SKYWATCH

Riccardo Falsini, Kerin Grewal, Tuna Sogut

Introduction

General aviation accounts for 93.6% of all fatal aviation accidents, despite this there is little knowledge on what factors lead to such incidents.
Our Project seeks to:

(1) Compile and pre-process Federal Aviation Administration (FAA) data and NASA's Aviation Safety Reporting System (ASRS) data regarding 'general aviation accidents' to allow for easier future analysis.

(2) Analyze said data to help gain a better understanding on what the most frequent factors that lead to such accidents are.

Raw Data

- The ASRS data consists of 1500 self-reported unstructured text data. This data has 50 samples for 30 different types of general aviation accidents.
- The FAA data contains 27 fields of information about reported incidents.
  - The dataset contains 100,000 reported accidents which occurred in the 50 United States as well as select US territories such as Puerto Rico and Guam.
  - These accidents date from 1978-2017.

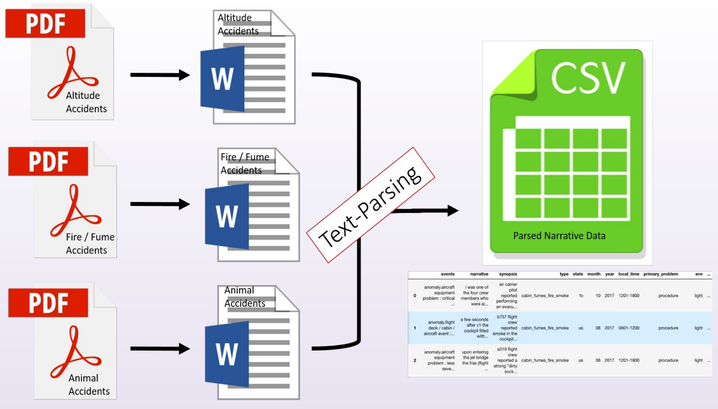

Figure 1:Pre-Processing Steps

Pre-processing for ASRS Data

- Extracted each report's contextual data from the unstructured text.
- Created a reusable parser to extract segments from these reports.
  - Formatted the data to be indexable.

Analysis and Results

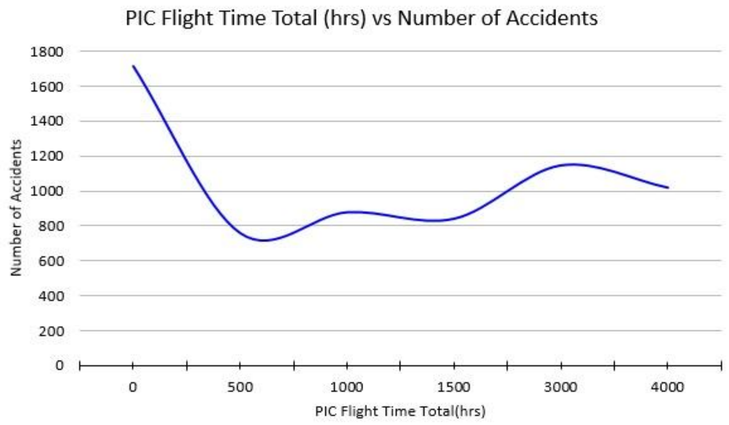

Figure 2: The Overconfidence Effect

Overconfidence Curve

- More experience equals less accidents, until we hit the overconfidence point. Then, overconfidence in ability causes more accidents

NASA ASRS Data Analysis

- Used SVM to predict accident type from flight context (environment, location, primary problem, season, plane model)
  - Accuracy: 26.3%
  - Low but significant since a random prediction would be 3% accurate as there are 30 possible accident types to predict
- Procedural errors and human factors make up 66% of accidents.
  - Figure 2 shows how as pilot confidence increases, so do accidents.
  - In both the FAA dataset and the NASA data we found evidence of this trend.

