

AVIATION ACCIDENT DATA INSIGHTS

Soudie Okwaro

Data Science Student

Flatiron College(USA), through Moringa School, Kenya

Part-time cohort 8

2024

edgarsoudie@gmail.com | +254710648100

BUSINESS PROBLEM

A company is expanding in to new industries to diversify its portfolio. Specifically, they are interested in purchasing and operating airplanes for commercial and private enterprises, but do not know anything about the potential risks of aircraft.

A Data Scientist is charged with determining which aircraft are the lowest risk for the company to start this new business endeavor.

S/he must then translate their findings into actionable insights that the head of the new aviation division can use to help decide which aircraft to purchase.

AIM: This project seeks to generate insights with a high level of specificity to inform aircraft purchase with the level of precision to a `Make` and a `Model` to consider.



AVIATION INDUSTRY OVERVIEW

- Aviation accidents are rare but catastrophic, causing loss of life, financial liability, and reputational damage.
- Common causes include human error, mechanical failures, poor weather conditions, and cybersecurity threats.
- Human error is the leading cause of aviation accidents, accounting for about 70% of incidents.
- Aging aircraft fleets and lapses in maintenance pose challenges.
- Per literature, Certain aircraft models have higher accident rates, such as the McDonnell Douglas DC-10 and Boeing 737 Max.



UNDERSTANDING AVIATION DATA

- Data analysis helps companies make informed decisions when purchasing aircraft, prioritizing those with proven safety records.
- The analysis of the aviation data was guided by CRISP-DM and Python PEP-8
- A Dataset 'AviationData.csv' was downloaded from kaggle.com
(<https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses?resource=download>)
- VS Code was used to load the required python libraries:
 - Pandas as pd
 - Numpy as np
 - Seaborn as sns
 - Matplotlib.pyplot as plt



