Aviation Business Analysis

BY Edwin Maina

Overview

As part of our strategic initiative to diversify our company's portfolio, we are venturing into the aviation industry. This expansion involves purchasing and operating airplanes for commercial enterprises. Given the high stakes associated with aviation, understanding the potential risks and ensuring the highest standards of safety is paramount for the success of this new business endeavor.

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. Delivering key recommendations for strategic aircraft purchases.This project seeks to minimize operational risks and enhance decision-making when purchasing and operating aircraft.

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. We need to identify which aircraft present the lowest risk at reasonable cost and successful entry to the market to make good investment decisions.

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. The goal is to provide insights that will guide the decision-making process for the acquisition of aircraft, ensuring that we select models with proven safety records.

Key Stakeholders

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

Success Criteria

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 a low risk aircraft with good return on investment

Data Understanding

Data Source:

•National Transportation Safety Board (NTSB) dataset covering aviation incidents from 1962 to 2023.

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
I handled Missing values were especially in critical columns such as Make, Model, and Total Injuries.

In this analysis, I aim to focus on the most relevant data to enable making of informed recommendations. As such, I decided to perform data reduction by dropping certain columns from this dataset. This step will simplify our data and make it easier to handle, without losing the information that is crucial to our analysis
filtered Data from 1982 onwards to focus on recent trends and patterns.
Created variables such as injury proportions for deeper insights.

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 extreme values or outliers compared to the mean.

Exploratory Data Analysis (EDA)
Descriptive Statistics:

Modeling and Analysis

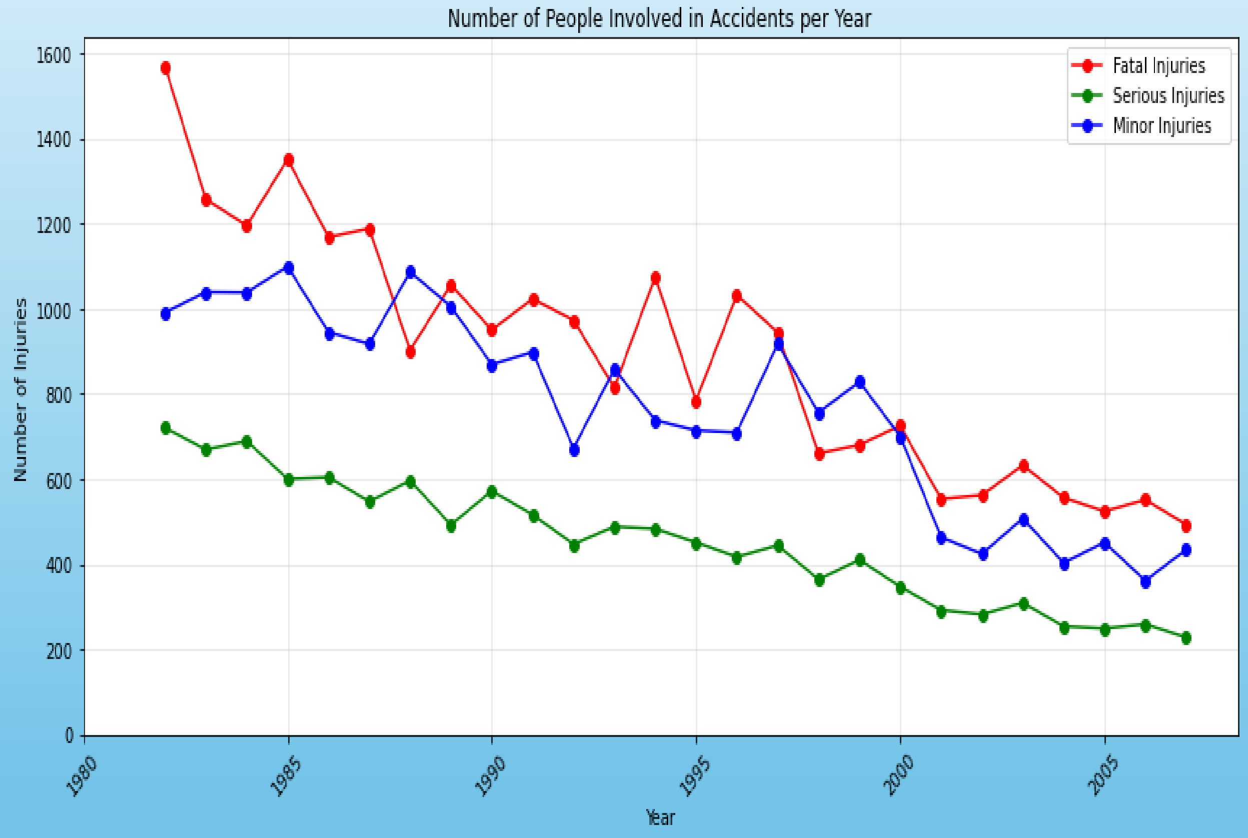
Feature Trend Analysis: I analyzed Injuries and fatalities over time to identify patterns.

