Aviation Business Analysis

BY Edwin Maina

Data scientist at Qwerty Ltd

Overview

As part of our company's strategic initiative to diversify its portfolio, we are venturing into the aviation industry.
This expansion involves purchasing and operating airplanes for commercial enterprise.

I Edwin Maina, a data scientist at QWERTY company ltd was tasked with:
Performing thorough research and analysis assessing aircraft safety through historical accident trends to identify aircraft makes and models with the least incidence of accidents.

Problem Statement

Our company is expanding its portfolio into the aviation industry but lacks knowledge of the risks associated with purchasing and operating aircraft for commercial and private operations.

Objective

identify the aircraft models that present the lowest risk in terms of safety and operational reliability. This involves a comprehensive analysis of the provided dataset to evaluate the safety performance of various aircraft manufacturers and models.

Success Criteria

- Provide insights that will guide the decision-making process for the acquisition of aircraft, ensuring that we select models with proven safety records.
- Identification of a good aircraft that fits the company's criteria of low risk accidental rate and fatalities for successful entry to the aviation industry with completion of purchase of an aircraft with good return on investment

Key Stakeholders

- Company stakeholders,
- potential investors
- Government policy makers,
- competing airline businesses
- customers

Data Understanding

Data Source:

•National Transportation Safety Board (NTSB) dataset covering aviation incidents from 1948 to 2022.

Description of data

Key columns

•**Aircraft Information**: Make, model, engine type, and number of engines.
•**Incident Information**: Injury severity, total injuries, accident causes, flight phase, weather conditions.
•**Outcomes**: Fatalities, serious injuries, minor injuries, and number of uninjured passengers.

•Event Date: The date when the aviation accident or incident occurred.
• Location: The geographic location where the incident took place.
•Aircraft Make and Model: The manufacturer and specific model of the aircraft involved in the accident.
•Broad Phase of Flight: The general phase of flight during which the accident occurred.
•Total Fatal Injuries: The total number of fatalities resulting from the accident.
•Total Serious Injuries: The total number of serious injuries (non-fatal) resulting from the accident.
•Total Minor Injuries: The total number of minor injuries resulting from the accident.

Data Cleaning

Checking for and Dropping duplicates to ensure the data is workable

In this analysis, I aim to focus on the most relevant data to enable making of informed recommendations. I decided to perform data reduction by dropping certain columns from this dataset that had a high percentage of missing values and those not important for this analysis.

I handled Missing values in columns containing categorical data by dropping them.

continuous data: I imputed the missing values using the median, as it provides a more typical measure of central tendency and is less sensitive to outliers compared to the mean.

Exploratory Data Analysis (EDA)

Descriptive Statistics:

Modeling and Analysis

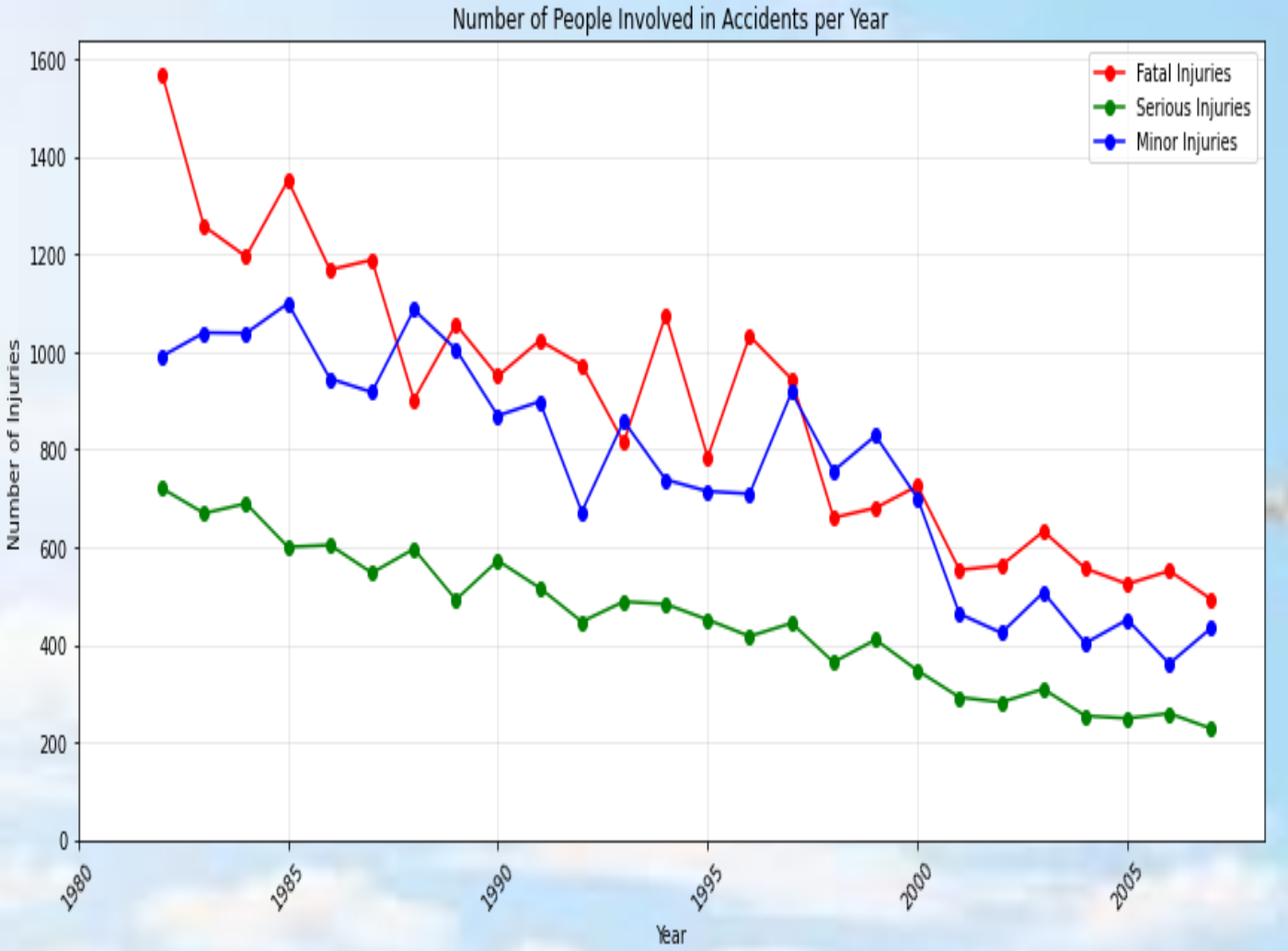
I analyzed Injuries and fatalities over time to identify patterns.

Visualizations: I used Graphs, such as bar plots, line plots, and pie charts to present data on injury severity, aircraft makes, and engine types.

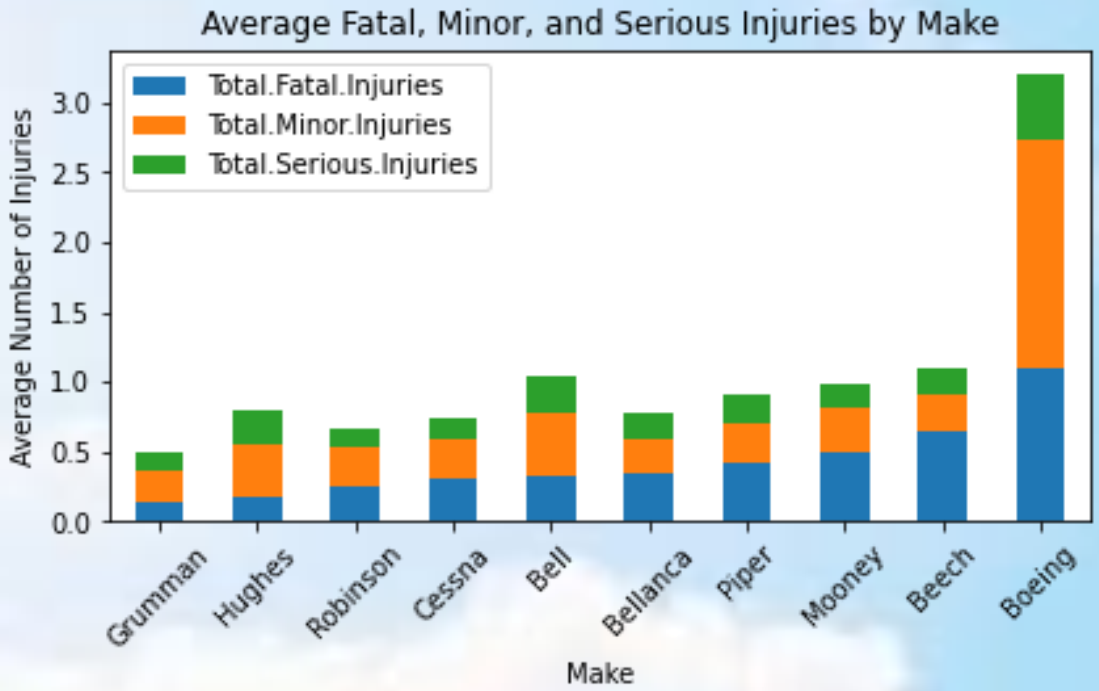
Injury Proportion Analysis: Created new features to compare the ratio of injuries to the total passenger count across different models.

