Analyzing Contributing Factors in Airplane Crashes

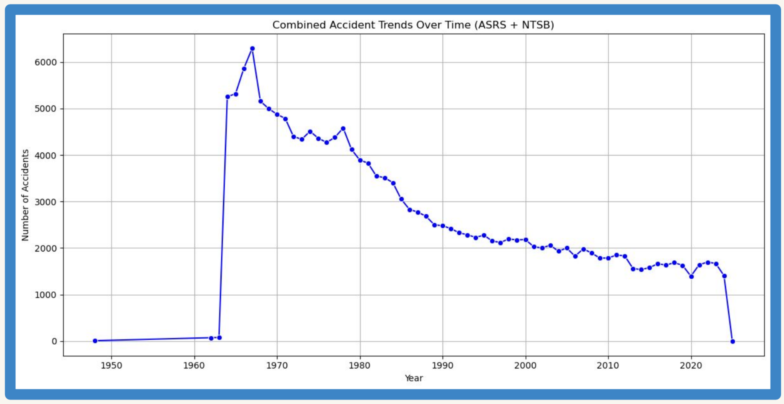
Mary Moor
Data Quality and Web/Text Mining
CS02625
Professor Breitzman
11 December 2024

Good evening everyone. My name is Mary Moor, and today I'll be presenting my project, Analyzing Contributing Factors in Airplane Crashes.
This study explores why airplane crashes occur, the contributing factors, and patterns observed across different aircraft types, regions, and weather conditions. By the end of this presentation, I'll share insights that can help improve aviation safety for pilots, manufacturers, and regulators.

Problem Statement

What's the Problem?

Crashes involving airplanes still occur despite advancements in technology. What are the key factors contributing to these accidents?



Why it Matters:

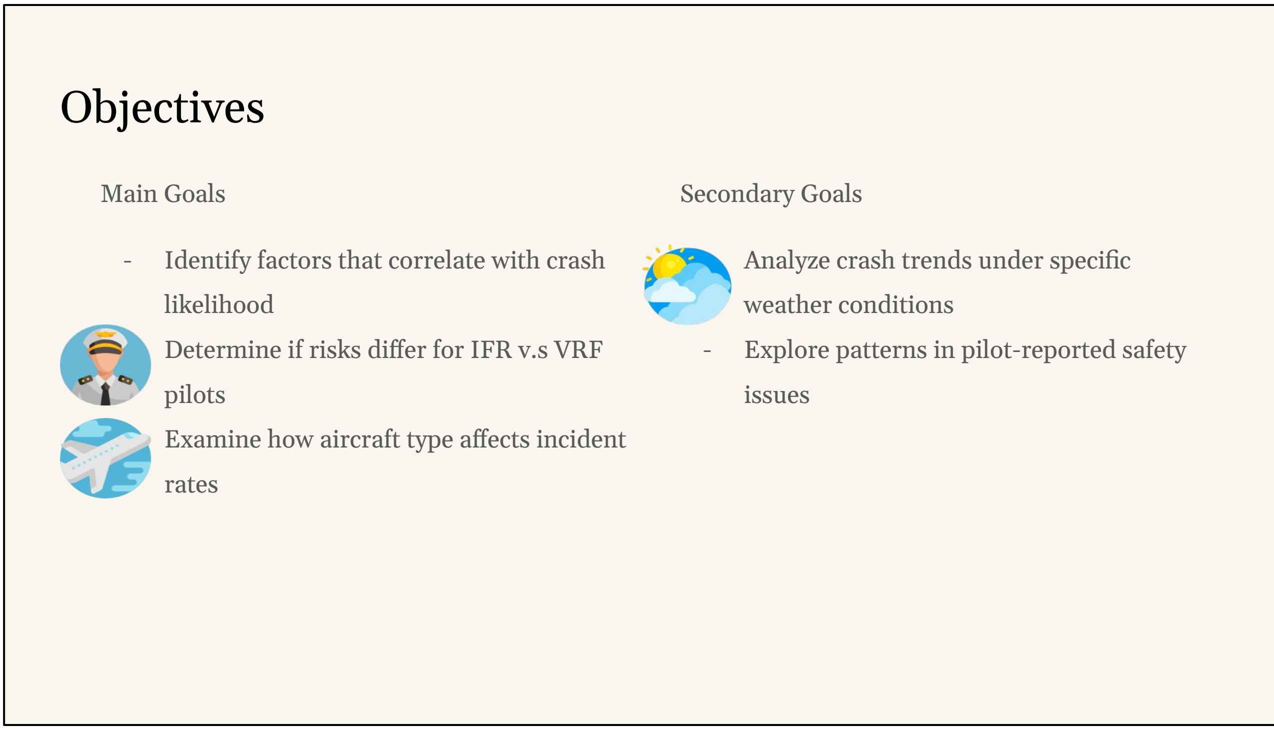
- Aviation safety is critical for reducing loss of life
- Insights can improve pilot training, aircraft design, and regulatory standards
- Highlights underexplored issues like IFR vs. VRF pilot risks

