# Aviation Accidents Analysis

By Catherine Maina


"Great flights and great investments both start with one thing: a solid runway and a smart plan."

# Business Overview

- The company is expanding into aviation to diversify portfolio
- It's exploring aircraft for private and commercial use
- There is limited knowledge on aircraft safety

**Overview**

•The company is expanding into the **aviation industry** to diversify its portfolio.
•Plans involve acquiring aircraft for **commercial and private operations**.
•Internal knowledge on **aircraft safety risks is limited**.

## Project Objectives

- Project analyzes accident data to identify low-risk aircraft to purchase
- Support data-driven decisions for new aviation division

**Project Objectives**

•This project analyzes aviation accident data to:

- Identify **low-risk aircraft models**.
- Support **data-driven purchasing decisions**.

•Insights will guide the **new aviation division** in safe and strategic aircraft investments.

# Tools and Technologies used

- **Python** (Pandas, Matplotlib, Seaborn) – Data cleaning, analysis, and visualization
- **Jupyter Notebook / VS Code** – Coding environment
- **Tableau** – Interactive dashboards and charts
- **Git & GitHub** – Version control and collaboration

# Data Overview

- The data used in this project is from the **National Transportation Safety Board (NTSB)** who maintain a comprehensive database detailing civil aviation accidents across various territories.

# Data Overview

The key features used in this project are;

- **Coverage Period:** 1962 to 2023
- **Data Scope:** Includes over 135,000 accidents and incidents.
- **Data Fields:** Contains information such as aircraft make and model, purpose of flight, number of injuries among others.

Key Features:
**Coverage Period:** Accidents from 1962 to 2023.
**Data Scope:** Includes over 135,000 accidents and incidents.
**Data Fields:** Encompasses information on aircraft make, model, weather conditions, purpose of flight, number of injuries and other contributing factors.

# Data Cleaning

Data Cleaning Summary;

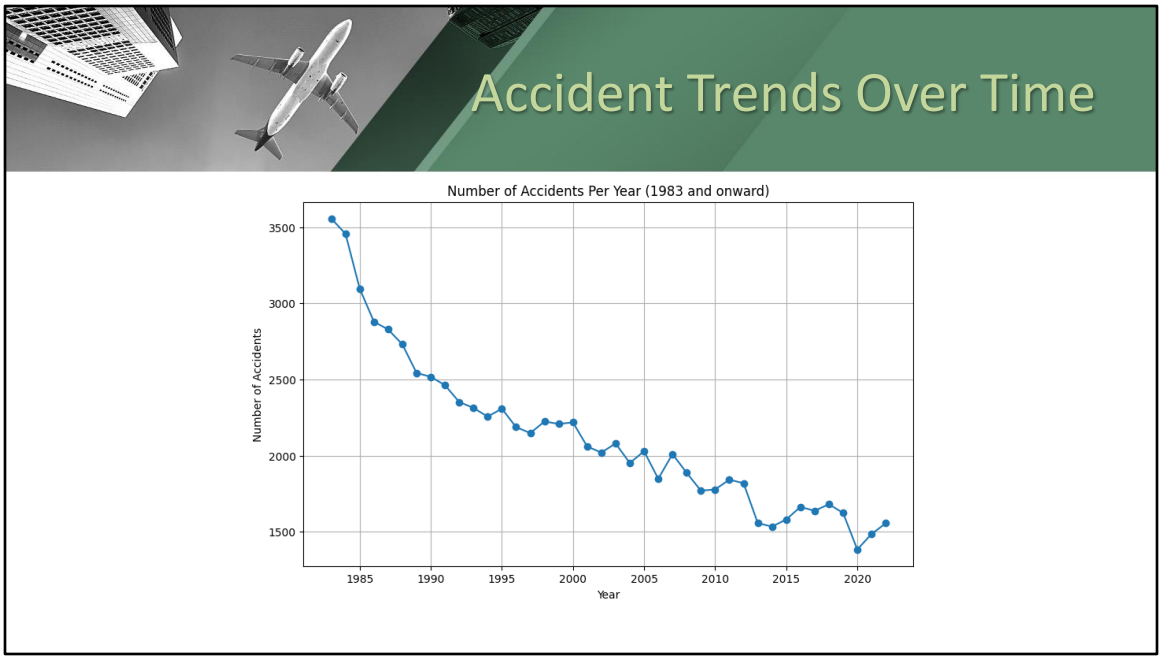
- Dropped columns with excessive missing data
- Selected only relevant columns for analysis
- Filled missing numeric values with 0, categorical with 'Unknown’
- Added a new column for total injuries