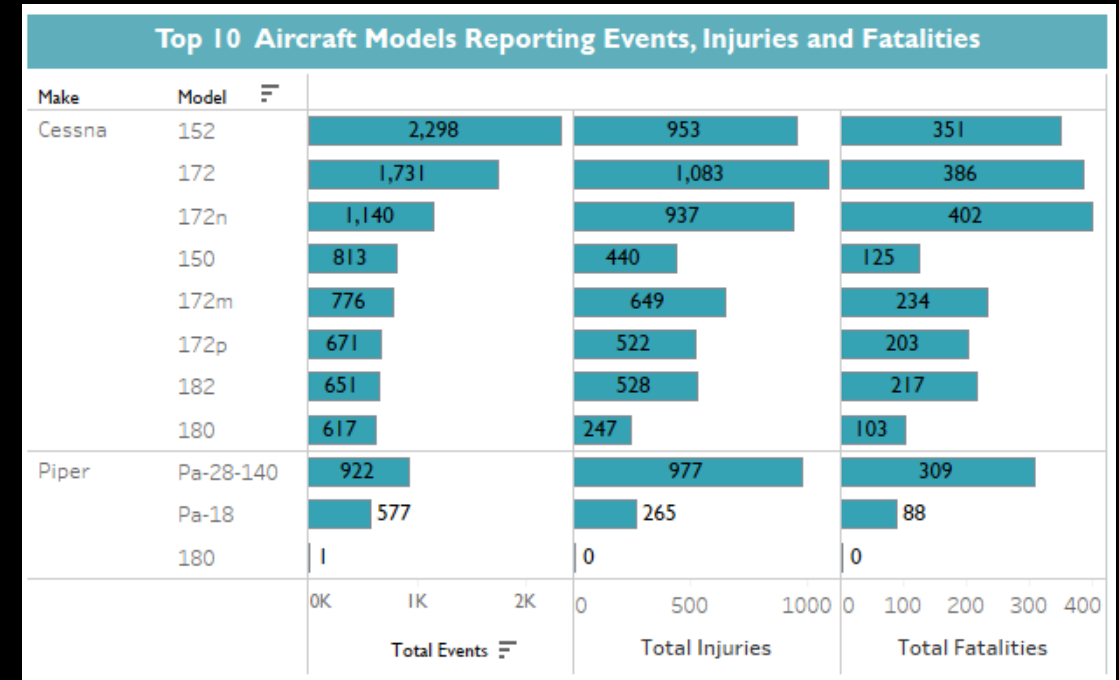
DATA PREPARATION & MODELLING

- Data cleaning was done, involving
 - Formatting headers
 - String formatting
 - Data type casting
 - Addressing missing values
 - String manipulation
 - Addressing Duplicate entries
 - Feature creation
 - Exploratory data analysis
- EDA:
 - Thinking through EDA questions
 - Thinking through coding logic and library function, modules and methods to use
 - Making observation on the code outputs
 - Troubleshooting code errors
 - Widely learning concepts through critical thinking
- Descriptive and deep-dive logical analytics revealed interesting data revelation and patterns
- Cleaned file was exported to the default folder as '.csv'
- The Cleaned CSV files



INSIGHTS

- Cessna models, particularly the Cessna 152 and Cessna 172, have the highest number of accidents and incidents.
- Aircraft models from Zwicker Murray R (Glastar) and Yuneec report the fewest accidents and injuries, with Yuneec being safer for commercial flights and Zwicker Murray R for private flights.



Bottom 5 Commercial Aircraft with least accidents and incidents:

| | Make | Purpose_Of_Flight | Accident_Count |
|------|--------------------------|-------------------|----------------|
| 9249 | Wheeler acft. co. | Business | 1 |
| 9251 | Wheeler technology, inc. | Business | 1 |
| 9442 | Wsk | Business | 1 |
| 9450 | Wsk pzl mielec | Business | 1 |
| 9511 | Yuneec | Business | 1 |

Bottom 5 Private Aircraft with least accidents and incidents:

| | | | |
|------|------------------|----------|---|
| ... | | | |
| 9561 | Zuber thomas p | Personal | 1 |
| 9562 | Zukowski | Personal | 1 |
| 9563 | Zwart | Personal | 1 |
| 9564 | Zwicker murray r | Personal | 1 |