1. Problem: Despite incredible advancements in aviation technology, airplane crashes still occur. The big question is what factors are causing these crashes? Is it pilot error, aircraft type, weather conditions, or something else?
2. Why It Matters:
    - Aviation safety is critical because every crash could mean lives lost. Understanding root causes helps us improve pilot training, design safer planes, and enforce better regulatory standards.
    - Additionally, certain overlooked issues like differences between IFR (Instrument Flight Rules) and VRF (Visual Flight Rules) pilots deserve deeper analysis to close potential safety gaps.
3. Graph: Combined Accident Trends Over Time (ASRS + NTSB)
    - X-Axis (Years): Represents time, starting from around 1945 and going up to 2023.
    - Y-Axis (Number of Accidents): Represents the number of airplane accidents recorded annually.
    - Data Source: Combined data from the NTSB (National Transportation Safety Board) and ASRS (Aviation Safety Reporting System).
    - Key Observations from the Graph
        - Pre-1950:
            - The number of recorded accidents is near zero.
            - This may be due to limited reporting mechanisms or less formalized crash recording systems during that time.
        - Sharp Increase Around the 1960s:

        - There’s a steep and sudden rise in accident counts, peaking around 1970 at over 6000 accidents.
        - This likely coincides with the rapid growth of the aviation industry post-World War II, including an increase in civilian and commercial flights.
    - Post-1970 Decline:
        - After the peak, accidents start gradually declining year by year.
        - This trend reflects improvements in:
            - Aircraft technology (better safety features and design).
            - Pilot training programs.
            - Regulatory standards (like mandatory safety checks).
                - For example, the FAA implemented stricter safety policies starting in the late 1970s, likely contributing to this drop.
    - Flat Trend Between 2000-2020:
        - Accident numbers appear to stabilize at a lower level, averaging around 300-500 accidents per year.
        - This suggests consistent safety improvements, though accidents have not been eliminated entirely.
    - Sharp Drop in 2020:
        - The graph shows a dramatic decline after 2020.
        - This is likely due to the COVID-19 pandemic, which drastically reduced global flight activity.
    - Recent Data (2021-2023):
        - While there’s a slight uptick post-pandemic, accidents remain lower than historical averages.
        - This indicates that reduced flight hours during COVID may have had lingering impacts on accident frequency.

