Statistical Data Analysis on Worldwide Airplane Crashes and Fatalities in Domestic Scheduled Passenger Flights: Causes, Occurrences, and Insights

By

Francisco Canas Fuentes, Camilo Quiceno Quintero
University of Aberdeen

Abstract

The aim of this report is to provide insights into fatal and non-fatal airplane accidentology events by examining the most common causes of accidents and occurrences for worldwide domestic scheduled passenger flights since 1919. The investigation also allowed us to obtain valuable insights into changing trends in aviation safety. Graphical representation of data was used to highlight the recurrence of three major categories: accident causes, aircraft manufacturers, and airlines involved in accidents. The results obtained show that 'Runway' and 'CFIT' - or 'Controlled Flight into Terrain' events were the most commonly reported causes of non-fatal, and fatal aircraft accidents, respectively. Boeing and Douglas airplanes, along with the airline Aeroflot, have been involved in the greatest number of accidents. Finally, we were able to infer from trending annual data that aviation safety has increased throughout the years.

CONTENTS

1.0 Introduction

Since the invention of the first airplane in 1903 (Crouch, 1998)[1], the history of modern aviation has been a testament of the triumphs and ingenuity of mankind, but also a stark reminder of some of the most catastrophic incidents which plague our historical records. The collection and analysis of information related to aviation accidentology is of paramount importance to understanding the causes, occurrences and obtaining insights into aviation safety, which were the main objectives of this report.

This analysis was done by utilizing a dataset on "Aircraft Accidents, Failures, and Hijacks" since 1919, which was downloaded from the Kaggle website, and imported into Wolfram Mathematica. The first step was to extract the data on 'domestic scheduled passenger flights' which was the scope of this report. This was followed by the pre-processing of the raw data, which was incomplete, and contained inconsistent formatting. The cleaning of this data consisted largely of data munging, which included data manipulation, eliminating unwanted information, and reconstructing data values into a consistent format suitable for the analysis. The integration of a second dataset allowed us to observe the correlation and trends between different variables which were normalized and plotted together.

Exploratory analysis using data visualization was the main approach used to summarize our results and obtain an understanding of the cleaned data. The major variables of interest which were examined in this report included: accident causes, airplane brands/manufacturers, and airlines. These categories were extracted, sorted, counted and graphically represented using bar charts showing their total recurrence in all reported accidents.

In the first part of this report we determined the most common categories of non-fatal and fatal accidents. The results showed that non-fatal accidents were most commonly caused by 'Runway' related incidents, and 'CFIT' was the most common cause of fatal accidents. The 'Result' category accounted for the majority of all accident types. Our results from the second part of this report indicated that Boeing and Douglas airplanes have been involved in the greatest number of accidents, and the airline category with most accidents was led by Aeroflot. A choropleth map also showed that the majority of airplane accidents occurred within the United States and Russia, respectively. In the final part of our analysis we are able to conclude that air transportation has, overall, become safer throughout the years.

2.0 Source and Data Model

2.1. Data Source

The source of the data used for this report came from the Flight Safety Foundation's Aviation Safety Network (ASN)[1], a private and independent initiative founded in 1996 (Ranter, 1996)[2]. This report utilized data from the 'ASN Accident Database', which is a database containing documented accidents for airliners (12+ passengers), military transport and corporate jets from 1919 up to the present day (Ranter, 1996)[2]. The information contained in this database was mostly obtained from official governmental agencies consisting of civil aviation authorities and air accident investigation boards (Ranter, 1996)[2], and is therefore, considered reliable.

In the final part of this report we utilized an additional dataset on annual air transport passengers. This dataset was sourced from the International Civil Aviation Organization, Civil Aviation Statistics of the World and ICAO staff estimates[2], and downloaded directly from 'The World Bank' website[3].

2.2 Data Structure/Model & Explanation

The data was accessed under the name 'Aircraft Accidents, Failures & Hijacks Dataset' from Kaggle[4] - an online crowd-sourced data science platform, and downloaded as a 'CSV' file type under the name 'Aircraft_Incident_Dataset' onto the local computer as an Excel file. The structure of the downloaded raw data consisted primarily of character strings, as illustrated in Figure 1:

Incident_Date	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category
03-Jan-22	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage
04-Jan-22	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage
05-Jan-22	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage
08-Jan-22	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-T	Cargo	Accident | hull-loss
12-Jan-22	Beechcraft 200 Super King Air		private	Illegal Flight	Criminal occurrence (sabotage, sh
22-Jan-22	Airbus A320-232	N760JB	JetBlue Airways	Domestic Scheduled Passenger	Accident | repairable-damage
31-Jan-22	Beechcraft B300 King Air 350	C-GEAS	Air Tindi, opf RCAF	Military	Accident | repairable-damage
05-Feb-22	Cessna 208B Grand Caravan	5H-TAZ	Care Aviation	Unknown	Accident | repairable-damage
08-Feb-22	Antonov An-26	SP-402	South Sudanese Air For	Military	Accident | repairable-damage
11-Feb-22	Antonov An-2R	RA-33599	Kamchatsky Krechet	Cargo	Accident | hull-loss
14-Feb-22	Let L-410UVP-E3	9S-GFA	Doren Air Congo	Cargo	Accident | hull-loss

Figure 1: Structure of the raw data downloaded from Kaggle. The dataset is structured as a list of associations consisting of 23 named columns and 23,520 rows.

[1] Aviation Safety Network > ASN Aviation Safety Database (aviation-safety.net)
[2] Statistics (icao.int)
[3] Air transport, passengers carried | Data (worldbank.org)
[4] Aircraft Accidents, Failures & Hijacks Dataset | Kaggle

A directory path [directoryPath] was defined on a Wolfram Mathematica notebook for the location of the data, and set using the 'SetDirectory' function. The data was imported onto Mathematica using the 'Import' function and defined as a variable called 'rawData'. The imported raw dataset was composed of a list of associations containing 23,520 rows corresponding to 23 named columns of information from 1919 to 2022. The columns used for this report included: "Incident Date", "Aircraft Model", "Aircraft Nature", "Incident Category", "Incident Causes", and "Departure Airport".