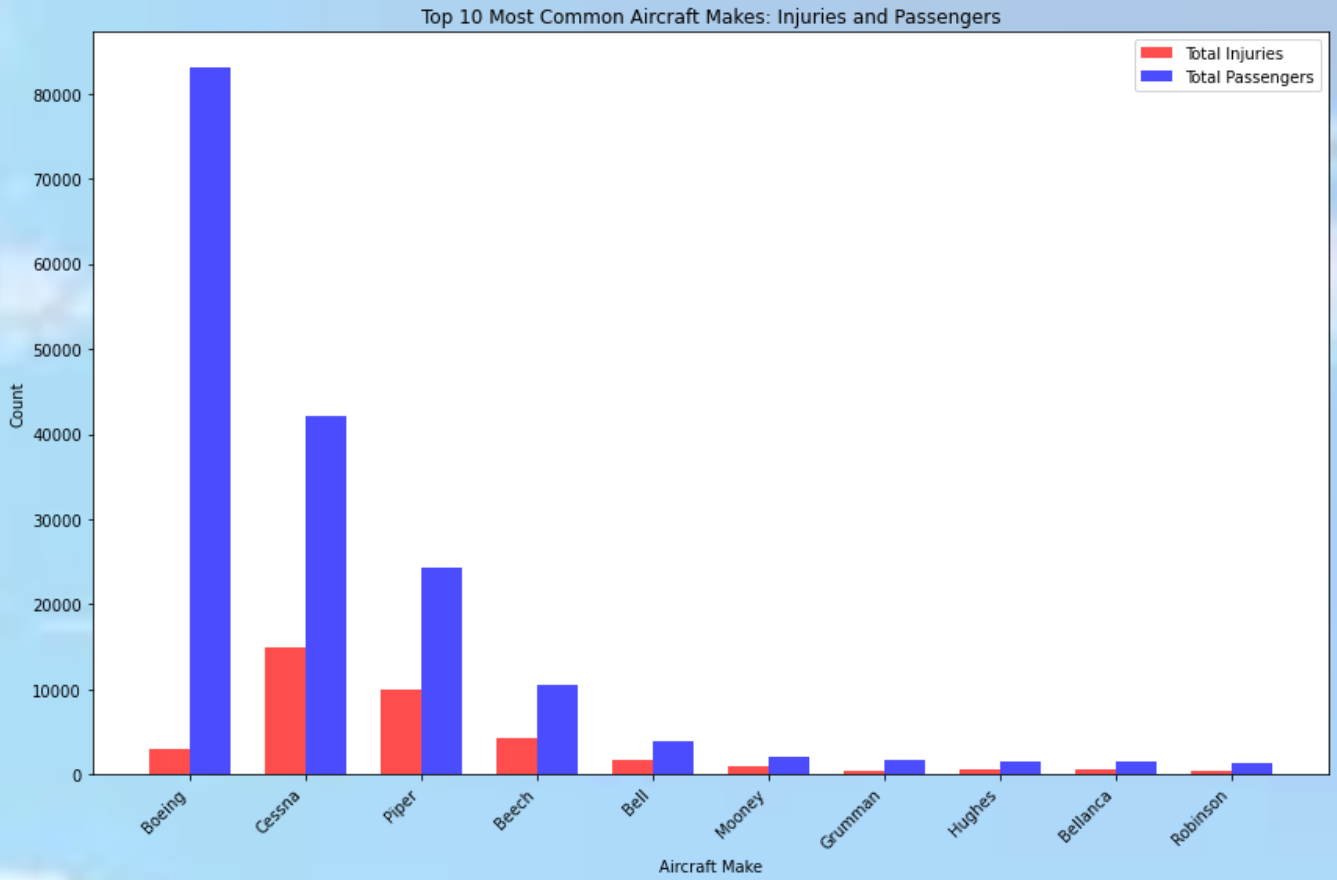
Grouping by Aircraft Makes: Identified the top 10 manufacturers (e.g., Boeing, Cessna, Grumman etc.) and compared their safety records.