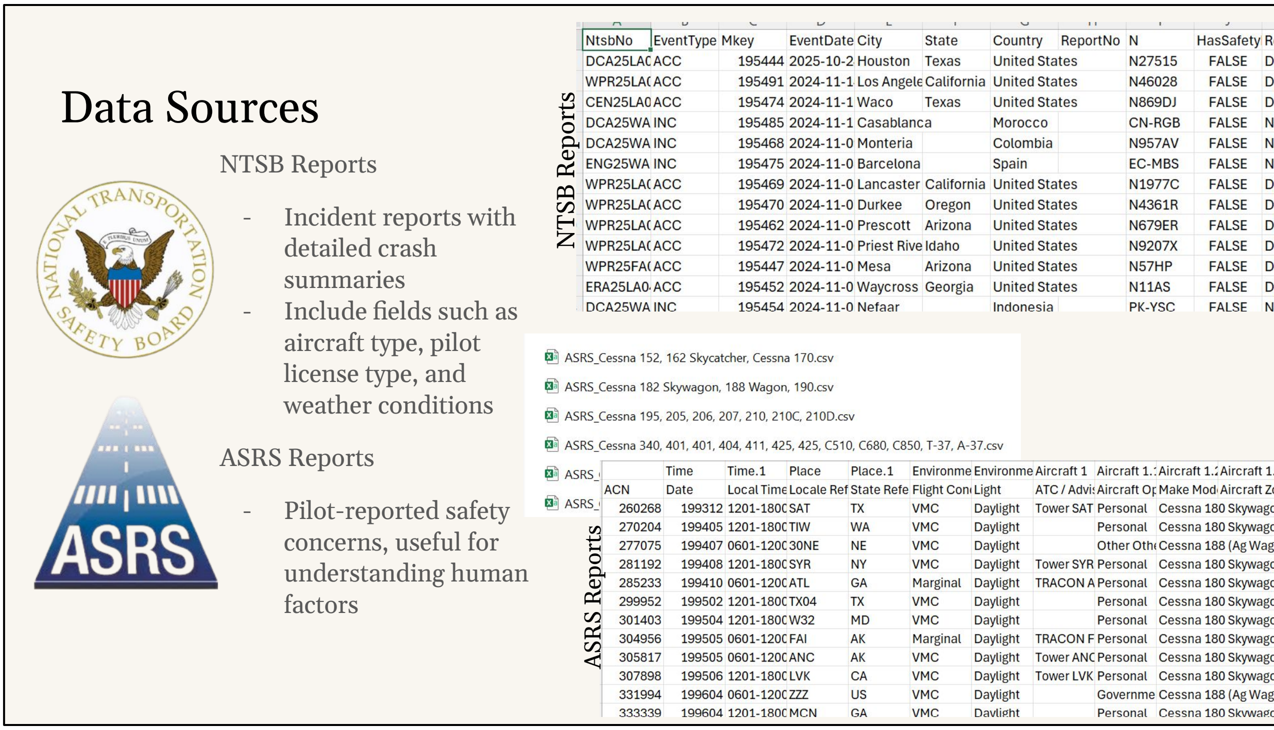
OVERALL THE GRAPH - highlights significant trends in aviation safety over time. While accidents peaked during the rapid expansion of air travel in the 1960s-70s, improvements in safety regulations, technology, and pilot training led to a steady decline.

However, the stabilization seen in the last two decades shows there's still room for improvement, particularly in addressing modern risks. The sharp decline in 2020 is notable and corresponds directly to reduced aviation activity during the COVID-19 pandemic.



Let’s walk through the objectives of this study:

1. Main Goals:
    - First, identify key factors that correlate with crash likelihood. For example, are certain aircraft models more prone to accidents?
    - Second, compare IFR and VRF pilot risks to understand how training and weather influence crash outcomes.
    - Lastly, explore aircraft type to determine which models appear frequently in crashes.
2. Secondary Goals:
    - I also analyzed crash patterns under specific weather conditions.
    - Lastly, I investigated pilot-reported safety concerns to examine human error factors not always visible in formal reports.



NTSB Reports:

- The National Transportation Safety Board (NTSB) maintains detailed records of every crash. These reports include structured fields like aircraft type, weather, pilot experience, and cause of incident.
- This is a goldmine for understanding structured patterns in crashes.

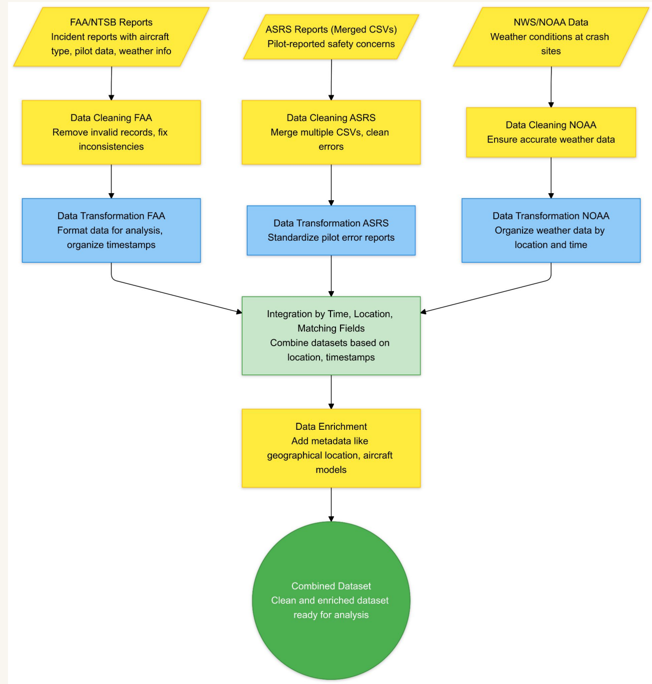
ASRS Reports:

- The Aviation Safety Reporting System (ASRS) is unique. Pilots anonymously self-report near-misses, safety concerns, or mistakes they've encountered in the cockpit. This gives us insights into human factors—like stress, fatigue, or risky decision-making—that may not appear in formal crash reports.

(Why combine these datasets?): By integrating NTSB and ASRS, I aimed to create a more holistic picture—linking human factors with hard crash data.

Flow Chart

- **Data Sources:** FAA/NTSB, ASRS, NWS/NOAA
- **Data Processing:** Cleaning, Transformation, and Integration of the three datasets
- **Data Enrichment:** Adding metadata for better analysis
- **Final Output:** Clean and combined dataset for analysis

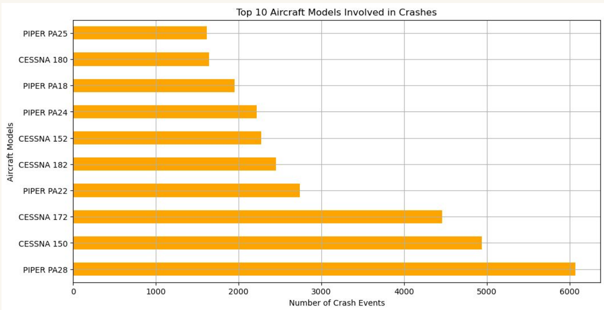


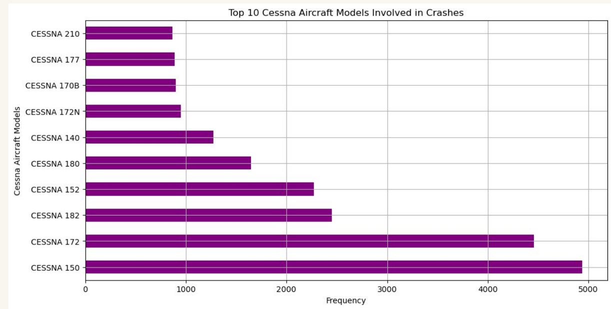
[Editor | Mermaid Chart]

1. Data Sources: I began with raw data from the FAA/NTSB and ASRS databases.
2. Data Processing:
   - The data needed significant cleaning—removing inconsistencies, null values, and duplicates to ensure reliability.
   - Transformation involved converting text data into meaningful numerical or categorical values, such as crash severity levels.
3. Data Enrichment: I enriched the datasets by integrating additional metadata, like state-specific aviation activity or weather trends. This allowed for normalized comparisons, which I'll explain shortly.
4. Final Output: After cleaning, transforming, and enriching, I built a clean dataset ready for analysis.

A clear process was essential to ensure the insights we derive are both accurate and actionable.

Top Aircrafts Involved in Crashes





PIPER PA28	6070	CESSNA 152	2272	CESSNA 150	4939	CESSNA 140	1274
CESSNA 150	4939	PIPER PA24	2217	CESSNA 172	4460	CESSNA 172N	952
CESSNA 172	4460	PIPER PA18	1953	CESSNA 182	2450	CESSNA 170B	900
PIPER PA22	2737	CESSNA 180	1647	CESSNA 152	2272	CESSNA 177	892
CESSNA 182	2450	PIPER PA25	1618	CESSNA 180	1647	CESSNA 210	867

This graph shows the raw counts of crashes for different aircraft models. The two dominant aircraft types are the Piper PA28 and the Cessna 150.