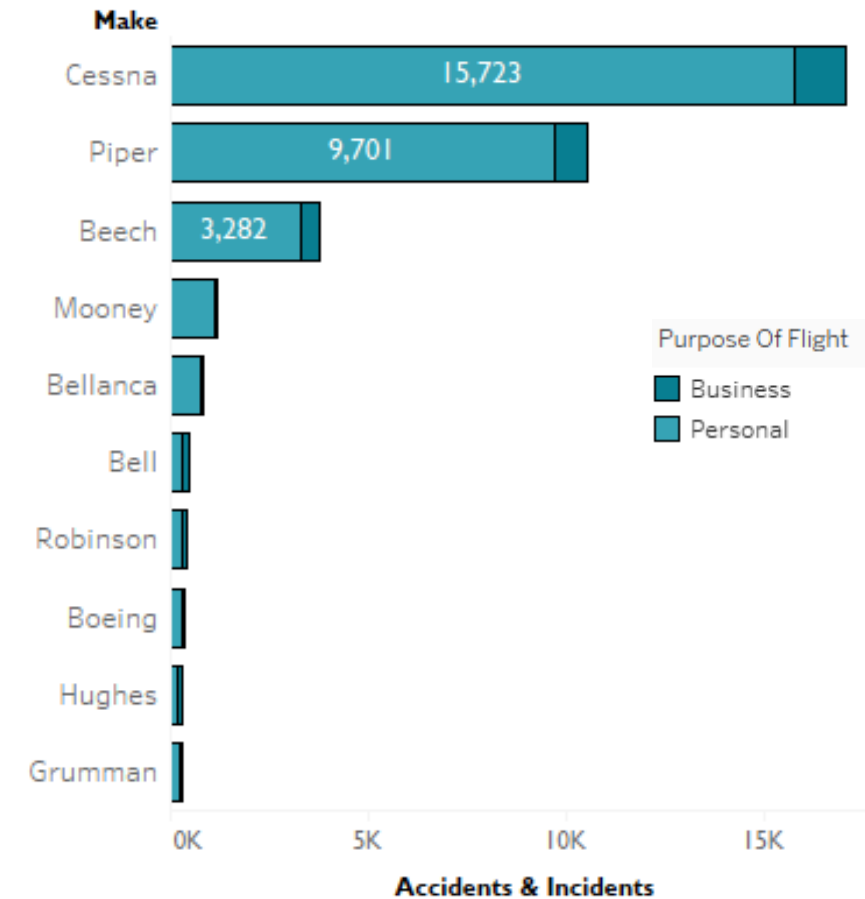
INSIGHTS

- Private aircrafts have a higher incidence of accidents and incidents compared to commercial flights, with the Cessna 172 and Piper models (e.g., Pa-28-180, Pa-28-181) being particularly accident-prone.
- Commercial flights also see high numbers of accidents, with the Beech A36 model leading in accidents and injuries.

Top 5 Commercial Aircraft with most accidents and incidents:

| | Make | Model | Purpose_Of_Flight | Accident_Count |
|-------|--------|-------|-------------------|----------------|
| 2806 | Beech | A36 | Business | 60 |
| 8322 | Cessna | T210n | Business | 52 |
| 18114 | Piper | Pa-18 | Business | 52 |
| 3706 | Bell | 206b | Business | 43 |
| 6870 | Cessna | 182 | Business | 36 |

Top 10 Aircrafts reporting events by Purpose of flight



Top 5 Commercial Aircraft with most injuries:

| | Make | Model | Purpose_Of_Flight | Total_Injuries_Sum |
|-------|--------|-----------|-------------------|--------------------|
| 2806 | Beech | A36 | Business | 79.0 |
| 8322 | Cessna | T210n | Business | 57.0 |
| 18694 | Piper | Pa-32-300 | Business | 53.0 |
| 3706 | Bell | 206b | Business | 51.0 |
| 7622 | Cessna | 414 | Business | 47.0 |

INSIGHTS

- Aircraft with reciprocating engines have a higher number of accidents, particularly during takeoff, cruise, and landing phases, particularly in VMC weather conditions. These accidents often involve Cessna models.
- Turbo engine types, such as those from Swearingen, Taylorcraft, and Waco, report significantly fewer accidents.
- Planes with older, reciprocating engines are more likely to have accidents, especially during takeoff, cruising, and landing. In contrast, planes with more modern turbo engines seem to have fewer issues.
- Private flights often face problems during landing and cruising, while commercial flights experience fewer issues overall during these phases.

Top 5 Landing (Private Purpose) with most accidents and incidents:

| | Make | Model | Engine_Type | Broad_Phase_Of_Flight | Purpose_Of_Flight \ |
|-------|--------|-------|---------------|-----------------------|---------------------|
| 9719 | Cessna | 172 | Reciprocating | Landing | Personal |
| 10154 | Cessna | 172n | Reciprocating | Landing | Personal |
| 9542 | Cessna | 152 | Reciprocating | Landing | Personal |
| 10569 | Cessna | 180 | Reciprocating | Landing | Personal |
| 10095 | Cessna | 172m | Reciprocating | Landing | Personal |

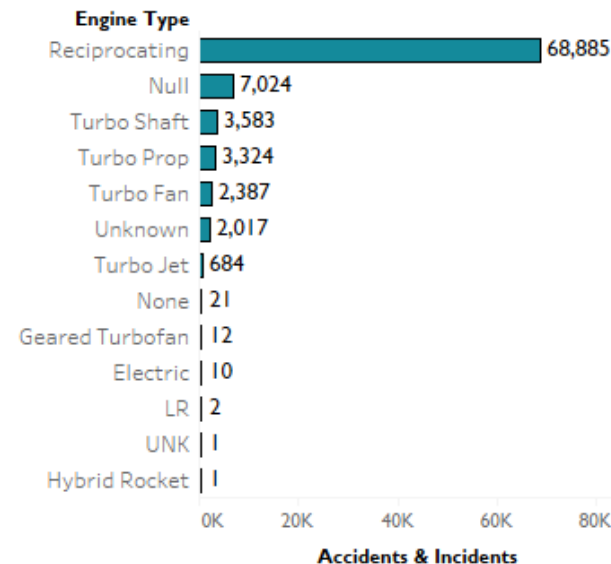
Accident_Count

| | |
|-------|-----|
| 9719 | 239 |
| 10154 | 203 |
| 9542 | 196 |
| 10569 | 171 |
| 10095 | 155 |

