# Investigating Aircraft Accidents in the United States of America

By Robin Stopa, Bruno Barreto, & Ian Stack

# Project Outline

- Project Source: Aviation Data

Part 1: Problem Statement

Part 2: Background

Part 3: Exploratory Data Analysis (EDA)

Part 4: Preprocessing & Modeling

Part 5: Conclusion

# Problem Statement

The purpose of this project is to design a model for the plane safety regulators and researchers that can predict the severity of an aviation accident based on contextual factors, and identify the most important factors as possible targets of regulation.

# Background

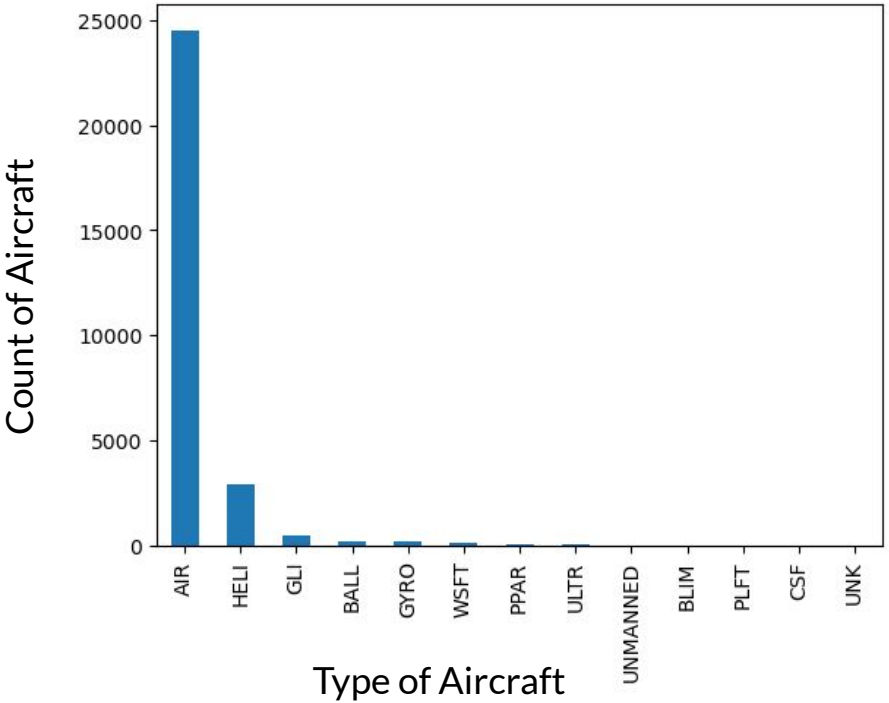
- Our NTSB aviation dataset contains data from 2002 - 2022
- Includes: accidents and incidents from across the United States.
- Flying Accidents are still dangerous and prevalent. Just recently, a plane crashed and 68 people were confirmed dead

# Preliminary Data Cleaning

- Convert given database to CSV
- Drop all columns with 30,000 + null values
- Unravel nested columns
- Convert data types


# EDA

## Aircraft per Count



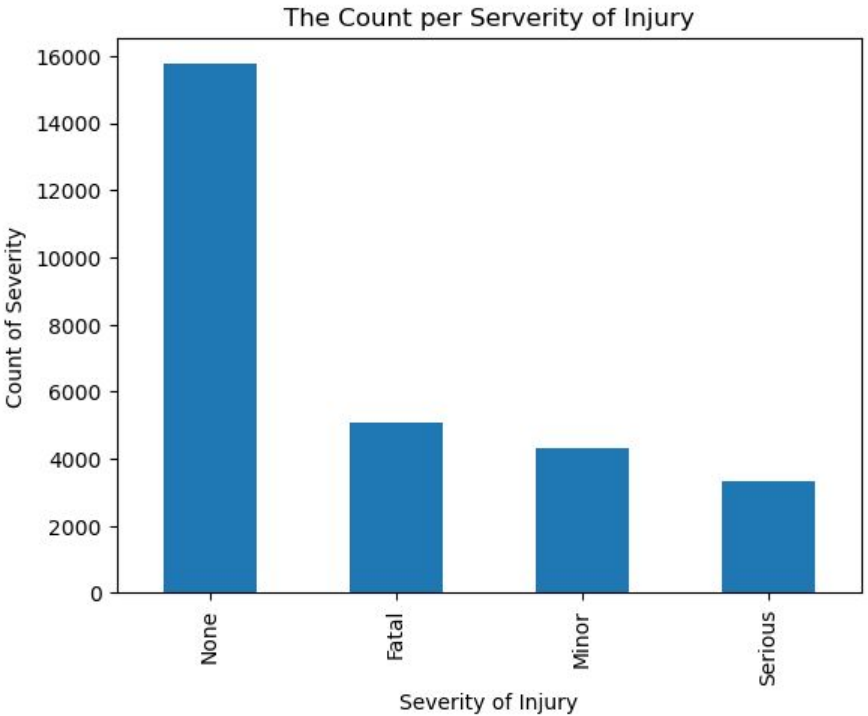

Takeaways:

- Dataset is majority airplanes

Next Step:

- What airplanes are involved?

# EDA




Takeaways:

- Most accidents have No injury

Next Step:

- What airplanes are involved in fatal accidents?