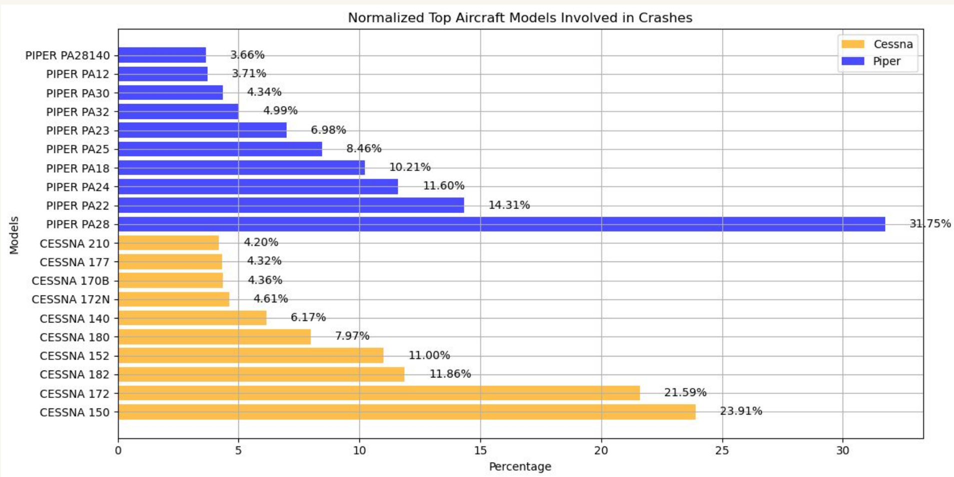
the top 10 aircraft models involved in crashes, based on data from the ASRS and NTSB datasets.
As you can see, Cessna and Piper dominate this list, with models like the Cessna 150 and Piper PA28 appearing frequently.
To create this visualization, we first cleaned and pre-processed the datasets to ensure consistent and accurate data.
For the NTSB dataset, we combined the make and model columns into a single field, make_model, allowing us to group crashes by aircraft type.
We then filtered the data to focus on Cessna and Piper models, as these were among the most common aircraft types.
The counts you see on the bars represent the raw number of crashes for each model, with the Cessna 150 leading the list.
This chart provides a high-level view of the frequency of crashes but doesn't yet account for proportional differences between Cessna and Piper aircraft, which leads us to the next slide.

1. What does this tell us?:
    - Cessna models, especially the Cessna 150 and 172, appear frequently in crash reports. This could be because they're widely used in training flights for new pilots.
    - The Piper PA28 also stands out, raising questions about its design,

    - usage, or other underlying factors.

However, these are raw counts. To truly assess crash risk, I normalized this data, which I'll show on the next slide.

Normalized Top Aircraft Models Involved in Crashes



Piper Models: Normalization reveals higher crash proportions for Piper PA28 (31.75%), indicating areas needing attention.

Cessna Models: Proportions show a clearer comparison of crash frequencies, highlighting dominant models like the Cessna 150 with 23.91% of crashes.

This slide shows the normalized crash proportions for specific aircraft models. Normalization allows us to compare crash risks relative to each model's operational use, rather than just raw counts.

We take the analysis a step further by normalizing the crash data to show the relative proportion of crashes for each model within their respective manufacturer groups. Normalization helps us compare the data fairly, especially since the total number of crashes involving Cessna and Piper aircraft differ significantly.
Here's how normalization was done:

- First, we calculated the total number of crashes for all Cessna models and all Piper models.
- Then, we divided the crash count for each model by the total crashes for its manufacturer and multiplied it by 100 to get a percentage.

The orange bars represent the normalized percentages for the top Cessna models, while the blue bars represent the Piper models.
For example, while the Cessna 150 accounts for nearly 24% of all Cessna crashes, the Piper PA28 accounts for over 30% of Piper crashes.
This visualization gives us a clearer picture of the relative risk associated with each model within its brand, rather than focusing solely on raw numbers.

Cessna Aircraft

1. Cessna 150:
   - The Cessna 150 has the highest crash proportion at 23.91% of total

- crashes. This is significant, as the Cessna 150 is widely used for training pilots, which increases its operational exposure.