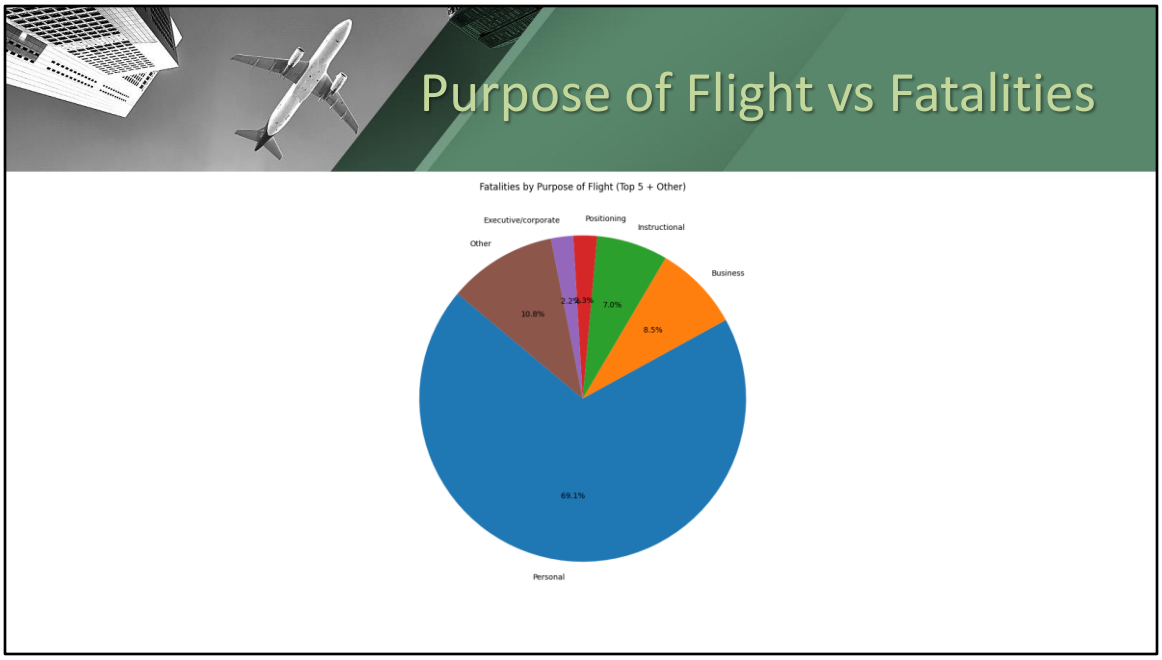
```
The following data cleaning steps were taken:
* Columns with too many missing values are dropped.
* Only the relevant columns were chosen for a cleaner analysis
• Numerical relevant columns with missing values were replaced
  with 0
• Relevant columns with categorical data with missing values
  were replaced with 'Unknown’
• A new column that combined the total injuries was added
```