Understanding the Results

- The accuracy of 26.3% was achieved using a small and inconsistent dataset (1500 samples of unstructured text)
- Our result is significant because it shows that with more and better data, we can begin to discern the predictors of accidents
- Because procedural errors and human factors are the largest primary causes of accidents, there exists opportunity to create tools to improve procedural checks and automate human efforts

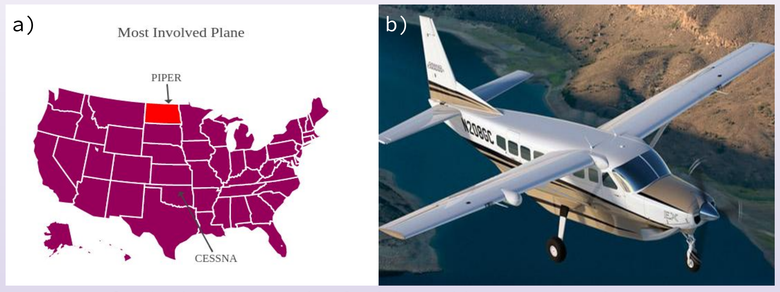

Figure 3: a) Aircraft most involved in accident per state b) Cessna Grand Caravan, the most common plane in General Aviation accidents in all States except North Dakota

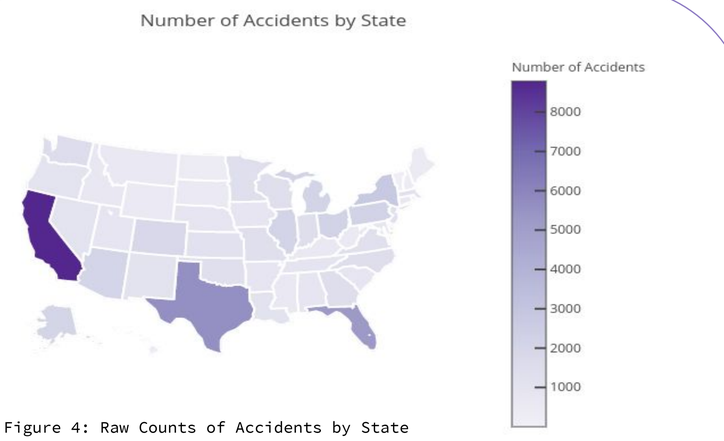

Figure 4: Raw Counts of Accidents by State

FAA Data Analysis

- The FAA Dataset did not require the same level of pre-processing because it was presented in an indexable format.
- We initially looked at raw counts to find what airplane make, US state, PIC Certificate Type, airport, and flight phase were the most common for these general aviation accidents.

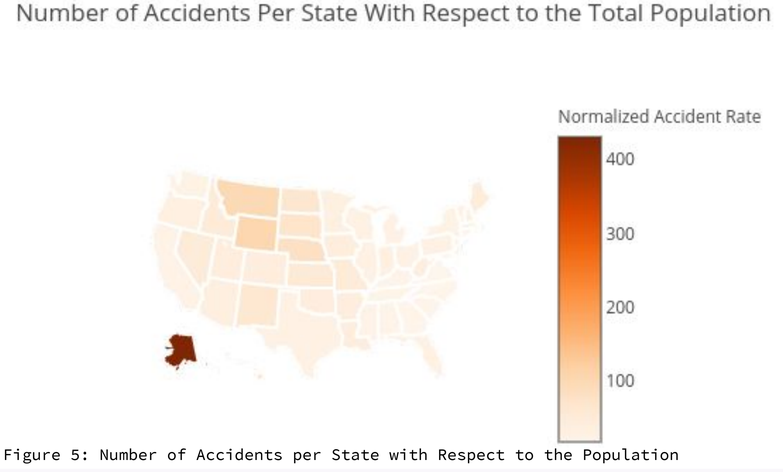

Figure 5: Number of Accidents per State with Respect to the Population

Understanding the Results

- As we started to notice trends, we created maps which pointed out which state is most common for GA accidents and which plane is most involved.
- We found that the data needed to be normalized by different constants depending on the factor in question.
- Figure 3 shows non-normalized counts of the make of each airplane in the incidents.
- Figure 4 shows the raw counts of accidents per state. We recognized that this data was skewed by state population, so we used census data to model this in Figure 5.

Further Research

-With access to a 'control' dataset we could not only properly normalize our data, but also find factors that tie disproportionately and significantly to accidents.
-Utilizing such a control dataset would also allow us to develop models to classify and predict various types of aviation accidents based on different flight features (weather, plane-type, location etc..)