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

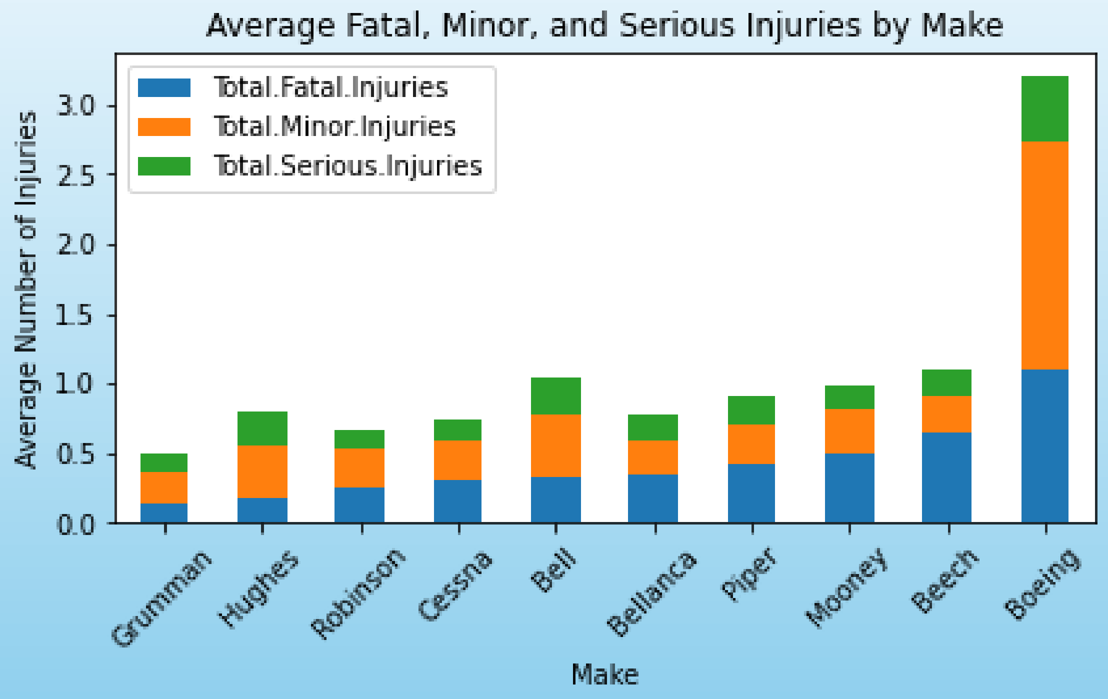
Grouping by Aircraft Makes: Identified the top 10 manufacturers (e.g., Boeing, Airbus) and compared their safety records.

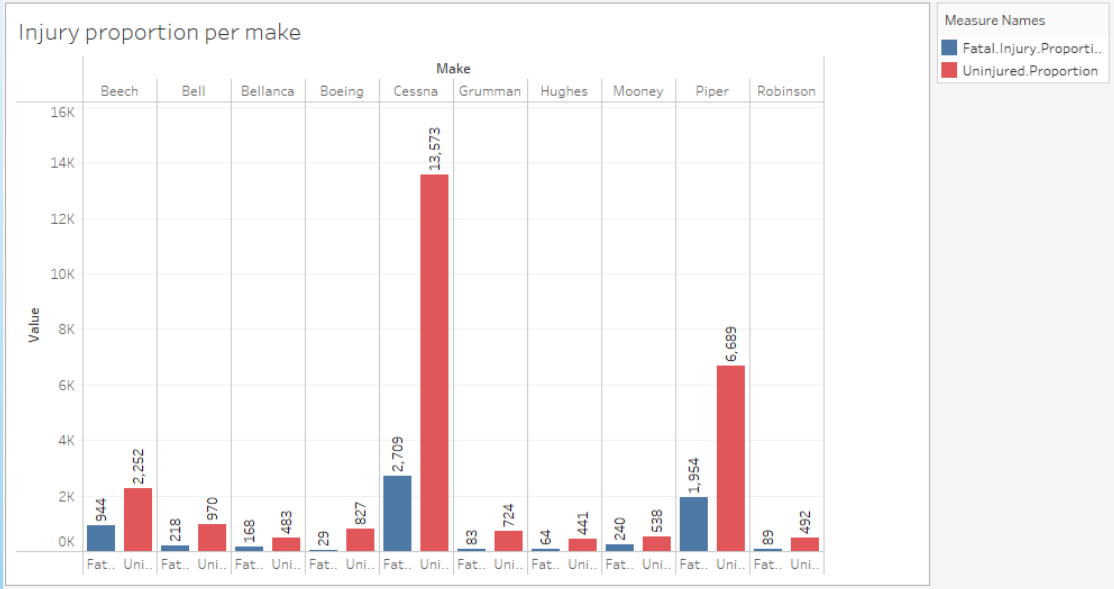
Analysis by Engine Type: Examined the relationship between engine type (e.g., Turbojet, Piston) and accident outcomes.