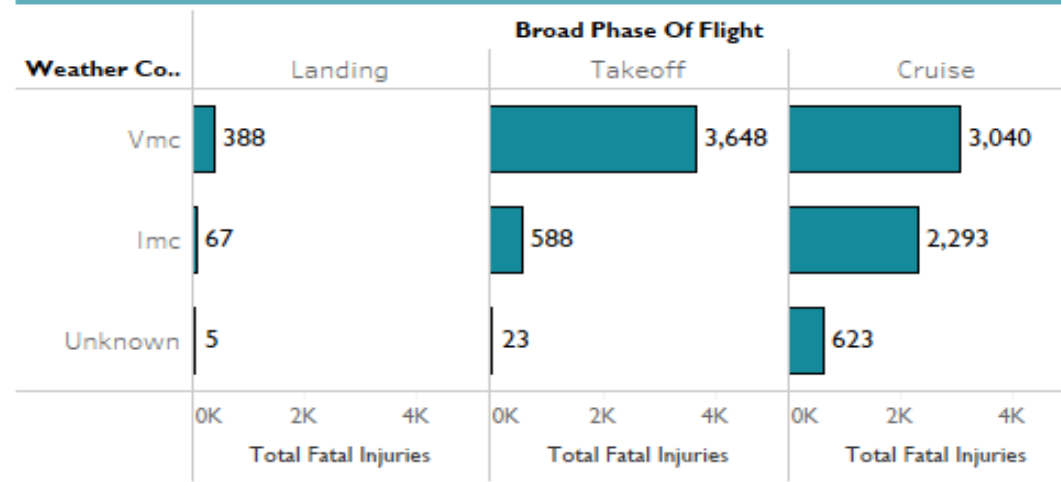
Top 5 Cruise (Private Purpose) with most accidents and incidents:

| | Make | Model | Engine_Type | Broad_Phase_Of_Flight \ |
|-------|--------|-----------|---------------|-------------------------|
| 9532 | Cessna | 152 | Reciprocating | Cruise |
| 27015 | Piper | Pa-28-140 | Reciprocating | Cruise |
| 10140 | Cessna | 172n | Reciprocating | Cruise |
| 27234 | Piper | Pa-28-181 | Reciprocating | Cruise |
| 9706 | Cessna | 172 | Reciprocating | Cruise |

Type of engine for aircraft reporting events

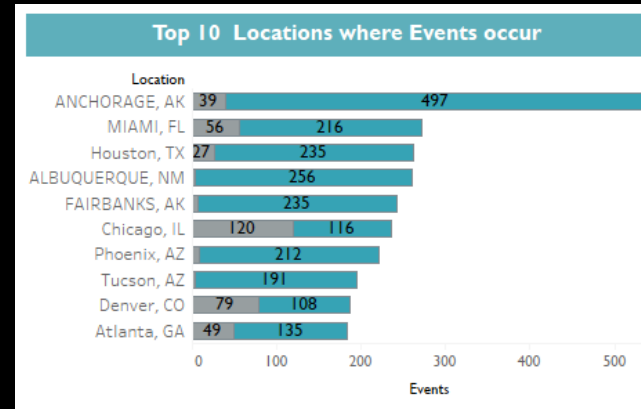


Total Fatalities are reported in Takeoff, Cruise and Landing Flight Phases by Weather Conditions

