**DS5K Final Project Pitch**

**Problem Statement #1**

Over the relatively short 118 year history of powered human flight, the human race has achieved the unachievable time and again. From the first slow short hops of the Wright brothers in 1903 to supersonic jet flight only 44 years later, and to the moon only 22 years after that, untold numbers of people have perished in the advancement of aviation technology. These days, we think almost nothing of boarding a machine that can fly us at altitudes approaching 40,000 feet at almost the speed of sound to get us from one side of the country to the other in mere hours. While the technology is proven and safe enough for us to not worry about the miniscule chance of a serious problem resulting in a devastating outcome, those problems do still arise. There are many causes for aviation mishaps and accidents, and whenever one occurs, investigators gather every possible data point that could lead to a better understanding of what happened and how to prevent it in the future.

**Data set(s)**

For this exercise, I will be analyzing aviation mishap data sets provided by the Federal Aviation Administration (FAA) Aviation Safety Information Analysis and Sharing (ASIAS) system. This system provides links to multiple source databases provided from the FAA, the National Transportation Safety Board (NTSB), the Bureau of Transportation Statistics (BTS), and the National Aeronautics and Space Administration (NASA), among others, that track various safety issues and incident and accident reports.

**Sample questions I seek to answer**

I will begin with exploring the data and producing summary statistics

Are there significant clusters of primary and secondary contributing factors in aviation mishaps when looking at different types of mishaps? For example, do stall or spin mishap types occur more frequently when there is a specific combination of contributing factors (e.g., meteorological, physiological, or mechanical, etc.)?

Are there specific contributing factors (or groups of factors) that, if focused on via training or other resources, would have a more significant impact on aviation mishap rates for the most common mishap types?

**Possible Challenges**

There are several data sets provided from multiple organizations that are each very large. I will need to study each data set thoroughly to determine the best combination of data and questions to ask.

The Federal Aviation Regulations (FAR) define “accident” and “incident” in such a way that incident reporting is voluntary in many cases, and jurisdiction over an accident/incident falls under different agencies (i.e., the FAA or the NTSB) depending on the circumstances. This complicates the selection of data sets and the questions I will ultimately decide to seek answers to, and could also complicate the reporting of results.

**Problem Statement #2**

**Data set(s)**

For this exercise, I will use the Amazon Pricing Data housed in the Sailfish Exchange repository.

**Sample questions I seek to answer**

I will begin with exploring the data and producing summary statistics

Are

**Possible Challenges**

There are several data