Analysis by Engine Type: Examined the relationship between engine type (e.g., Turbojet, Reciprocating etc.) and accident outcomes.



The time series graph shows that the number of accidents has been on the decline over the years hence air travel is becoming a safer means of transport and good venture for our company

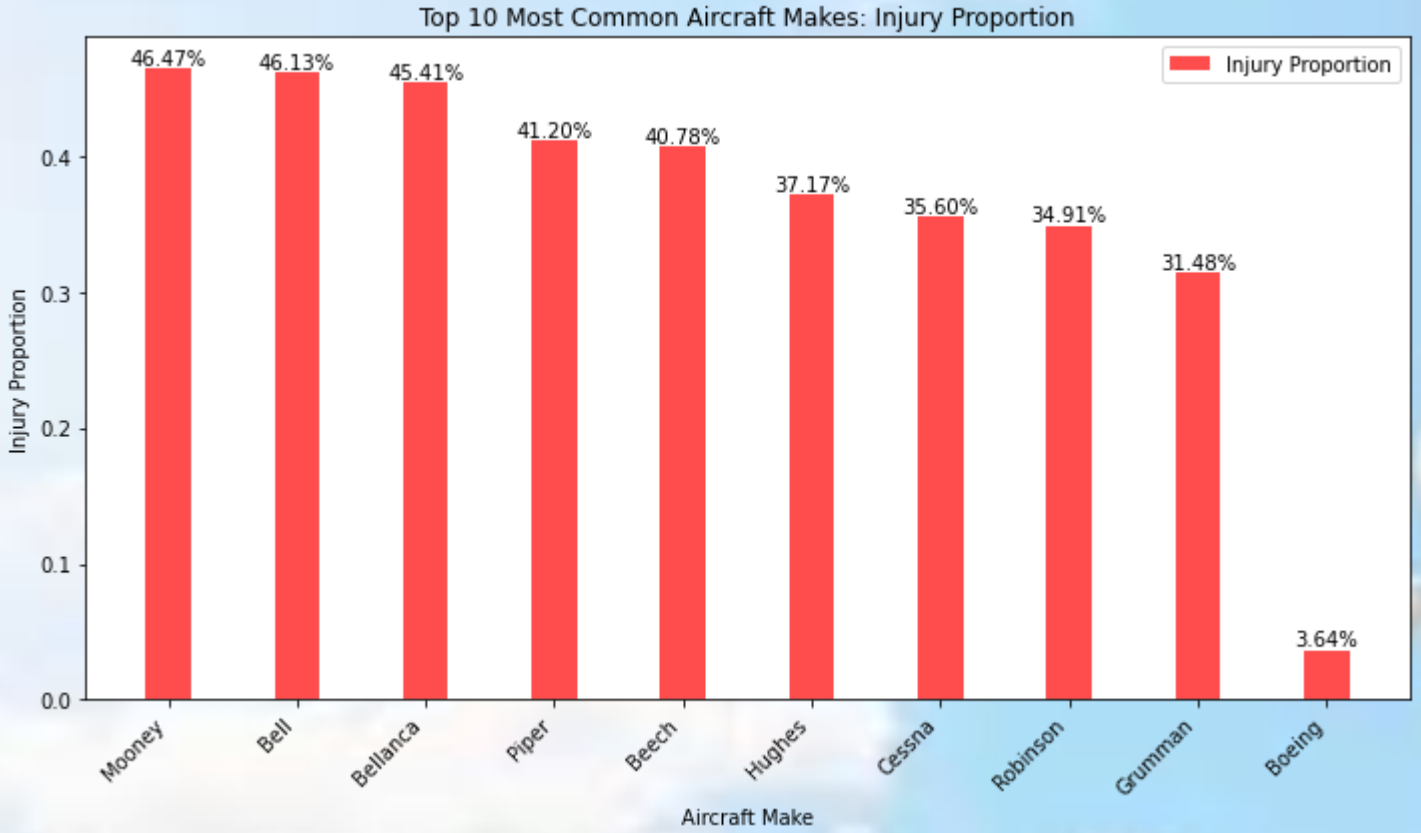




Analysis of injuries of the top ten aircrafts by make shows that that:

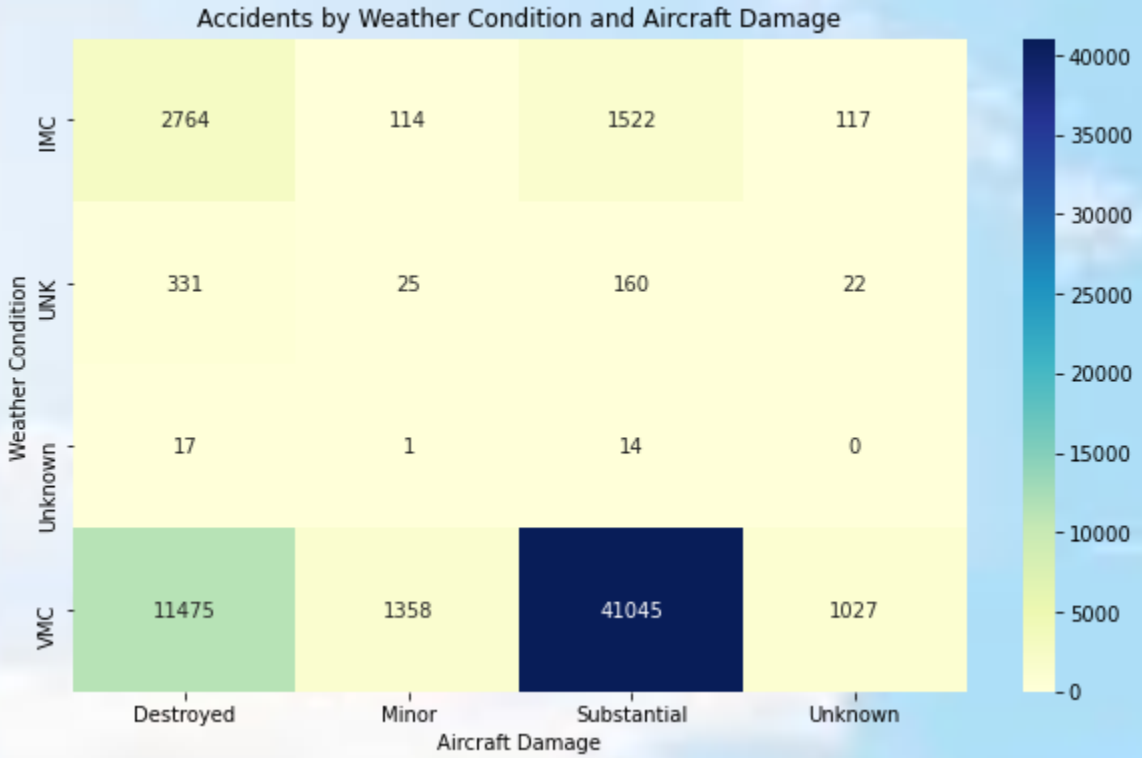
- Although Boeing has the highest number of injuries it also has the highest number of minor and uninjured per total passengers this is due to its large carrying capacity.

- Aircrafts like Piper, Beech and Mooney have high fatality rates compared to their low carrying capacity