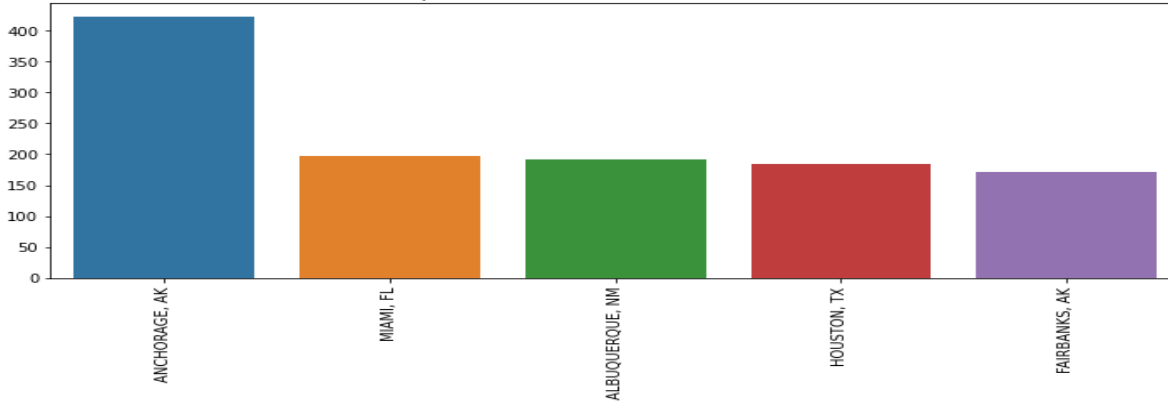


INSIGHTS

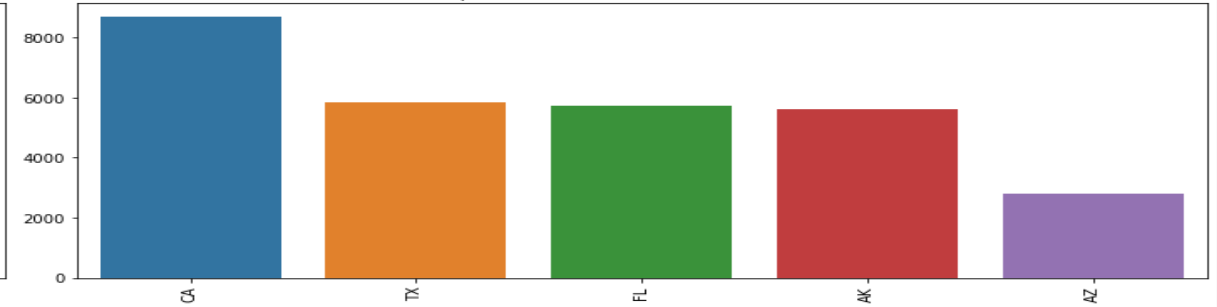
The highest accident rates are reported in the USA, especially in airports in Alaska and mostly in Anchorage city, while locations with the fewest accidents are predominantly outside the USA.



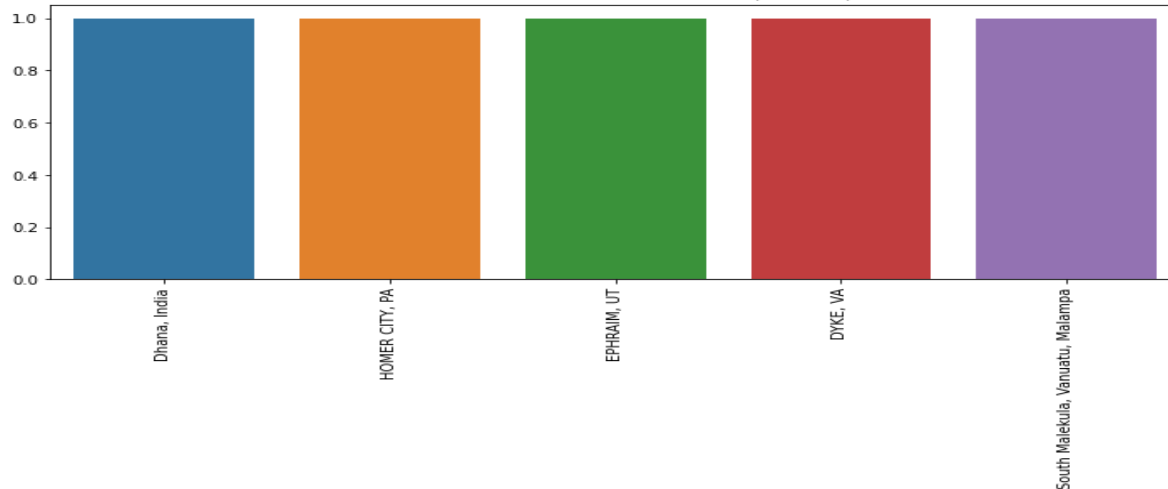
Top 5 Cities with Most Accidents (Location)