This study focused exclusively on the data corresponding to the 'Domestic Scheduled Passenger' category, therefore, all other data related to a different aircraft nature was disregarded. Missing or unknown values were also found throughout the dataset which were likely due to lost data or information which was unattained from accident reports.

3.0 Methodology

The initial state of the imported raw data used for this analysis was characterized as 'dirty' due to the fact that it was incomplete, inconsistent, and contained an abundance of information not relevant to the scope of this study. In particular, the data contained missing element values, duplicate data, and discrepancies in attribute formatting. For these reasons, the cleaning and preprocessing of the data was an essential requirement for this analysis. The steps and methods which were followed using Wolfram Mathematica are summarized below.

3.1. Data Preprocessing

Data Cleaning Tasks & Methods

1. *Data Segmentation & Reduction:*

   The first step in our data cleaning process was to eliminate non-desired information (i.e. data reduction) by extracting all rows of data containing the 'Domestic Scheduled Passenger' category. This was performed using the *Cases* function with pattern matching for two distinct groups: non-fatal accidents, and accidents which resulted in at least one fatality (i.e. fatal accidents). Throughout this analysis, columns of interest were segmented via extraction from the dataset using the *Parts* function in order to examine each unique area that was addressed in the report. Individual character strings within list of lists were also extracted to obtain tallied counts of specific variables.

2. Data Munging & Manipulation

Many of the elements found in the dataset contained multiple character strings, duplicate values, delimiters, and inconsistencies in formatting which made data munging a necessary part of the data cleaning process. The cleaning of this data required extensive manipulation using various functions as follows: for attributes containing multiple duplicate values, elements were separated by the "," delimiter using the *StringSplit* function. This was done for categories of accidents, accident subcategories (causes), aircraft manufacturers, and airlines. The *StringReplace* function was used to manipulate the data into a consistent format by eliminating unwanted strings resulting from inconsistencies in naming conventions. Non-desired information resulting from unknown or duplicate entries was eliminated using the *DeleteCases* and *DeleteDuplicates* functions, respectively. Finally, manipulated string names corresponding to airlines and countries were concatenated using the *StringJoin* function.

3. Unified Date Format:

The dataset contained dates which were formatted with both strings and numerical values (e.g. 01-Jan-10) and needed to be reformatted to allow the *DateObject* function to be applied. This was done by converting the expression for months into a numerical value (e.g. 2010,1,10) by using the *ToExpression* and *StringSplit* functions.

3.2. Techniques

1. *Data Integration & Transformation*

In the final part of this report we examined annual trends in fatalities and accidents, utilizing the second dataset containing numbers of annual passengers from 1971 to 2021. One of the challenges from the integration of both datasets was data value conflicts resulting from different scales in attribute values. A min-max normalization was applied to the data when plotting on a histogram to ensure all variables were plotted on a normalized scale between 0.0 to 1.0.

2. *Exploratory & Diagnostic Analysis*

Visualization techniques involving the plotting of data on bar chart graphs were utilized throughout this report to obtain an understanding of the relationships and distributions of different categories. A diagnostic analysis was performed in the final part of this

report by plotting historical annual data on accidents, fatalities, and passengers on a single histogram. This was done by creating an association for 'year' and number of fatalities, which allowed the total number of annual fatalities to be counted and plotted. The aim of this type analysis was to observe the correlations between variables to obtain insights into flight safety based on their relative changes.

4.0 Results

The results that were obtained from this analysis are presented in three separate parts, each addressing a unique area which was investigated in this report.

4.1. Part 1: Accident Categories and Causes

In this section of the analysis our aim was to determine which category of accidents was most commonly reported to have occurred on domestic passenger flights, and what the most common underlying causes of these accidents were. The analysis was done for accidents which resulted in zero fatalities (i.e. non-fatal accidents) as well as those which resulted in at least one fatality (i.e. fatal accidents).

The most common type of accident category for both non-fatal and fatal accidents was determined to be a 'Result' type accident, which accounted for a total of 1,756 non-fatal accidents (Figure 2.) and 1,281 fatal accidents (Figure 3.) between 1919 and 2022. The five most common accident categories are displayed below in Figures 2-4.

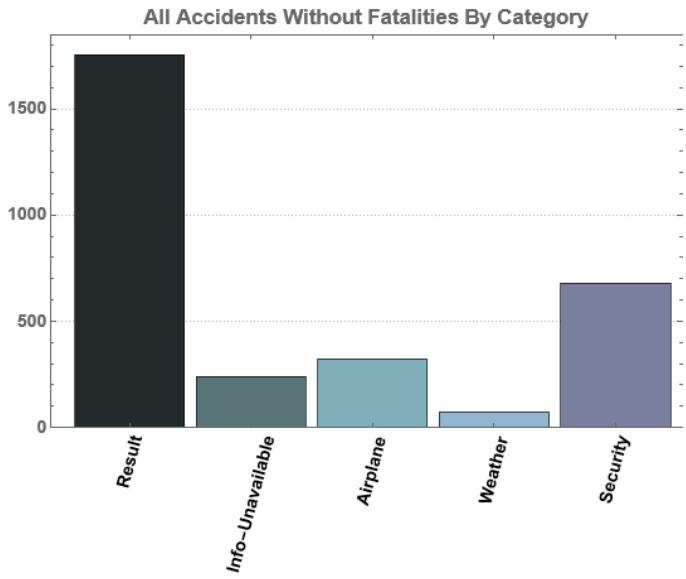

Figure 2. A bar chart showing the 5 most common accident categories for non-fatal accidents. The 'Result' category was found to be the most common type, and accounted for 1,756 non-fatal accidents from 1919 to 2022.

