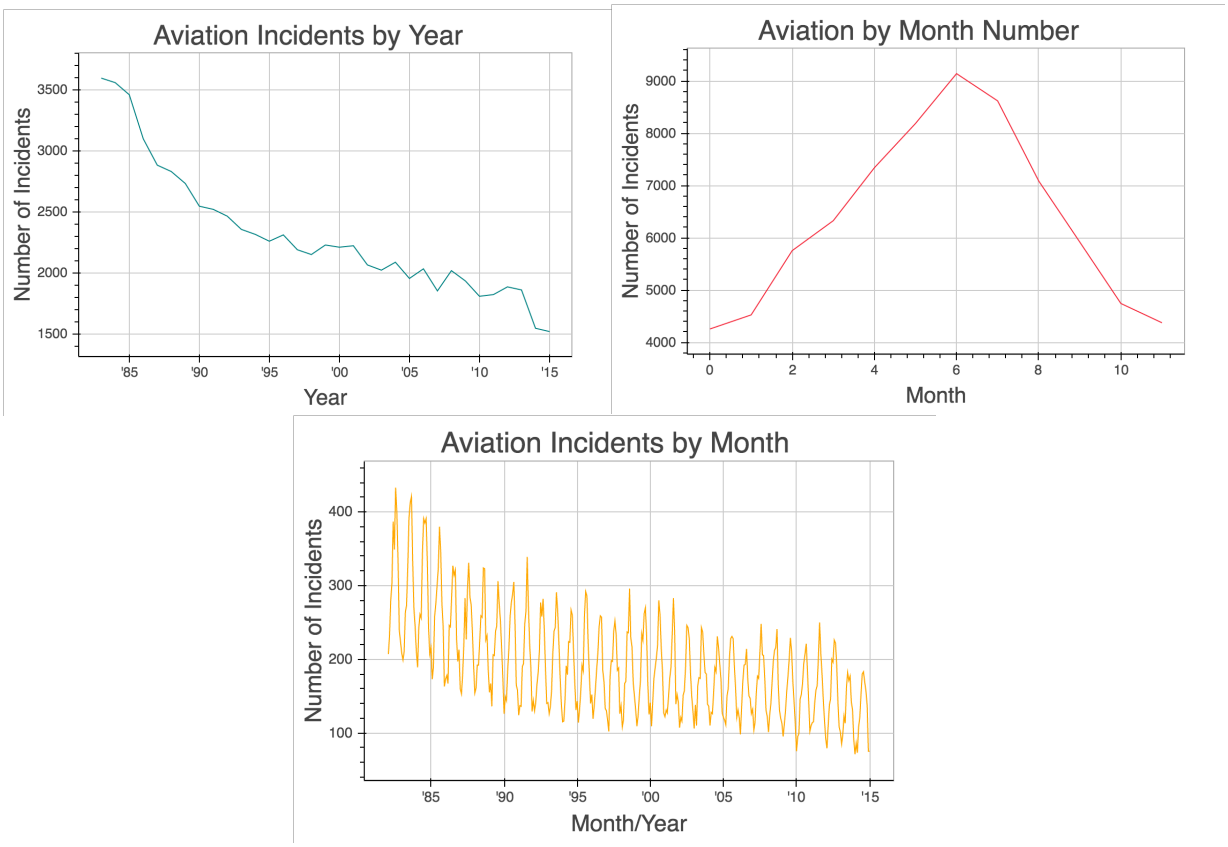
Aviation Accidents Data

Data taken from the National Transportation Safety Board (NTSB) was briefly analyzed for initial insights as well as for identifying future areas of research. Available in the data was records for each flight incident from 1982 to 2015, along with an accompanying data set of NTSB investigation findings.

Trends in the Data

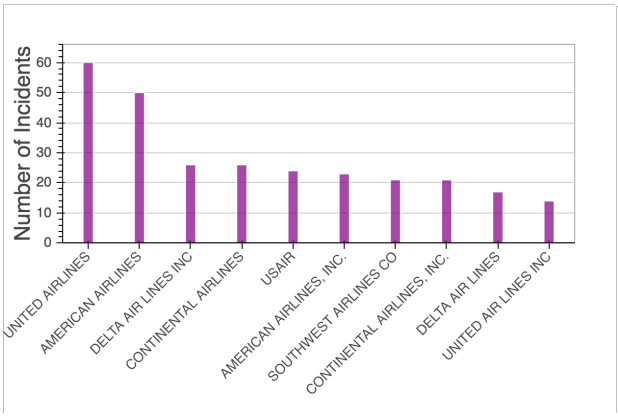
Since 1982, the number of aviation incidents has followed a downward trend. As well, aviation accidents appear to follow very seasonal trends, with summer months being higher than the rest of the year.

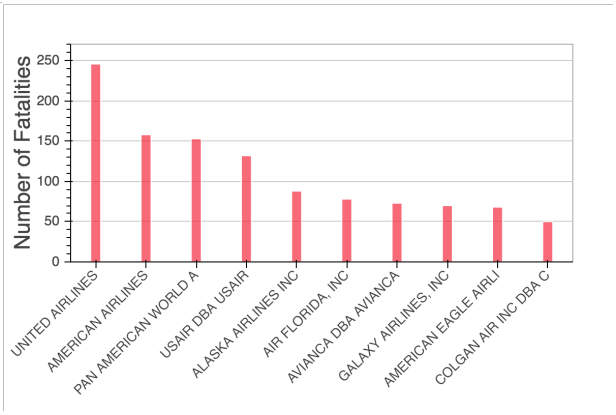


This could be for many reasons, including the increase in travel and recreation aviation over the summer months. However, further statistical investigation should be taken to determine the reasons behind the summer spike.

Air Carrier Performance

That data also poses interesting questions around air carrier performance. When not accounting for other factors such as time, or events beyond the control of the carrier (i.e. – terrorism), similar carriers appear at the top of the list for most incidents and most fatalities.





To further evaluate carrier performance, outside data on all flights would be necessary in order to standardize these metrics. Is United Airlines' high number of incidents all that high when compared to the number of flights they carry out? Are there certain types of events that certain carriers are more prone to having based on the number of flights they are operating?

NTSB Probable Cause Text

Investigation in to the types of incidents that occur leads into analyzing the NTSB investigation data. The data is in text form, which is often difficult to group without manual labor. However, after doing very cursory exploration, the data appears to have underlying groupings that can be modeled and described with future work. By grouping NTSB probable cause reports based on word counts, patterns appear for each group that can give insight into the general narratives for each group.

For example, one of the cursory groups created from the text showed higher correlations with the words "fuel", "exhaustion", "engine", and "inadequate power". It could likely be determined that this group of probable causes are mostly related to fueling of the aircraft. More research should be done to investigate the natural groupings of probable cause text in the data and to profile those groups so that they can be used in other capacities such as predictive modeling.