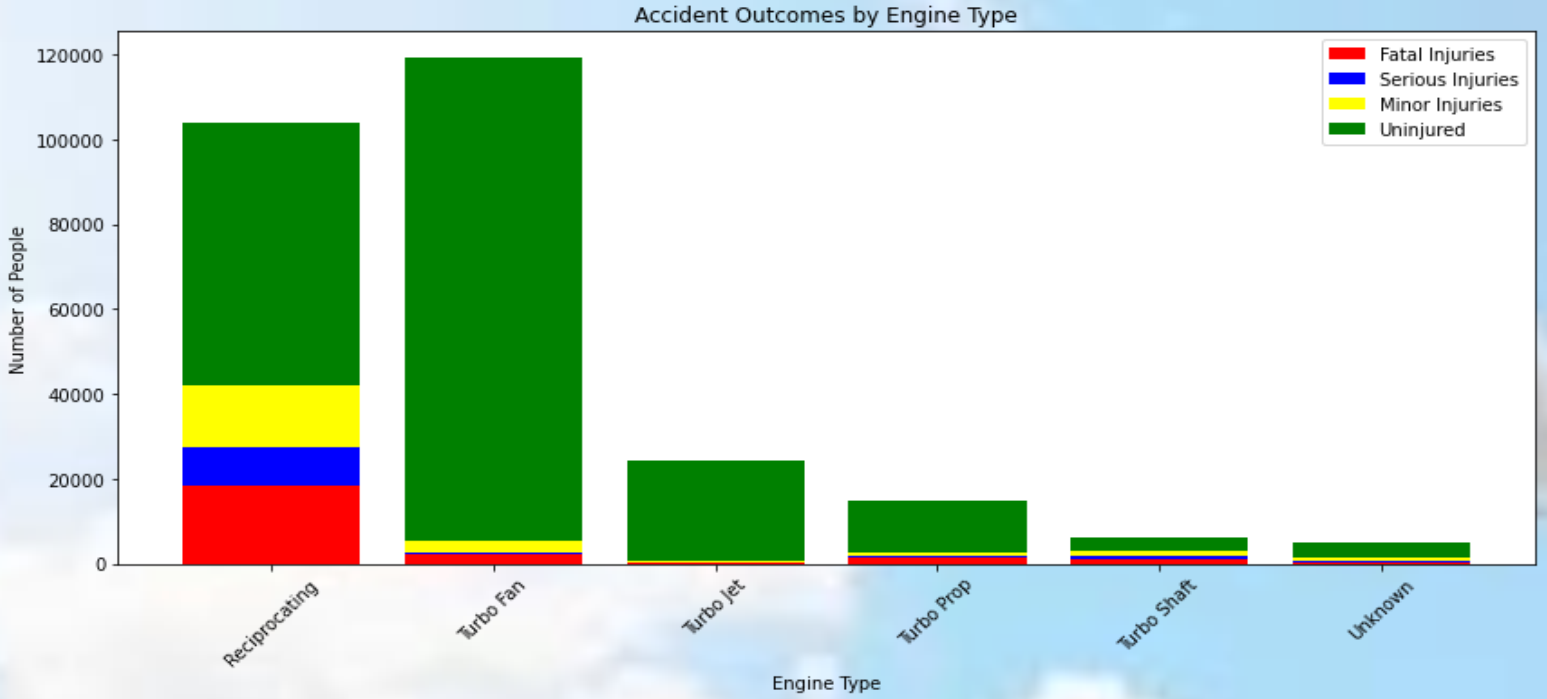
- Plotting the injury proportion of the top 10 makes confirms that Boeing and Grumman aircrafts are good purchases as they have low-risk affiliated with them by number of injuries.

- Mooney and Bell aircrafts should be scrutinized with a lot of caution and may not be a good purchase recommendation due to their high proportionality of accidents and injuries.



- Most accidents occur under clear conditions when the visibility is very clear (VMC).
- This shows human error and mechanical issues This necessitates adequate and continuous training of the pilots.

ANALYSIS BY ENGINE TYPE



- Reciprocating engines have the highest number of fatal injuries, followed by a significant number of serious and minor injuries.
- Turbo fan engines are associated with the highest overall number of incidents, but most of these result in no injuries (uninjured category).
- Turbo jet, turbo prop, and turbo shaft engines show a mix of outcomes, but generally fewer injuries compared to reciprocating and turbo fan engines but significantly fewer Passengers.

Limitations

Missing Values - Significant portions of the data are missing in many columns. Missing data about the location or aircraft category can lead to biased or incomplete analysis.

Lack of Contextual Information - The dataset doesn't include detailed information on factors such as maintenance records or pilot experience.

Incomplete Data on Injuries - The columns Total.Fatal.Injuries, Total.Serious.Injuries, Total.Minor.Injuries, and Total.Uninjured have many missing values.

Geographical Bias - The data primarily covers accidents in the US and international waters limiting the applicability of the findings to international operations.

Recommendations

- Aircraft Acquisition: Cessna 172P model: This model demonstrates the best safety record among the analyzed aircraft. The 172P specifically has the lowest injury count (246) despite a moderate number of passengers (829). It's an excellent choice for smaller Flights

- Turbofan-powered aircraft: aircraft with turbofan engines showed a majority of uninjured outcomes in accidents. For medium to large-scale operations, modern turbofan-powered aircraft from manufacturers like Boeing are good options.

- Multi-engine aircraft: The data shows a positive correlation between number of engines and passenger capacity. For commercial operations, consider multi-engine aircraft.

- Avoid - The data indicates that PA-28-140 and PA-28-180 models have significantly higher injury proportions.

- Pilot training : Invest in comprehensive pilot training programs, especially for operations in Visual Meteorological Conditions (VMC) where most accidents occur.

FURTHER STEPS

Further Research: Additional investigation into maintenance and operational costs for selected models to ensure cost-effectiveness.

Implement a system for continuous monitoring and analysis of safety data.

THANK YOU

Open to Any questions

Email: mainaedin039@gmail.com

LinkedIn: https://www.linkedin.com/in/edwin-maina-08e05m/