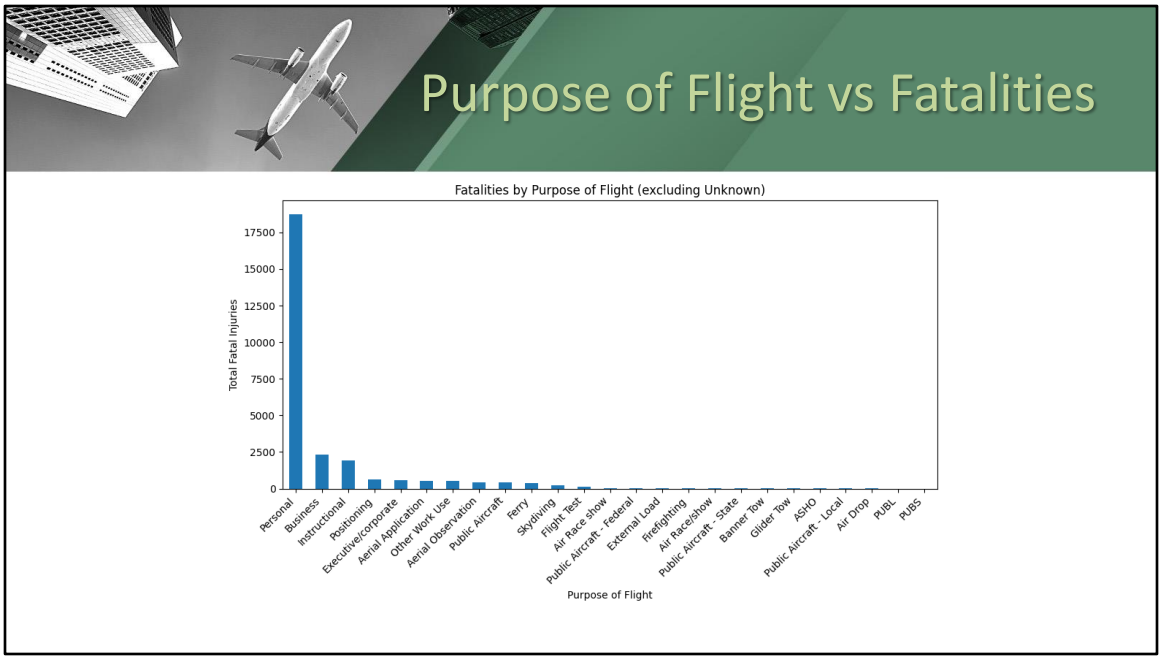
# Exploratory Data Analysis

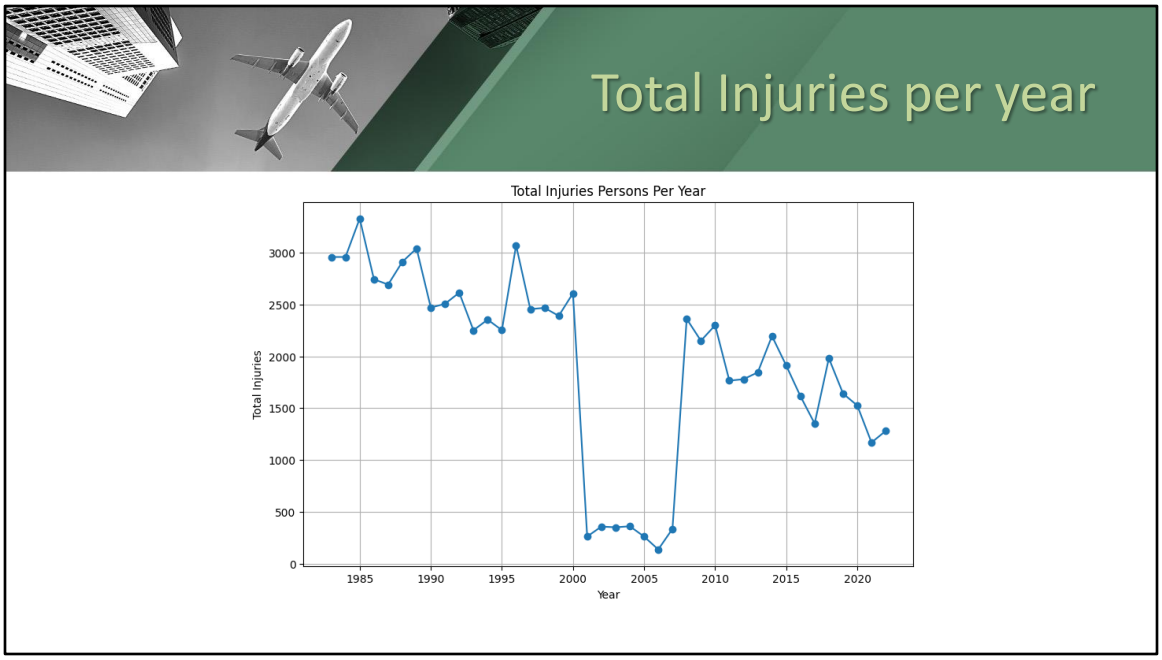