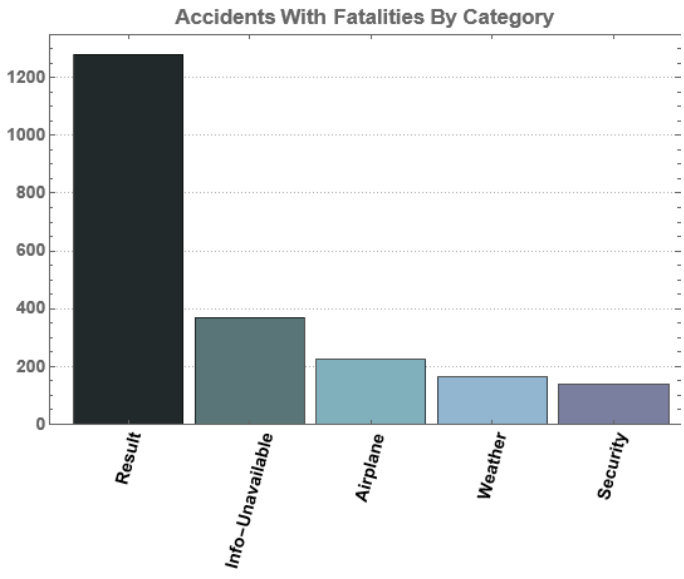

Figure 3. A bar chart showing the 5 most common accident categories for fatal accidents. The 'Result' category was found to be the most common type, and accounted for 1,281 accidents with fatalities from 1919 to 2022.

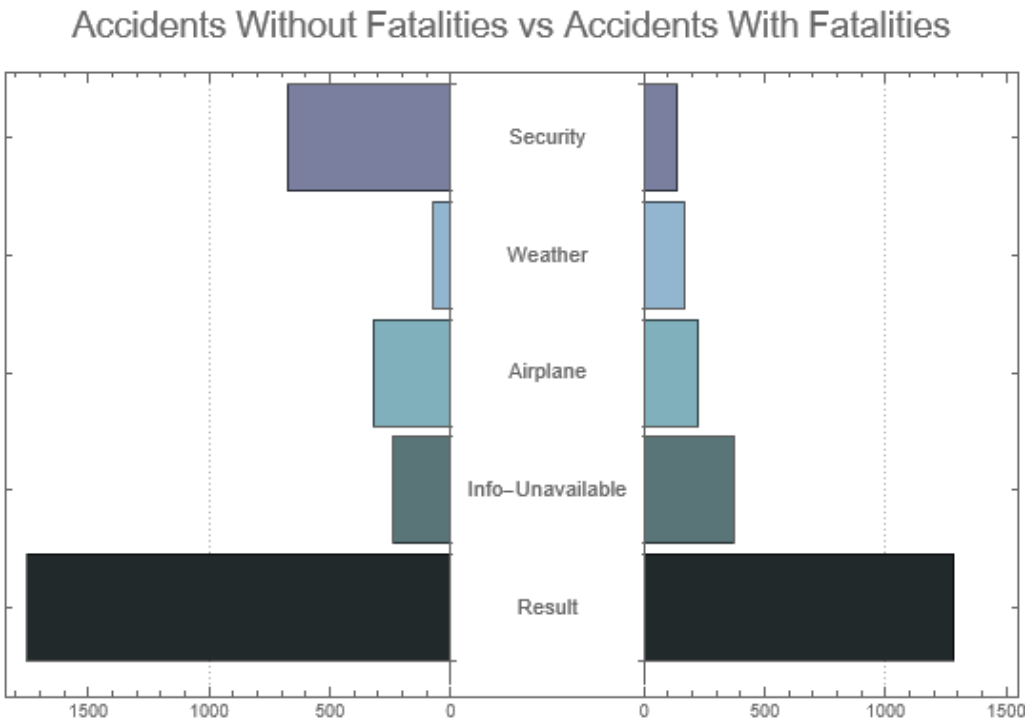

Figure 4. A paired bar chart showing a comparison between the 5 most common accident categories for non-fatal accidents and accidents with at least one reported fatality. For both cases, the 'Result' category was found to be the most commonly occurring category of all airplane accidents.

The 'Result' category was composed of many different types of accident subcategories or 'causes', which were also examined in this report. The results indicate that 'Runway' related incidents were the most common cause of non-fatal accidents, accounting for 743 reported incidents (Figure 5.). These results are consistent with claims made by (Eekeren et al., 2018, p. 261)[3], which indicate that the majority of aviation accidents are related to the runway. Furthermore, according to a runway safety report conducted by (IATA, 2015)[4], 39 percent of all commercial aviation accidents between 2010-2014 occurred on the runway.

The most commonly occurring cause of fatal airplane accidents was due to 'Loss of Control' or 'LOC-I' events (Figure 6.), which are categorized as accidents in which the flight crew loses control of the aircraft's intended flight path (IATA, 2019)[5]. This was followed closely by 'CFIT' or Controlled Flight Into Terrain and 'Runway' accidents as the most fatal causes. These results are illustrated in Figures 5-7. The results we obtained also correlate with a peer-reviewed publication from the University Aviation Association (UAA), which, according to (Huang, C., 2020)[6], "Loss of Control In-Flight (LOC-I), runway excursion, and Controlled Flight into Terrain

are the three most common fatal accident categories in scheduled commercial jet airplanes" (p.89).

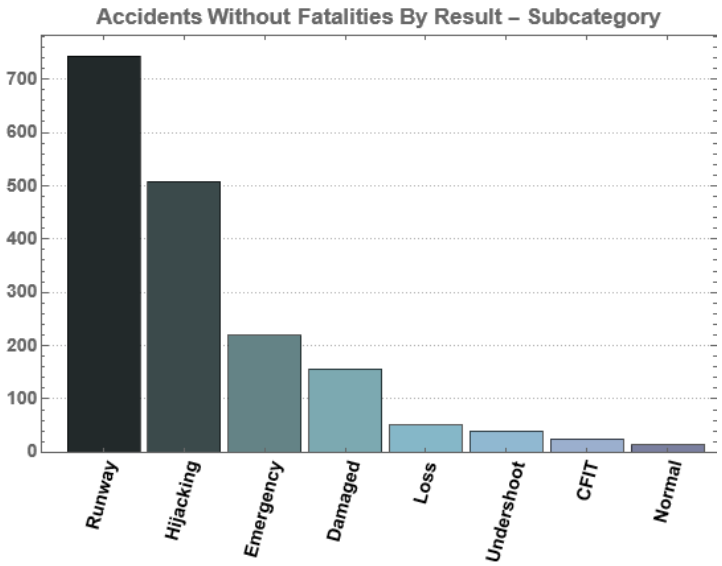

Figure 5. A bar chart showing the 5 most common accident causes by 'Result' subcategory for non-fatal accidents. Runway accidents were found to be the most common cause, accounting for 743 non-fatal accidents.