2. Other Notable Models:
   - The Cessna 172, another popular training and general aviation aircraft, follows closely with 21.59%.
   - The Cessna 182 and Cessna 152 account for 11.86% and 11.00%, respectively.
3. Implications:
   - These findings indicate that Cessna models are frequently involved in crashes, likely due to high operational use, particularly in flight schools, or design and maintenance factors that may need further scrutiny.

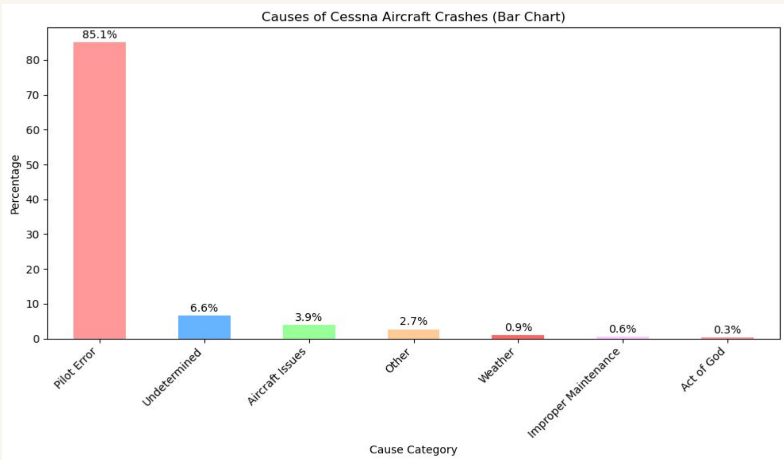
Piper Aircraft

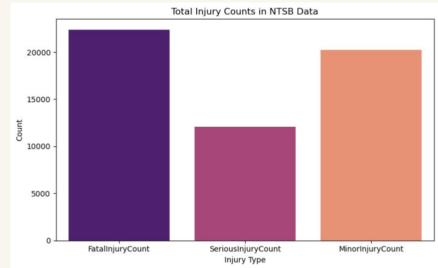
1. Piper PA28:
   - The Piper PA28 stands out with the highest proportion of crashes at 31.75%. This model is popular among private pilots, which may explain its exposure.
2. Other Piper Models:
   - Other Piper models, such as the PA22 (14.31%), PA24 (11.60%), PA18 (10.21%), and PA25 (8.46%), also contribute significantly to crash statistics.
3. Implications:
   - The higher crash proportions in Piper aircraft suggest areas that may require enhanced preventive measures, like improved maintenance checks, pilot training specific to these models, or design assessments.

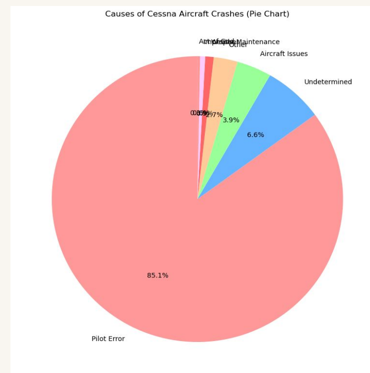
Key Takeaway

Both Cessna and Piper models dominate the crash statistics. While their high usage in general aviation and training explains part of this trend, the findings underscore the need for targeted safety interventions—whether through improved maintenance protocols, better pilot training, or model-specific inspections.

Causes of Cessna Aircraft Crashes







This slide breaks down the key contributing causes of crashes involving Cessna aircraft, using multiple visualizations to provide a clear and detailed perspective.

1. Bar Chart: Causes of Cessna Aircraft Crashes
   - X-Axis: Represents different cause categories of crashes.
   - Y-Axis: Shows the percentage of crashes caused by each category.
   -

Observations from the Bar Chart:

1. Pilot Error (85.1%):
   - The overwhelming majority of Cessna aircraft crashes (over 85%) are attributed to pilot error.
   - Pilot error may include issues like:
     - Poor decision-making.
     - Loss of control.
     - Misjudgment of weather conditions.
   - This aligns with the fact that many Cessna models, such as the Cessna 150 and 172, are used for pilot training, increasing the likelihood of errors.
2. Unknown Causes (6.6%):
   - About 6.6% of crashes are listed under unknown causes, likely due to insufficient data or incomplete investigations.
3. Aircraft Issues (3.9%):
   - Mechanical or technical failures in the aircraft contribute to 3.9% of crashes.

