergon.2019.102852

[11] T. Brady, "The aircraft accident investigation that never was," *J. Aviat. Aerosp. Educ. Res.*, vol. 23, no. 3, pp. 1–39, 2014. https://doi.org/10.15394/jaaer.2014.1610

[12] J. Oehling and D. J. Barry, "Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data," *Saf. Sci.*, vol. 114, pp. 89–104, 2019. https://doi.org/10.1016/j.ssci.2018.12.018

[13] K. Herbst, G. Beckers, L. Harper, D. Bägli, A. Nieuwhof-Leppink, M. Kaefer, M. Fossum, and N. Kalfa, "Don't be mean, be above average: Understanding data distribution and descriptive statistics," *J. Pediatr. Neurol.*, vol. 16, no. 5, p. 712, 2020. https://doi.org/10.1016/j.jpurol.2020.09.007

[14] H. B. Sankaranarayanan, G. Agarwal, and V. Rathod, "An exploratory data analysis of airport wait times using big data visualisation techniques," in *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE, 2016. https://doi.org/10.1109/csitss.2016.7779379

[15] R. Passarella and S. Nurmaini, "Behavioral evidence of public aircraft with historical data: The case of Boeing 737 MAX 8 PK-LQP," *J. Appl. Eng. Sci.*, vol. 20, no. 4, pp. 1254–1262, 2022. https://doi.org/10.5937/jaes0-38696F

[16] E. Paradis, B. O'Brien, L. Nimmon, G. Bandiera, and M. A. Martimianakis, "Design: Selection of data collection methods," *J. Grad. Med. Educ.*, vol. 8, no. 2, pp. 263–264, 2016. https://doi.org/10.4300/jgme-d-16-00098.1

[17] H. Benhar, A. Idri, and J. Fernández-Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Comp. Methods Programs Biomed.*, vol. 195, p. 105635, 2020. https://doi.org/10.1016/j.cmpb.2020.105635

[18] A. Famili, W. Shen, R. Weber, and E. Simoudis, "Data preprocessing and intelligent data analysis," *Intell. Data Anal.*, vol. 1, no. 1–4, pp. 3–23, 1997. https://doi.org/10.1016/s1088-467x(98)00007-9

[19] S. García, J. Luengo, and F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," *Knowl. Based Syst.*, vol. 98, pp. 1–29, 2016. https://doi.org/10.1016/j.knosys.2015.12.006

[20] N. Donges, "Intro to descriptive statistics," 2021. https://towardsdatascience.com/intro-to-descriptive-statistics-252e9c464ac9

[21] E. Allaj, "Measuring variability and association for categorical data," *Fuzzy Sets Syst.*, vol. 421, no. 1, pp. 29–43, 2021. https://doi.org/10.1016/j.fss.2020.11.018

[22] S. S. Durduran, "Automatic classification of high resolution land cover using a new data weighting procedure: The combination of k -means clustering algorithm and central tendency measures (KMC–CTM)," *Appl. Soft Comput.*, vol. 35, pp. 136–150, 2015. https://doi.org/10.1016/j.asoc.2015.06.025

[23] J. Frost, "Measures of variability: Range, interquartile range, variance, and standard deviation," 2018. https://statisticsbyjim.com/basics/variability-range-interquartile-variance-standard-deviation/

[24] D. N. Joanes and C. A. Gill, "Comparing measures of sample skewness and kurtosis," *J. R. Stat. Soc. Ser. D*, vol. 47, no. 1, pp. 183–189, 1998. https://doi.org/10.1111/1467-9884.00122

[25] A. T. Jebb, S. Parrigon, and S. E. Woo, "Exploratory data analysis as a foundation of inductive research," *Hum. Resour. Manage. Rev.*, vol. 27, no. 2, pp. 265–276, 2017. https://doi.org/10.1016/j.hrmr.2016.08.003

[26] C. Kneale and S. D. Brown, "Uncharted forest: A technique for exploratory data analysis," *Talanta*, vol. 189, pp. 71–78, 2018. https://doi.org/10.1016/j.talanta.2018.06.061

[27] D. Kelly and M. Efthymiou, "An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017," *J. Saf. Res.*, vol. 69, pp. 155–165, 2019. https://doi.org/10.1016/j.jsr.2019.03.009

[28] M. Claiborne, "What is the ICAO, and what do they do?" 2021. https://aerocorner.com/blog/icao/, accessed June 10, 2021.

[29] G. Marshall and L. Jonker, "An introduction to descriptive statistics: A review and practical guide," *Radiography*, vol. 16, no. 4, pp. e1–e7, 2010. https://doi.org/10.1016/j.radi.2010.01.001

[30] C. Almonte, "Small sample confidence intervals for the mean of a positively skewed distribution," *FIU Electronic Theses and Dissertations*, 2008. https://doi.org/10.25148/etd.FI13101599

[31] K. A. Adamson and S. Prion, "Making sense of methods and measurement: Measures of central tendency," *Clin. Simul. Nurs.*, vol. 9, no. 12, pp. e617–e618, 2013. https://doi.org/10.1016/j.ecns.2013.04.003

[32] J. Frost, "Mean, median, and mode: Measures of central tendency," 2018. https://statisticsbyjim.com/basics/measures-central-tendency-mean-median-mode/

[33] S. Manikandan, "Measures of dispersion," *J. Pharmacol. Pharmacother.*, vol. 2, no. 4, pp. 315–316, 2011. https://doi.org/10.4103/0976-500x.85931

[34] The GoodData Community, "Normality testing - Skewness and kurtosis," 2021. https://community.gooddata.com/metrics-and-maql-kb-articles-43/normality-testing-skewness-and-kurtosis-241

[35] G. G. Gomes, "Descriptive statistics: Expectations vs. reality (exploratory data analysis)," 2021. https://towardsdatascience.com/descriptive-statistics-expectations-vs-reality-exploratory-data-analysis-eda-8336b1d0c60b

[36] M. Xu and P. Shang, "Analysis of financial time series using multiscale entropy based on skewness and kurtosis," *Phys. A*, vol. 490, pp. 1543–1550, 2018. https://doi.org/10.1016/j.physa.2017.08.136

[37] CFI Team, "Kurtosis-definition, excess kurtosis, and types of kurtosis," 2021. https://corporatefinanceinstitute.com/resources/knowledge/other/kurtosis/

[38] M. J. Blanca, J. Arnau, D. López-Montiel, R. Bono, and R. Bendayan, "Skewness and kurtosis in real data samples," *Methodology*, vol. 9, no. 2, pp. 78–84, 2013. https://doi.org/10.1027/1614-2241/a000057

[39] A. Bejan, J. D. Charles, and S. Lorente, "The evolution of airplanes," *J. Appl. Phys.*, vol. 116, no. 4, p. 044901, 2014. https://doi.org/10.1063/1.4886855