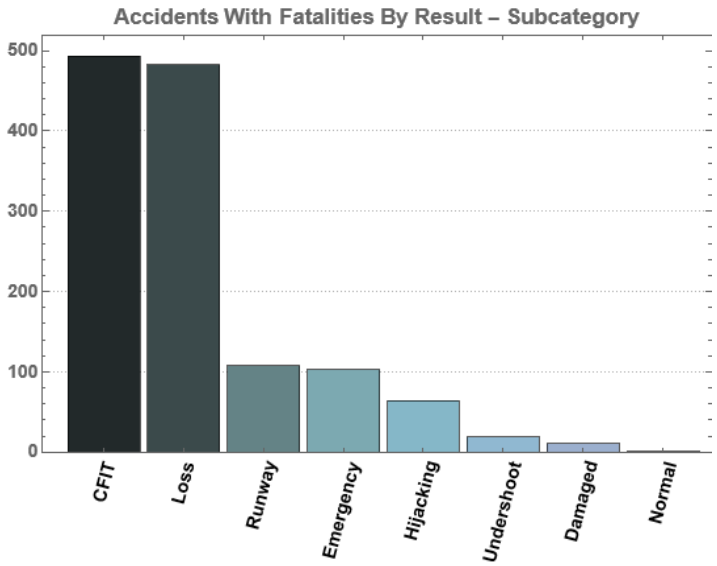

Figure 6. A bar chart showing the 5 most common accident causes by 'Result' subcategory for fatal accidents. CFIT accidents were found to be the most common cause, accounting for 493 fatal accidents.

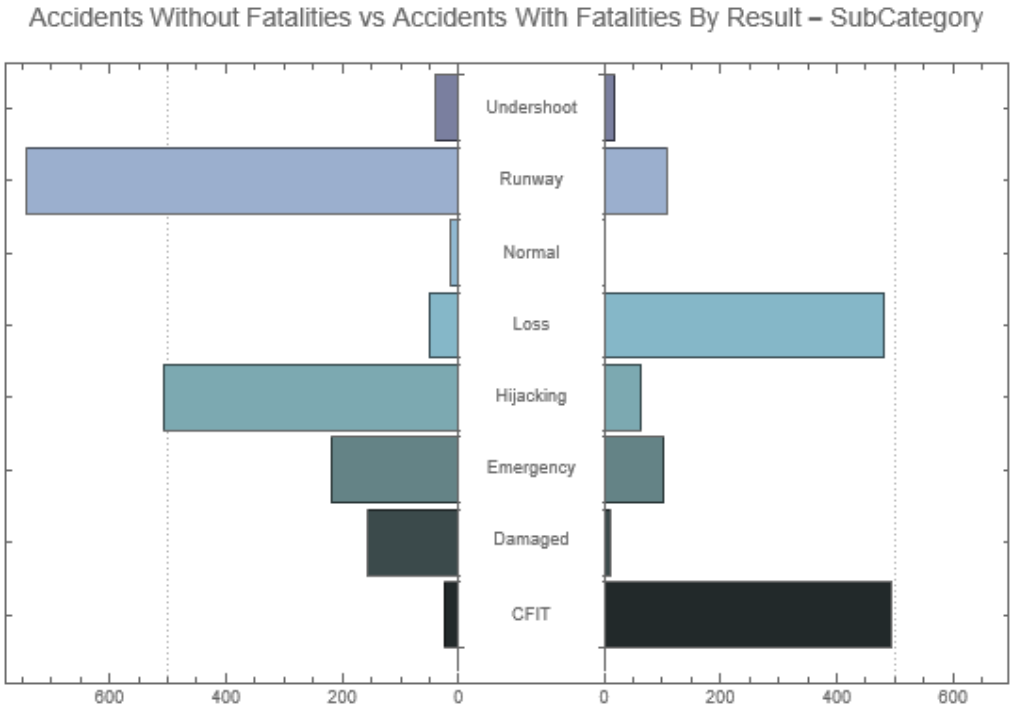

Figure 7. A paired bar chart showing a comparison between the 5 most common accident causes by 'Result' subcategory for non-fatal and fatal accidents.

4.2. Part 2: Airplane Brands, Airlines and Accident Locations

In the second part of this study we determined which aircraft manufacturers and which airlines had been involved in the greatest number of accidents. Our results showed that 'Boeing Commercial Airplanes' had the greatest occurrence of accidents without fatalities, with a total of 498 non-fatal documented accidents (Figure. 8). Douglas Aircraft Company was involved in the greatest number of fatal accidents, with a total of 399 cases, shown in Figure 9 below.

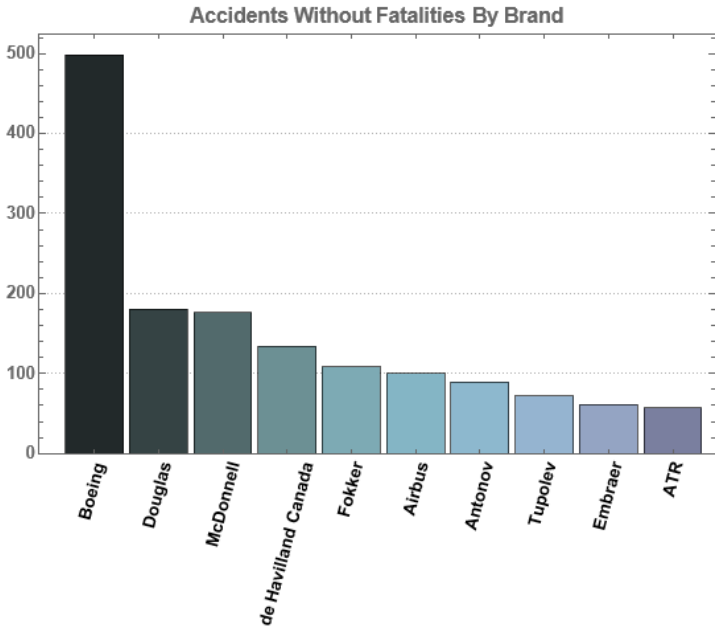

Figure 8. A bar chart showing the 10 aircraft manufacturers involved in the greatest number of non-fatal domestic passenger accidents from 1919-2021. Boeing accounted for a total of 498 accidents.