    - These may include engine failure, control surface malfunctions, or structural issues.
2. Other Causes:
    - Weather (0.9%), Improper Maintenance (0.6%), and Acts of God (0.3%) contribute much smaller proportions.
    - These suggest that while weather and maintenance are factors, they are far less common contributors compared to pilot error.

2. Pie Chart: Causes of Cessna Crashes (Simplified View)

- The pie chart visually complements the bar chart by showing the same categories of causes, but in a proportional manner.
- Key Highlight:
    - The dominance of pilot error is immediately clear, as it takes up the vast majority of the pie (85.1%).
    - The smaller wedges for other causes (like unknown causes and aircraft issues) reinforce their comparatively minor contributions.

3. Injury Counts Bar Chart (Top-Right)

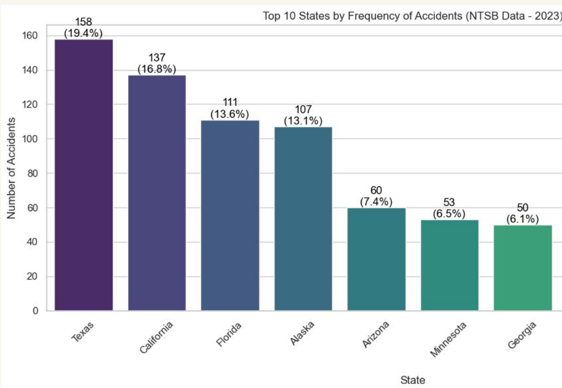
- This chart shows the total injury counts based on the NTSB dataset:
    - Fatal Injury Count: Highest number of injuries, indicating the severity of crashes.
    - Serious Injury Count: Lower, but still significant.
    - Minor Injury Count: The lowest among the three categories.

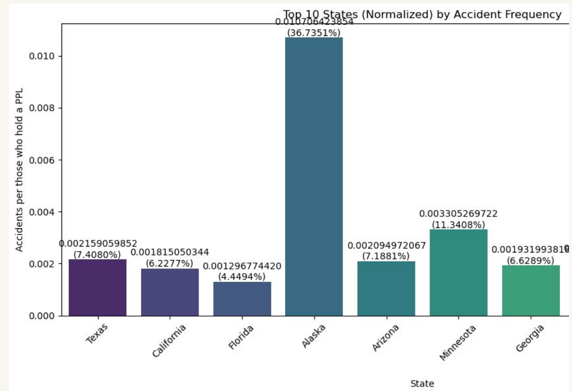
Key Takeaway:

- The high proportion of fatal injuries suggests that pilot errors and other crash causes often lead to severe consequences, emphasizing the importance of improving pilot training, operational safety, and maintenance procedures.

This slide highlights that the majority of Cessna aircraft crashes are caused by pilot error (85.1%). Given the widespread use of Cessna models in pilot training, this finding underscores the critical need for improved pilot education and advanced safety measures. While other causes like unknown issues and mechanical failures exist, their impact is much smaller in comparison.

Normalization of Top 4 States by Frequency of Accidents (2023)





Now that we've looked at total accident numbers, this slide shows a normalized view of the data for the top four states. Normalization adjusts for those who have their private pilots licence in 2023, giving us a per capita accident rate instead of raw totals.

On the left, we see normalized accident counts, showing how accidents would scale across states if populations were equal. While California still leads, Texas, Florida, and Alaska follow closely.

On the right, we focus on accidents per capita. Here, Alaska stands out with a significantly higher rate than the other states. This is because Alaska has a much smaller total of private pilots in 2023, so even a moderate number of accidents translates to a disproportionately high rate per person.