**Objective:**

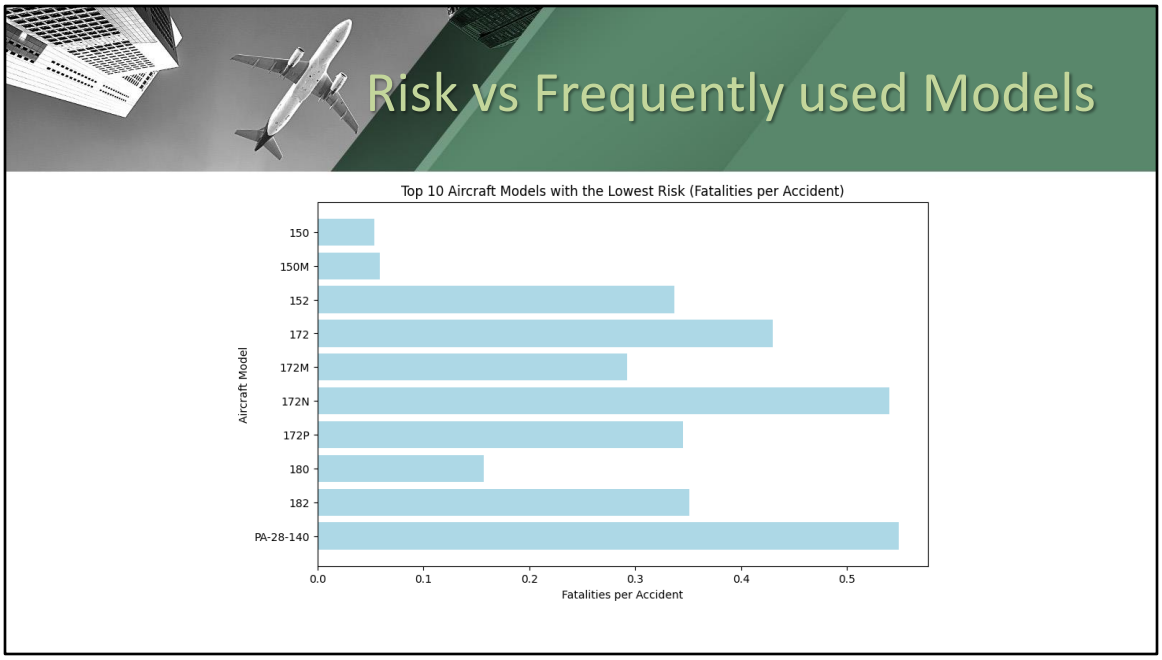
To understand key patterns and distributions in the aviation accident dataset.