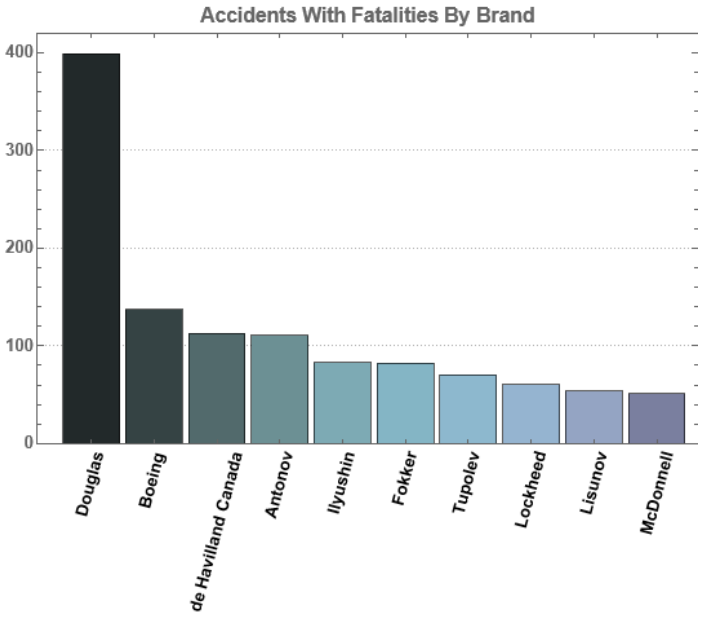

Figure 9. A bar chart showing the 10 aircraft manufacturers involved in the greatest number of fatal domestic passenger accidents from 1919-2021. Douglas accounted for a total of 399 accidents.

It is important to note that these results were not normalized based on fleet sizes. The number of existing airplanes for each aircraft manufacturer will vary largely depending on when each company became operational, defunct, or merged with other companies. An independent study conducted by (Herrera, J.M, et al., 2009, p.125)[7], showed that Boeing airplanes accounted for the greatest number of accidents from 2000-2007, before normalizing values. It is also important to note that an airplane manufacturer's involvement in an accident does not necessarily imply a lack of mechanical or structural safety from the manufacturer.

Our analysis on airline companies showed that ‘Aeroflot’ was the airline that had been involved in the greatest number of both fatal and non-fatal accidents. Aeroflot was involved in a total of 158 non-fatal accidents, and 282 accidents with fatalities. These results are shown in Figures 10-11.

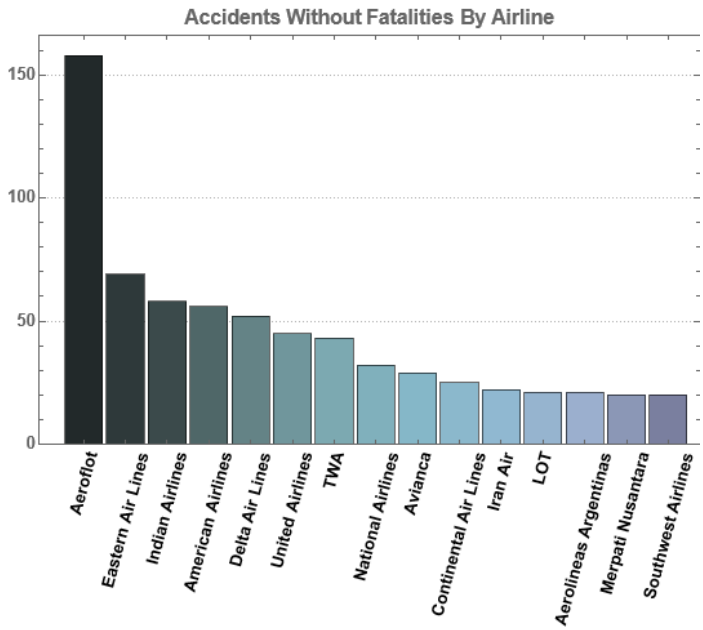

Figure 10. A bar chart showing the 15 airlines involved in the greatest number of non-fatal domestic passenger accidents from 1919-2022. Aeroflot was involved in 158 accidents.

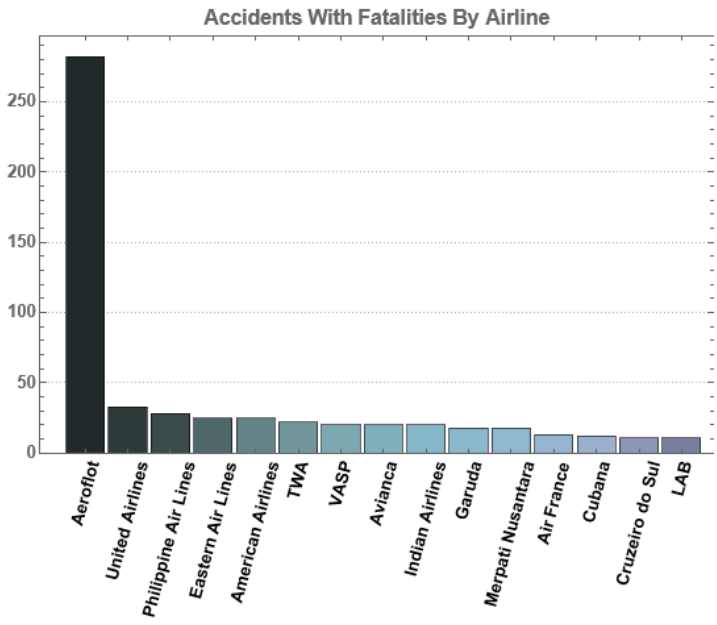

Figure 11. A bar chart showing the 15 airlines involved in the greatest number of fatal domestic passenger accidents from 1919-2022. Aeroflot was involved in 282 accidents

Once again, it is worth noting that these results are not normalized based on the number of commercial flights fulfilled by each airline. A greater involvement in accidents may be expected for an airline which conducts a larger amount of commercial flights, in comparison to those which fly less frequently or have a smaller fleet size.