Injury Proportion Analysis: Created new features to compare the ratio of fatalities and uninjured passengers across different models and years.

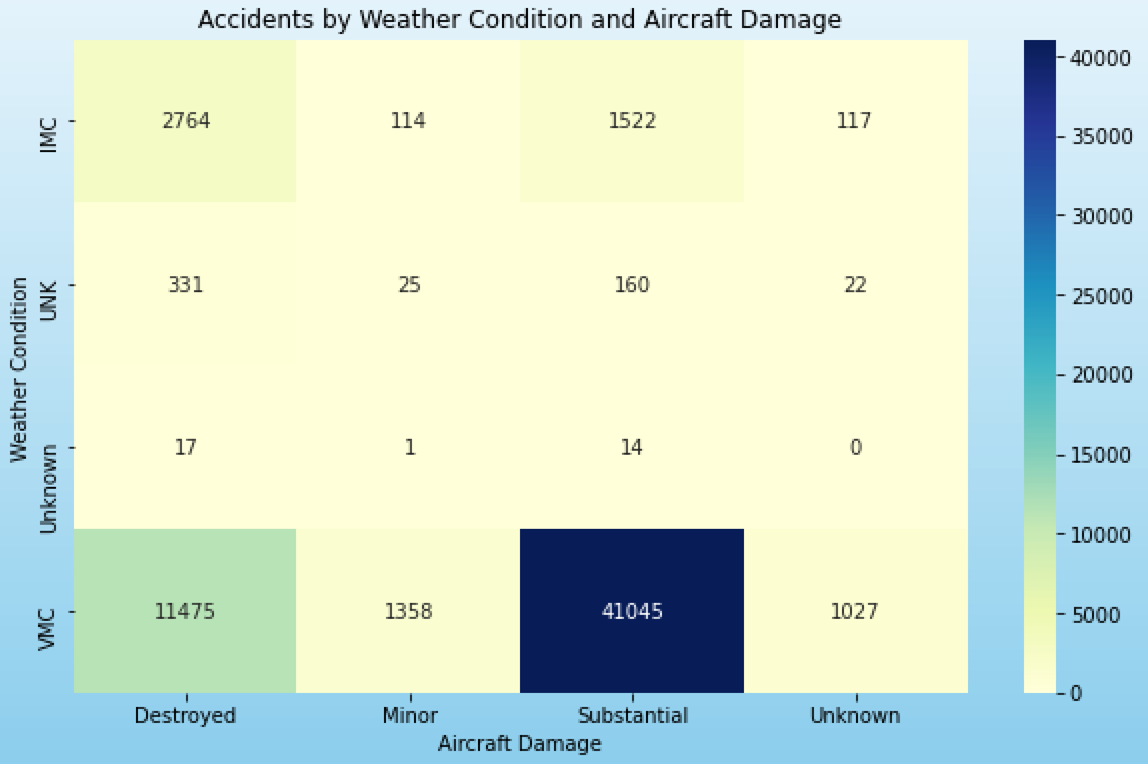


As seen in the time series graph the number of accidents has been on the decline over the years hence air travel is becoming a safer means of transport and good venture for our company