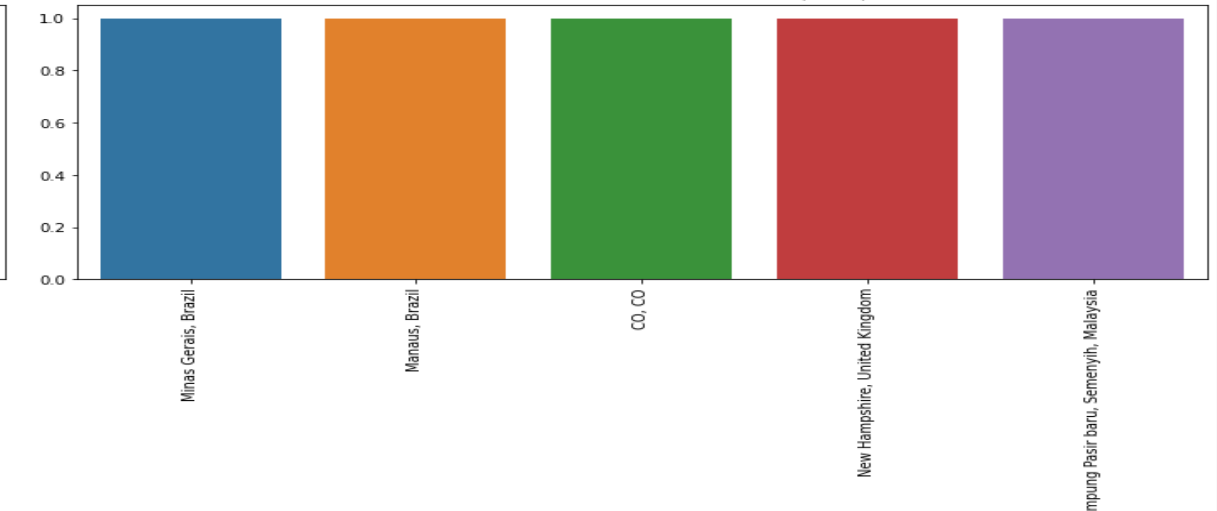
Top 5 States with Most Accidents (State)



Bottom 5 Cities with Least Accidents (Location)