The geographic distribution of accidents was also examined in this report to determine which country had the greatest number of reported fatal aircraft accidents. Accident locations are shown on a choropleth map (Figure 12.) as an aggregate summary of data showing the number of domestic fatal accidents which occurred within each country.

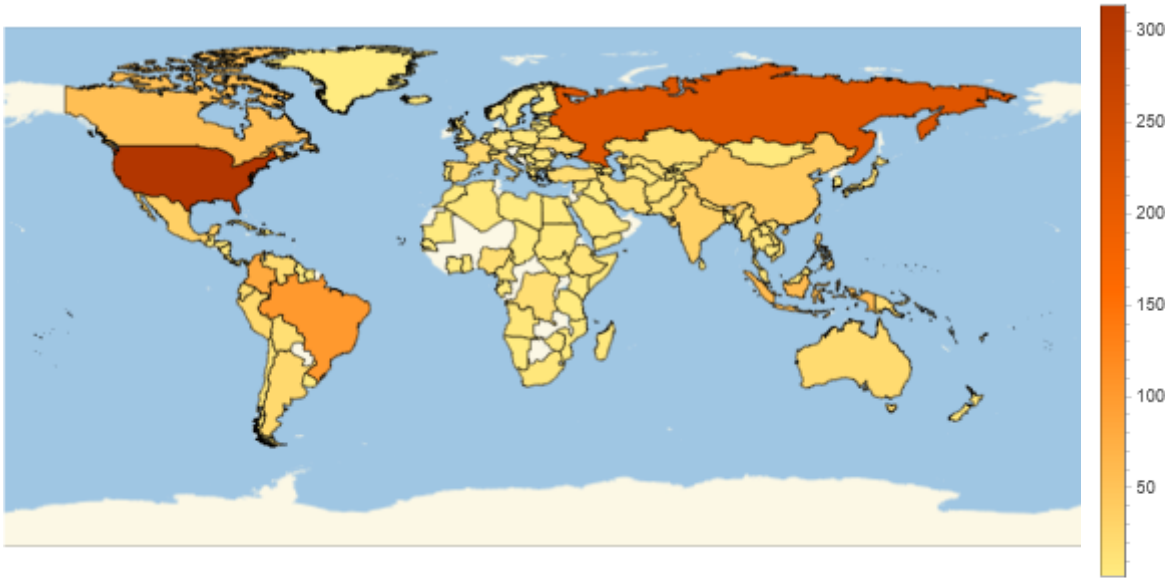

Figure 12. Choropleth map showing the global distribution of fatal accidents for domestic passenger flights per country. The United States was the country with the largest amount of fatal airplane accidents.

As seen in Figure 12, the greatest amount of fatal airplane accidents occurred within the United States. This was followed by Russia, and Brazil, respectively. Historically speaking, the United States has always been the largest aviation market in the world for international and domestic passenger traffic (ICAO, 2019)[8] until it was surpassed by China in recent years due to the impact of the Covid-19 virus (CAPA - Centre for Aviation, 2020)[9]. The results obtained from this report are therefore expected.

4.3. Part 3: Insights Into Air Transportation Safety

The objectives in the final part of this report were to determine if air transportation has become safer over time. Figures 13-14 show histograms of annual fatalities and accidents from 1928 to 2021 for domestic scheduled passenger flights, respectively. As expected, a general correlation between both variables can be observed, since fatalities are directly caused by airplane accidents. These variables are, therefore, causally related. Of particular interest is the peak in fatalities observed in 2001 (Figure 13), which corresponded to 3,088 airplane-related fatalities. This outlier was the result of the 9/11 terrorist attacks, which resulted in the deaths of 2,750 victims from four hijacked aircrafts (Bergen, 2003)[10].

The overall distribution of both graphs appear to follow a slightly left-skewed bell curve with peak fatalities and accidents occurring in the 1970's, and decreased numbers at the tail ends of both histograms. The lower number of fatalities and accidents which occurred in the early

1930's was likely the result of fewer aircraft flights happening during that time period. Figure 15 shows an exponentially increasing number of passengers from the 1970's up to 2021, which would naturally correspond to an increased number of airplane flights. The inverse relationship observed in Figure 15 between the decreasing annual accidents and fatalities against increasing passengers strongly suggests that accidents have decreased despite an increase in air transportation. We can conclude from this data, therefore, that air transportation safety has increased in recent years, and that fewer flights are resulting in accidents.