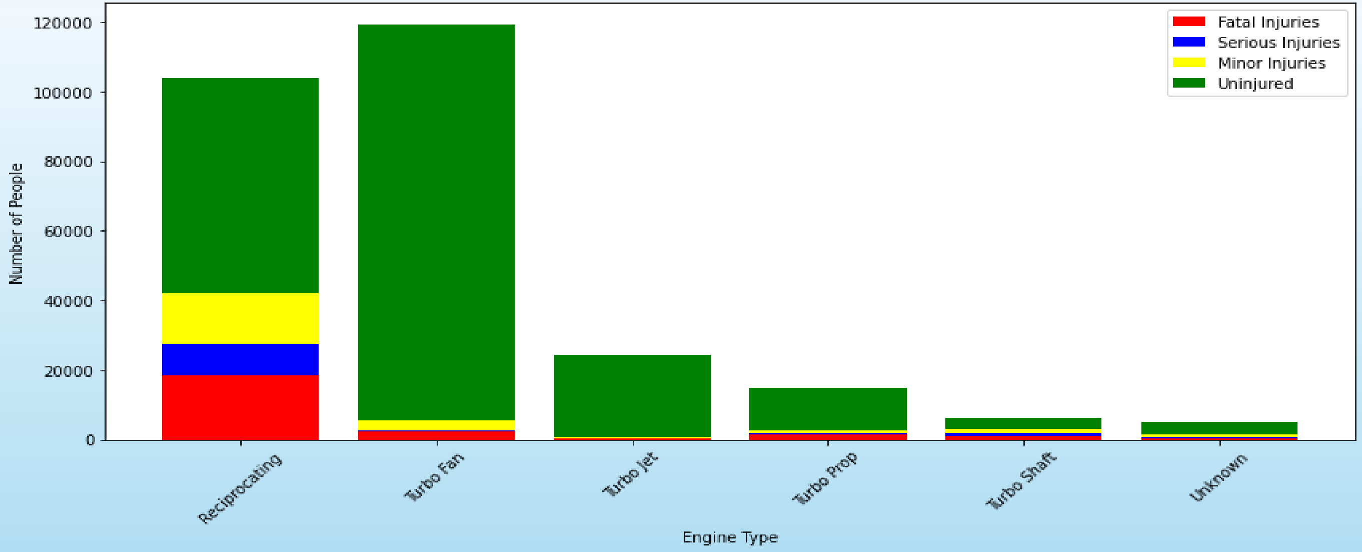
analysis of injuries by make shows that that although Boeing has the highest number of injuries ii also has the highest number of uninjured per total passengers this is due to its large carrying capacity while aircrafts like Cessna have high fatality rates with low carrying capacity hence Boeing and Grumman aircraft are a better choice for the company than Cessna and piper



From the Weather Condition and aircraft damage heatmap, it is evident that:

Pilots cause accidents under clear conditions when the visibility is very clear (VMC).

- This necessitates adequate and continuous training of the pilots.



- 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 progressively fewer incidents, with a mix of outcomes but generally fewer injuries compared to reciprocating and turbo fan engines.

Limitations

Missing Values - Significant portions of the data are missing in many columns. Missing data can lead to biased or incomplete analysis. Missing information about the location or aircraft category might not reveal important patterns related to accident frequency or severity.

Lack of Contextual Information - The dataset doesn't include detailed information on operational factors such as maintenance records, pilot experience, or specific environmental conditions at the time of the accident. These factors are crucial in understanding the full context of accidents and making informed recommendations. Without this data, this analysis may have overlooked critical contributors to aviation safety.

Incomplete Data on Injuries - The columns Total.Fatal.Injuries, Total.Serious.Injuries, Total.Minor.Injuries, and Total.Uninjured have many missing values. The risk assessment might be skewed if severe accidents are underreported or missing.

Geographical Bias - The data primarily covers accidents in the US and international waters. This geographical focus might not represent global aviation safety trends, limiting the applicability of the findings to international operations and thereby potentially affecting the recommendations I've made.

Recommendations

Aircraft Acquisition: Focus on aircraft models from manufacturers with a proven safety record, such as Boeing and Airbus.Engine Type
Preference: Consider models with turbojet or turboprop engines, as they have lower fatality rates.
Operational Considerations: Improve safety protocols, especially during high-risk phases like takeoff and landing.

The analysis also highlighted specific aircraft makes and models, such as Cessna and Piper, which have higher accident frequencies, possibly due to mechanical issues.

FURTHER STEPS

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

THANK YOU

Open to Any questions

Email: mainaedin039@gmail.com

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