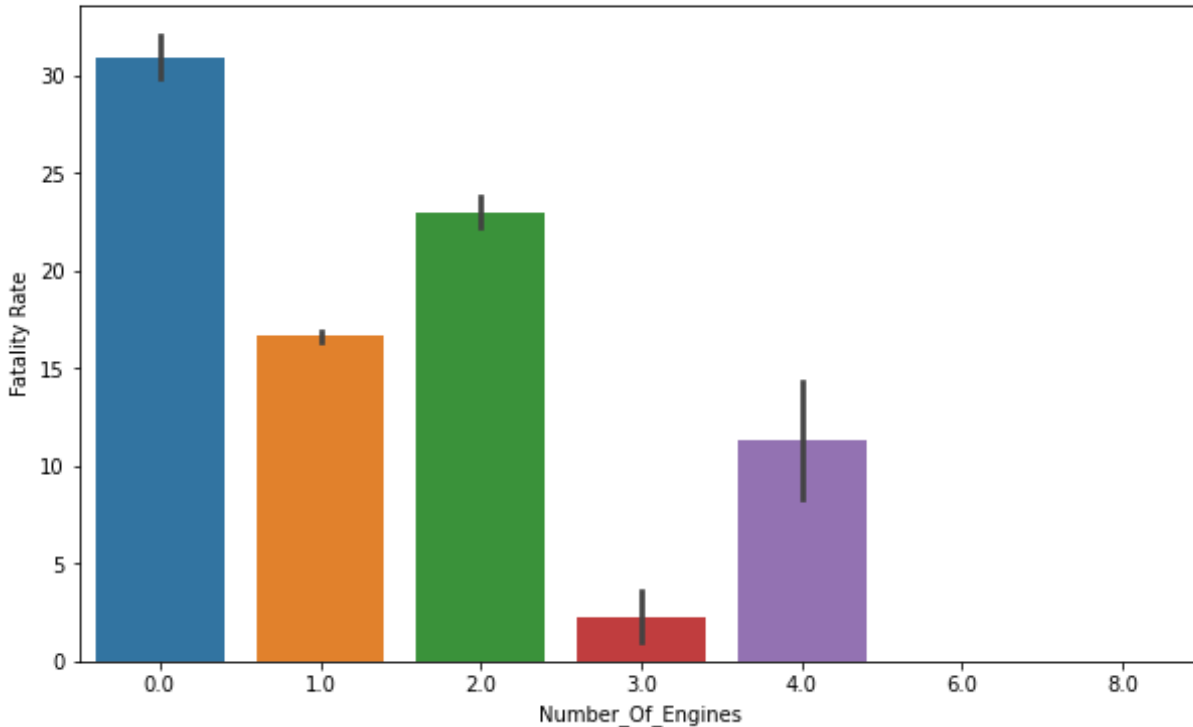
Bottom 5 States with Least Accidents (State)