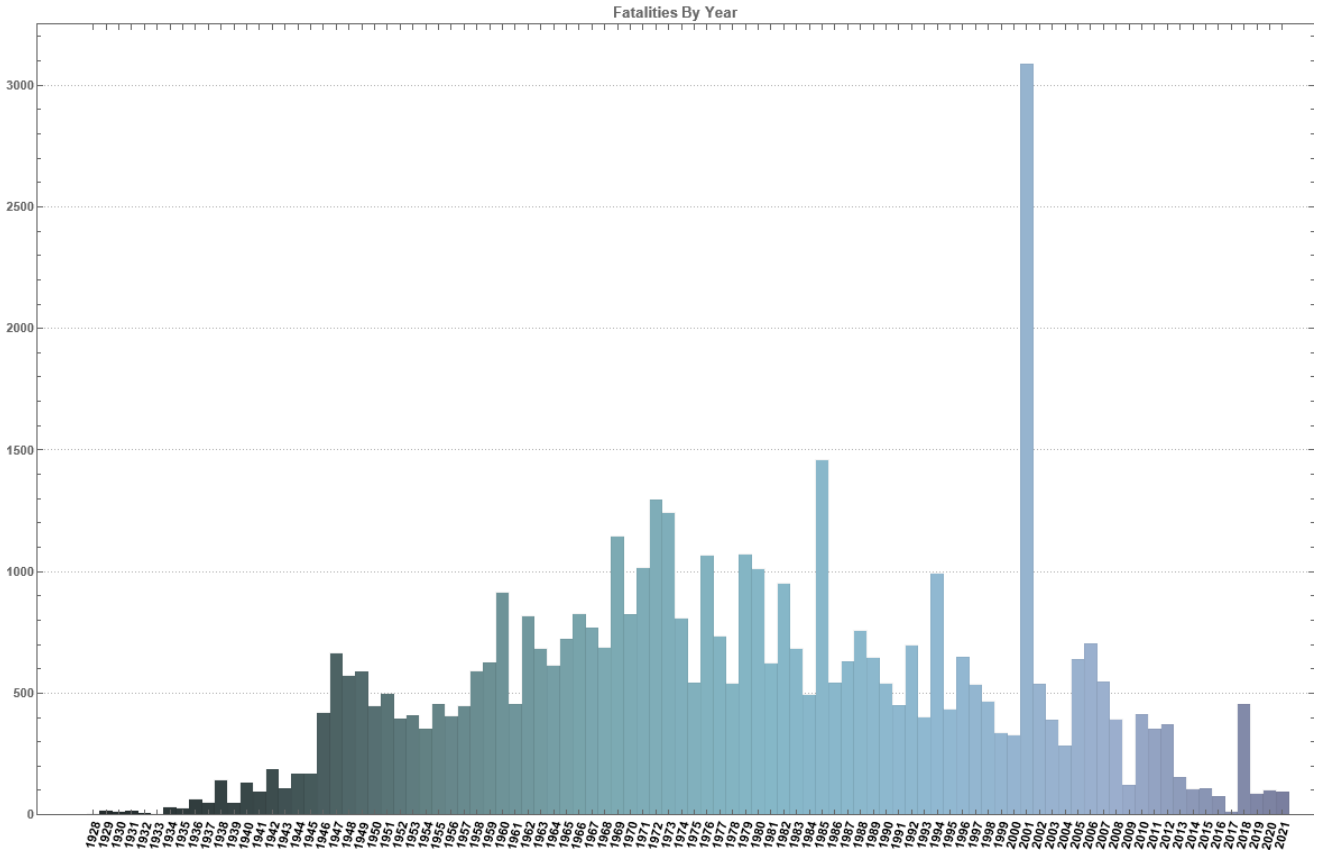

Figure 13. Histogram showing the number of annual airplane fatalities from 1928 to 2021 for domestic scheduled passenger flights.

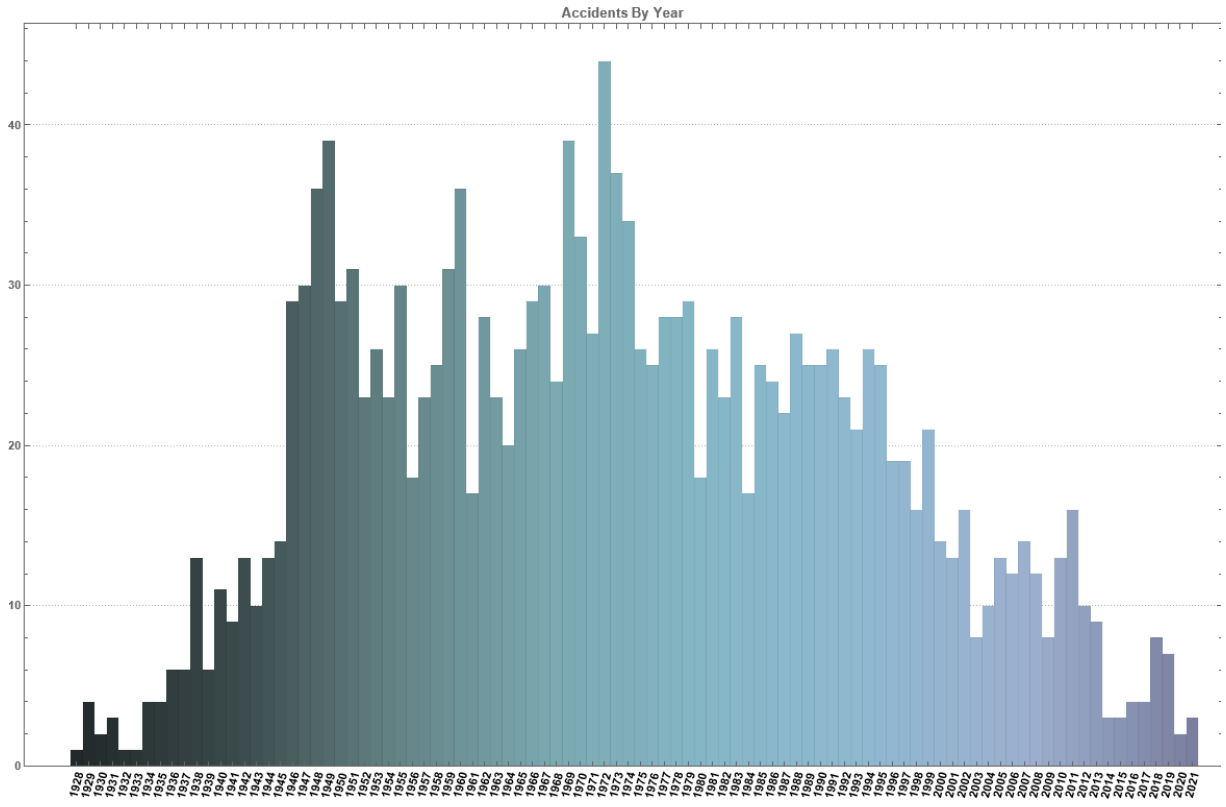

Figure 14. Histogram showing the number of annual airplane accidents from 1928 to 2021 for domestic scheduled passenger flights.