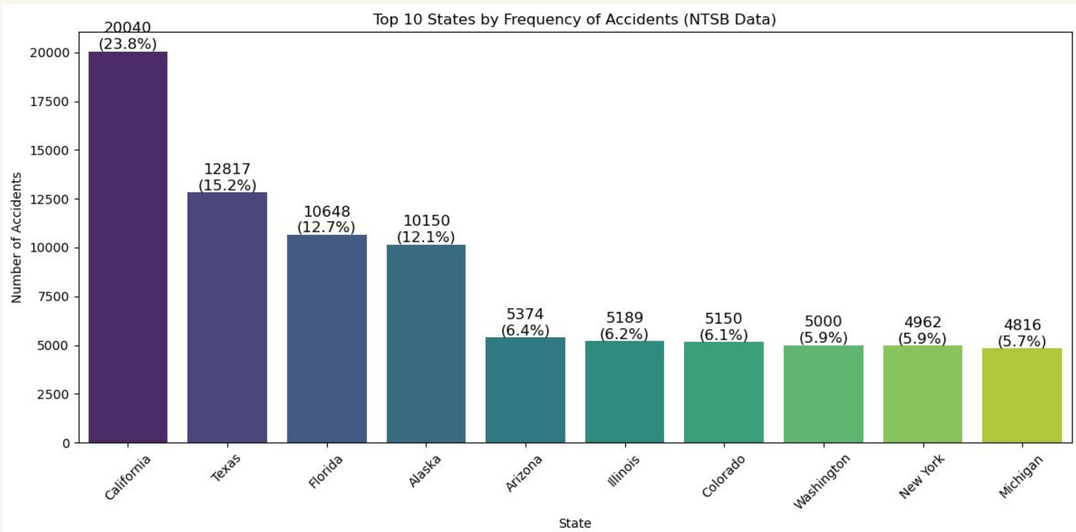
Alaska's high per capita accident rate can be attributed to its unique challenges:

- A small total of private pilots amplifies the per capita impact.
- Harsh weather, rugged terrain, and limited road infrastructure make transportation riskier.
- Long travel distances and reliance on nontraditional transportation, like bush planes and snowmobiles, increase exposure to accidents.

On the other hand, states like California, despite having the highest total accidents, have a much lower per capita rate due to their larger number of private pilots in 2023 and better-developed infrastructure.

This analysis emphasizes the importance of tailoring safety measures to the specific needs of each state rather than relying solely on raw numbers.

Top 10 States by Frequency of Accidents



This bar chart identifies the top 10 states with the highest number of reported accidents, based on data from the National Transportation Safety Board (NTSB). Each state is represented by a bar, with the height reflecting the total number of accidents. The percentage next to the state name shows the proportion of accidents that occurred in that state relative to the total. Key observations:

- California leads significantly with 20,040 accidents (23.6% of total accidents).
- Texas, Florida, and Alaska follow, with 12,817 (15.4%), 10,648 (12.7%), and 10,150 (12.1%) accidents, respectively.
- The remaining states (5th to 10th place) show fewer accidents, with Nebraska having the least in this top 10 list, at 4,816 accidents (5.7%).

Overall Findings

Aircraft Models:

- Piper PA28 and Cessna 150 dominate crash incidents.
- Normalization reveals Piper PA28 has disproportionately high crash rates (31.75%).

Geographical Insights:

- Top accident-prone states include California, Texas, and Florida.
- Normalized data highlights areas with unexpectedly high incident rates.

Crash Causes:

- Common causes for Cessna crashes include engine failures, weather conditions, and pilot errors.

IFR vs. VRF Pilots:

- IFR-rated pilots demonstrate lower crash risk under adverse weather compared to VRF pilots.

After diving into the data, here's what I found. The key contributors to airplane crashes are human factors, weather conditions, and mechanical failures. But what's really interesting is how these factors play out differently depending on the context.

For example, California has the highest number of total accidents, which isn't surprising given its large number of private pilots and high air traffic volume. But when I normalized the data, Alaska stood out with a much higher accident rate per capita. This makes sense when you think about Alaska's unique challenges harsh weather, rugged terrain, and heavy reliance on air travel for transportation.

Human error came up as a major factor in many crashes, often tied to challenging weather conditions. This highlights the importance of better training and preparation for extreme scenarios. Mechanical issues also played a role, but they weren't as frequent as human or environmental factors.

What this really shows is the need to focus on improving pilot training, investing in infrastructure, and developing systems to better handle adverse conditions especially in areas like Alaska where the risks are higher.

Overall, this project helped me uncover patterns that aren't always obvious and pointed toward actionable ways to make air travel safer. Thanks for listening, and I'm happy to take any questions!

Resources