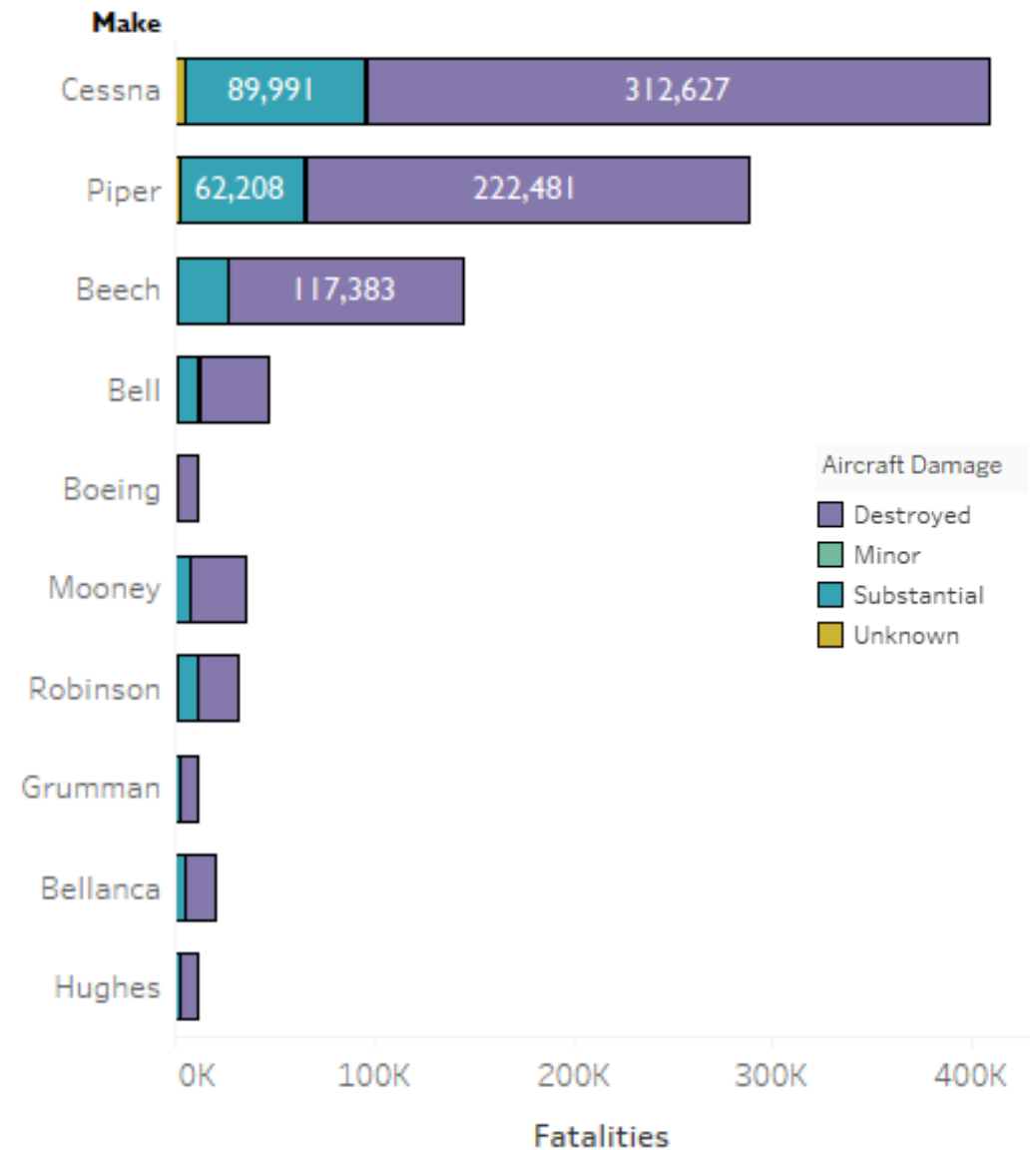
INSIGHTS

- Most accidents are non-fatal, but significant fatal incidents are recorded, especially when the aircraft is destroyed during the accident. In both VMC and IMC weather conditions, aircraft are often substantially damaged or destroyed, with increasing fatalities in the latter case.
- Aircraft with one or two engines are more likely to record higher fatality rates.

Fatality Rate by Number of Engines



Fatality from Top 10 Aircraft reporting events by extent of Aircraft Damage



RECOMMENDATIONS

1. Aircraft Models to Avoid (High Risk)

- Cessna (High Incidence of Accidents and Injuries)

Key Models:

- Cessna 152 (most accident-prone)
- Cessna 172 (high number of accidents and incidents for private flights)
- Cessna 180 series (significant incidents in private aviation)

- Piper (High Injuries)

Key Models:

- Piper Pa-28-180, Pa-28-181, Pa-28-140 (record the highest number of injuries from accidents)

- Boeing (High Fatality Risks)

Key Models:

- Boeing 737 (leads in both total and fatal injuries)

2. Aircraft Models with Lower Risk (Recommended for Purchase)

- Zwickler Murray R (Low Incidence of Accidents and Injuries)

Key Models:

- Zwickler Murray R, Glastar (fewest injuries and accidents in private aviation)

- Yuneec (Low Incidence of Accidents for Commercial Operations)

Key Models:

- Yuneec (fewest accidents and incidents for commercial flights)

- Swearingen, Thunder and Colt, Waco (Low Risk for Engine-Related Incidents)

Key Models:

- Aircraft from Swearingen, Taylorcraft, Waco (low risk with Turbo engine types)



ACTIONABLE INSIGHTS

- **For Private Flights:** Start with low-risk aircraft like the Zwicker Murray R, Glastar for private operations, given their excellent safety records and low accident rates.
- **For Commercial Flights:** Consider Yuneec models for commercial operations due to their low incident rates and solid performance in accident prevention. Avoid high-risk models like the Boeing 737 and aircraft with reciprocating engines.
- **Engine Choice:** Focus on Turbo engine aircraft, as they tend to be safer with fewer recorded accidents, especially during critical flight phases.
- **Geography Consideration:** If operating in high-risk areas like Alaska, choose aircraft models that perform well in difficult conditions, such as Yuneec or Swearingen models.
- **Further analysis:** It will be useful to use more advanced analysis techniques e.g ML algorithms like NLP to analyze the sentiments from the `Report Status` to pin point specific details leading to the reported accidents or incidents.



THANK YOU

Safe Skies!



Tableau Dashboard