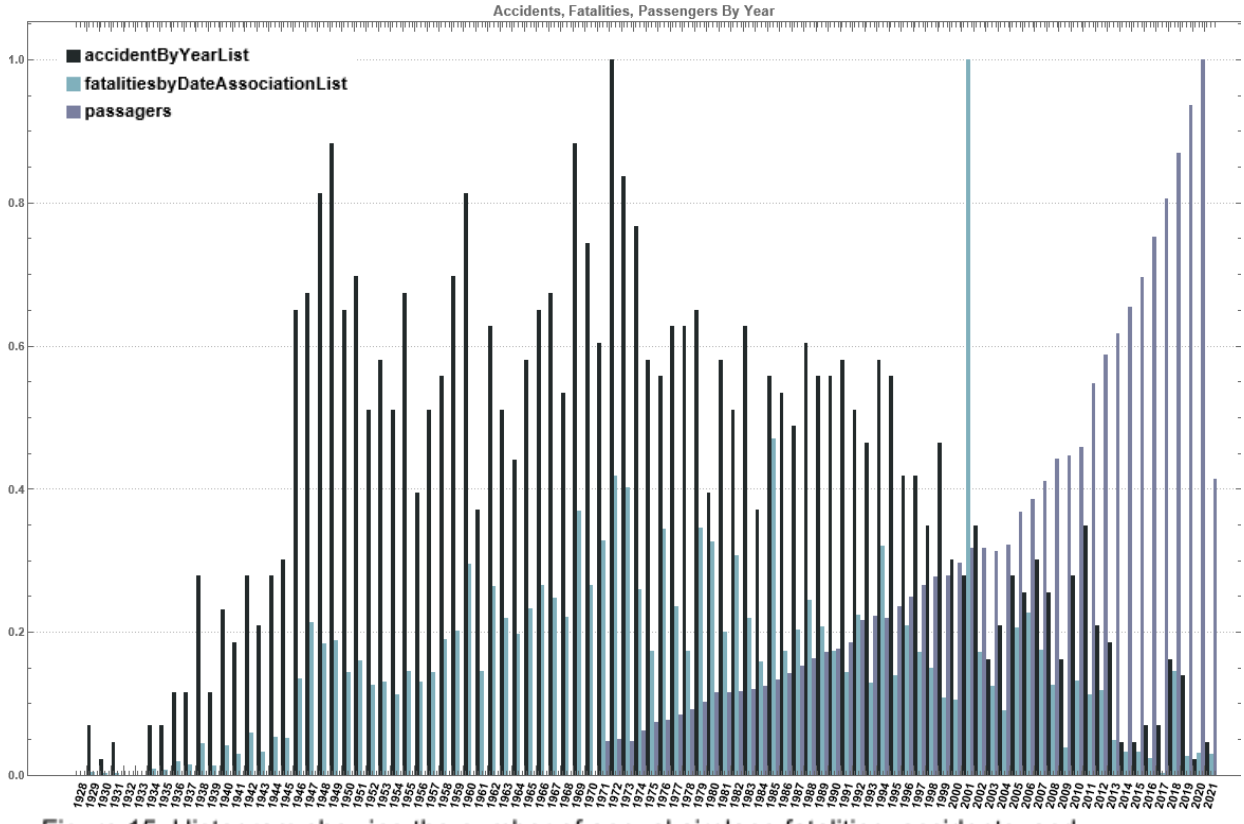

Figure 15. Histogram showing the number of annual airplane fatalities, accidents, and passengers on a normalized scale from 0.0 to 1.0.

5.0 Conclusion

The objectives of this report were to provide insights into airplane accidentology and safety by examining three major variables: accident causes, airplane manufacturers, and airlines. The results we obtained indicate that the most common cause of non-fatal accidents were related to runway incidents, while CFIT - 'Controlled Flight Into Terrain' was the most common cause of fatal accidents. These results were shown to be consistent with other publications in the field of airplane accidentology, which indicated that runway related excursions accounted for the majority of aircraft accidents. Our results also showed that Boeing airplanes have been involved in the greatest number of accidents, as well as the airlines, Aeroflot. The majority of airplane accidents were also observed to have occurred in the United States. Finally, by correlating data on annual airplane accidents, fatalities, and passengers, we were able to conclude that although air transportation has become more ubiquitous throughout the years, it has also become safer. This has likely been the result of increased safety and control measures implemented in commercial aviation.

It is important to acknowledge the various limitations of this study. Firstly, the scope of this report was limited to accidents resulting from domestic scheduled passenger flights only. Therefore, the results obtained may not be applicable across all flight nature categories. It is also important to consider that other variables such as the *total* number of flights - including non-accidented flights - and airline fleet sizes were not included in this dataset. These results are therefore not considered to be definitive, but rather insightful and contributory in nature.

6.0 References

[1] Crouch, T. D. (1998). *Wright Flyer of 1903*. Encyclopædia Britannica. Retrieved April 8, 2022, from https://www.britannica.com/topic/Wright-flyer-of-1903

[2] Ranter, H. (1996). *Aviation Safety Network > ASN Aviation Safety Database*. Aviation Safety Network > Retrieved April 8, 2022, from https://aviation-safety.net/database/

[3] Van Eekeren, R., Wright, S., & Čokorilo, O. (2018). Early cost safety analysis of runway events. *International Journal for Traffic & Transport Engineering*, *8*(3), 261-270.

[4] IATA. (2015). *Runway Safety Accident Analysis Report*. Retrieved April 8, 2022, from http://robinsoncenter.org/contentassets/7a5cd514de9c4c63ba0a7ac21547477a/rsar-1st-2015-final-version.pdf

[5] IATA. (2019). *Loss of control in-flight accident analysis report 2019 ...* Retrieved April 10, 2022, from https://www.iata.org/contentassets/b6eb2adc248c484192101edd1ed36015/loc-i_2019.pdf

[6] Huang, C. (2020). Further improving general aviation flight safety: analysis of aircraft accidents during takeoff. *The Collegiate Aviation Review International*, *38*(1), 88-105.

[7] Herrera, J. M., & Vasigh, B. (2009). A basic analysis of aging aircraft, region of the world, and accidents. *Journal of Business & Economics Research (JBER)*, *7*(5), 121-132.

[8] ICAO. (2019). *The World of Air Transport in 2019 //* . The World of Air Transport in 2019. Retrieved April 12, 2022, from https://www.icao.int/annual-report-2019/Pages/the-world-of-air-transport-in-2019.aspx

[9] CAPA - Centre for Aviation. (2020, April 16). *China becomes the largest aviation market in the world*. CAPA - Centre for Aviation. Retrieved April 12, 2022, from https://centreforaviation.com/analysis/reports/china-becomes-the-largest-aviation-market-in-the-world-521779

[10] Bergen , P. L. (2003). *September 11 attacks*. Encyclopædia Britannica. Retrieved April 16, 2022, from https://www.britannica.com/event/September-11-attacks