Research Idea 2: Mini Project

Elsa Hritz

Research Question:

Are there distinct commonalities between various environment, mechanical, and physical components of aircraft and their flight between planes that crash and those that do not and could plane crashes be predicted?

Data Information and Retrieval

The data comes from a few different places. The Federal Aviation Association holds many different data sets for flights. Here I would get all my information for non-crash flights including the pressure, temperature, winds, turbulence, flight locations, and type of aircraft. For my crash data, I would use the airplane crash dataset from Kaggle. This data frame includes the aircraft type, location of the crash, departure airport, expected arrival airport, time, date, and number of fatalities of the crash. I would use the location, date, and time information to then merge a weather data table with the crash data to have the pressure, temperature, and winds for that flight. The data is tangible, as it is clean and only needs to be merged with an easily accessible weather table in order to have all the variables I need. The data is extensive and allows me to consider multiple different models for my analysis. It would be retrieved through the locations mentioned above using a simple download and then rearranged and merged in R myself.

Model

I plan to merge all my data into one data frame with both the successful and crashed flights--including a binary column representing if the plane crashed. I then would use both clustering and regression models to analyze the data. I would create a new data frame that does not include information about the crash itself only the airplane, weather, and departure/arrival airport columns. Then, I would cluster all the flights and evaluate them based on their assigned cluster. I am curious whether some clusters have considerably more crashes and why. I would also train and test a regression model to determine if it is possible to predict whether or not a plane will crash--if it can, we know something must be changed in our choice to fly or how we build our planes. Otherwise, I would use it to supplement the clustering in finding connections or the most impactful variables in plane crashes.

Stakeholder Implications

Stakeholders would be interested in this project because it affects the well-being of airlines--as crashes in one airline could influence people to buy tickets from a different airline. Additionally, if a model exists that could predict crashes, then airplane mechanics can use that information to enhance the parts of the airplane that counteract the influence of that variable. Additionally, if certain weather conditions affect crashes, the FAA could be more on top of grounding flights before something goes wrong. In general, stakeholders want people to buy tickets to their airlines and to be known for their safety--preventing crashes allows for just that.

Ethical Considerations

Some ethical considerations when using this data include the description column on my crash dataset. This column uses some identifying features of the pilots in the crashes. Those combined with the aircraft tail number, date, and time of the flight could be used to determine exactly who was flying the aircraft. This could not only negatively affect the pilot's career but their personal life as well. If people knew about their crashes and treated them poorly, even if the crash was an accident, that pilot may never choose to fly again. Additionally, if models are used to predict potential crashes, but they are not incredibly accurate, what should be done? Should the flight be grounded, or should they risk a potential crash? Then, if the plane flies and there is a crash, how should that be handled? Moreover, we would not want to be too conservative and ground many planes unnecessarily-- as this could also give a negative impression for some airlines.