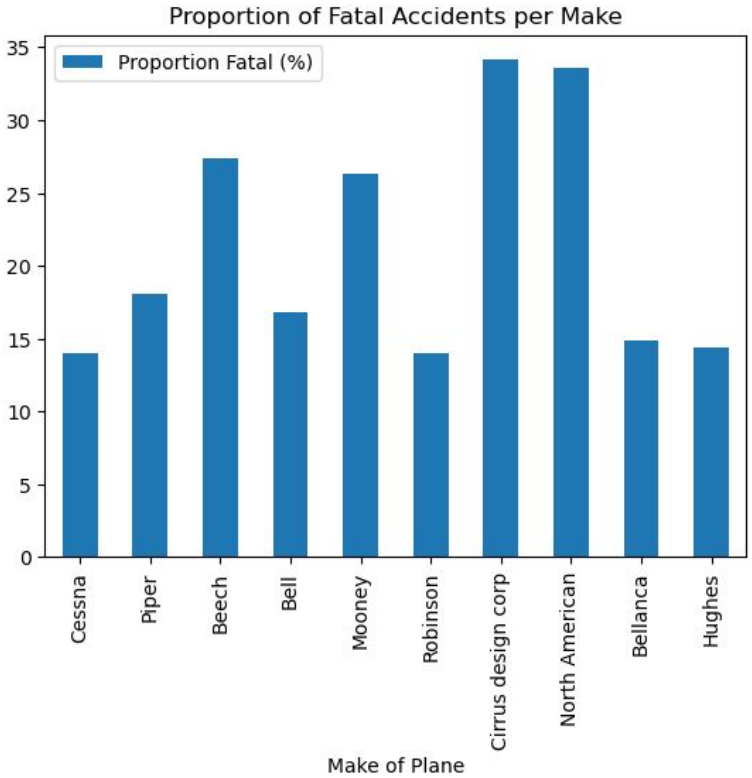
# Common Airplanes in the dataset

Beech Airplane
(Beechcraft Aircraft)


Cirrus Design Corp  Airplane
(Cirrus Aircraft)

Cessna Airplane

# EDA




Takeaways:

- Cessna is most popular plane, but proportionally, not most dangerous
- 2 types of planes have high rates of fatality

Next Step:

- Can we predict injury severity based off other factors?

# Dataset Filtering

- Majority of data includes private aircrafts
- We focused on commercial airplanes as they are more relevant to the population
- Accomplished by filtering for planes with 2 pilots

# Data Selection

- Full aviation dataset is too large to work with
- Relevant data was selected from our dataset

- **Data of Interest:**
  - Accident Report Text:
    - Reports of probable cause
    - Accident narrative reports
  - Contextual Flight Data:
    - Position information
    - Accident date and time
    - Aircraft manufacturer

# NLP (Natural Language Processing)

- NLP makes it possible for computers to read text and compute an output
- NLP was used on text-based columns of interest

NLP Tools used:

1. Removed short words and conjugation
2. Converted text to vectors of word counts

# Logistic Regression

- Used to estimate the relationship between a dependent variable and one or more independent variables

Modelling Tools used:

1. Represented each model of plane numerically
2. Scaled data down to one order of magnitude
3. Iteratively trained and selected an optimal classifier model

# Logistic Regression Results

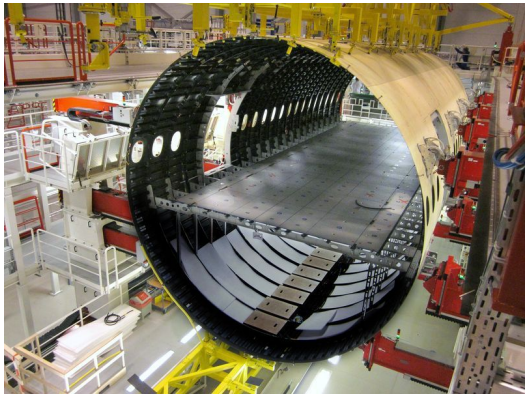
| Training Accuracy | Testing Accuracy | Baseline |
|---|---|---|
| 0.95 | 0.88 | 0.69 |

# Results and Interpretations

“Airframe”: the mechanical structure of an aircraft

- High feature importance and frequency
- Many serious/fatal accidents contained flight narrative detailing airframes
- Often used as proof that mechanical malfunction *did not* occur
- Used for accident investigation along with radar and radio transmission data

*example:*
“examination of available airframe and engine components revealed no evidence of [pre-impact] mechanical malfunction”



# Results and Interpretations

"Installation"

- High feature importance and frequency
- Cause for more than a handful of severe/fatal accidents

*example 1:*
"A shorted terminal lug on the landing gear hydraulic pump which resulted in a cabin fire. Contributing to the accident was the lack of clear installation procedures for the hydraulic pump."

*example 2:*
"A loss of engine power due to the in-flight separation of the 1-3-5 cylinder induction tube elbow, which was caused by the improper installation of the induction tube elbow by maintenance personnel."

# Conclusions and Recommendations

- Most aircraft are built safe
- Weather was not as important as we expected
- Radar, radio, and airframe data appears frequently in accident reports
- Improving mechanical installations is the biggest target for regulation

Regulators and researchers should:

1. Focus on trainings for maintenance personnel with regards to installations
2. Audit installation procedures to make sure they are correct, standardized, and well-documented
3. Research particular types of installations which highly influence airplane accidents

# Limitations

- Run time,  doc objects
- Notebook memory consumption
- Dataset dimensions
- Inconsistency in report styles
- Over-representation of private planes
- Lack of non-accident data

# Sources


Database: https://www.ntsb.gov/Pages/AviationQuery.aspx

Photo's: https://www.flyingmag.com/aircraft-pistons-evolution-cessna-172/

https://www.charterhub.com/blog/fun-stuff/2017/11/beechcraft-history

https://cirrusaircraft.com/

https://aviation.stackexchange.com/questions/45566/how-thick-is-the-skin-of-an-aircraft-like-the-airbus-a350-or-boeing-b777

http://www.aviationpartnersboeing.com/services_retrofit.php

https://illustoon.com/?id=444

Information: https://en.wikipedia.org/wiki/Airframe