Accident Trends Over Time
Number of Accidents Per Year (1983 and onward)
Number of Accidents
3500
3000
2500
2000
1500
1985
1990
1995
2000
2005
2010
2015
2020
Year


Purpose of Flight vs Fatalities
Fatalities by Purpose of Flight (Top 5 + Other)
Executive/corporate
Positioning
Instructional
Other
Business
10.8%
2.2%
4.3%
7.0%
8.5%
69.1%
Personal


Purpose of Flight vs Fatalities
Fatalities by Purpose of Flight (excluding Unknown)
Total Fatal Injuries
0
2500
5000
7500
10000
12500
15000
17500
Personal
Business
Instructional
Positioning
Executive/corporate
Aerial Application
Other Work Use
Aerial Observation
Public Aircraft
Ferry
Skydiving
Flight Test
Air Race show
Public Aircraft - Federal
External Load
Firefighting
Air Race/show
Public Aircraft - State
Banner Tow
Glider Tow
ASHO
Public Aircraft - Local
Air Drop
PUBL
PUBS
Purpose of Flight


Total Injuries per year
Total Injuries Persons Per Year
Total Injuries
Year
0
500
1000
1500
2000
2500
3000
1985
1990
1995
2000
2005
2010
2015
2020




- I sorted my data to be based on usage. Based on how often the aircraft were used as well as the number of fatalities reported for those models, the above chart shows the models with the lowest risk. This is shown the Risk vs Frequently used Models slide.


- **Accident trends** – Notable decline over the years
- **Aircraft models with fewer accidents**
- **Purpose of flight** – Commercial has fewer accidents

**Key Insights;**
**Accident trends:**
Over the years, there has been a notable decline in aviation accidents with occasional peaks and declines.
**Purpose of flight:**
Flights categorized under **Personal** or **Instructional** purposes are associated with a **higher number of injury incidents** compared to Commercial or Business flights.
**Aircraft Models:**
I sorted my data to be based on usage. Based on how often the aircraft were used as well as the number of fatalities reported for those models, the above chart shows the models with the lowest risk. This is shown the Risk vs Frequently used Models slide.


- Consider Aircraft with a low Accident rate
- Align Investment with Use-Case Insights (Purpose of flight)
- Prioritize Aircraft Models with No or Minimal Fatalities
- Consider Aircraft that have shown to have the least damage

**1. Consider Aircraft with a low Accident rate**
Aircraft such as the Mooney M-22 and Piper PA-28 as seen in Figure 7, consistently appear in accident data with lower fatality rates per incident, suggesting they are statistically safer and reliable for general aviation or training purposes.

**2. Align Investment with Use-Case Insights**
Use the Purpose of flight data to guide investment. Flights categorized under "Personal" or "Instructional" showed higher fatality rates compared to commercial operations as shown in Figure 6.Therefore, aircraft with safer outcomes in commercial or charter operations may be preferable if the business strategy is oriented toward such use cases. Avoid High-Risk Models Common in Personal and Instructional Flights.

**3. Prioritize Aircraft Models with No or Minimal Fatalities**
Aircraft with consistently low or zero fatality rates indicate higher safety and reliability.Aircraft Models such as Cessna 150, Cessna 152M and Cessna 172 showed minimum fatality rates and can be considered.

**4. Consider Aircraft that have shown to have the least damage**

This shows that the aircraft is generally a safer investment as lower damage can also signify lower injury rates.

## Next Steps

- Add a total number of flights column for each aircraft.
- Analyze manufacturer reliability further
- Conduct a Cost-Benefit analysis
- Incorporate Updated Data
- Perform Risk simulations

In the future,

1. **Add a total number of flights column.**

Having a column that shows the total number of flights taken for each model of Aircraft in order to get a clearer analysis of the safety of each model.

**2. Further Analyze Manufacturer Reliability**

Combine Make and Model to assess safety records by manufacturer, offering broader investment insights.

**3. Conduct Cost-Benefit Analysis**

Compare operational costs, maintenance history, and acquisition costs for the recommended models.

**4. Incorporate Real-Time or Updated Data**

Use newer accident reports or real-time flight safety databases for a more current risk evaluation.

**5. Run Risk Simulations**

Apply statistical or machine learning models to simulate accident likelihood under various conditions.

Thank you!