This analysis explored trends in aviation accidents from 1982 to 2015 and examined significant factors of increased risk of accident using structured data as well as text descriptions.

The amount of total aviation accidents as well as fatal accidents have decreased dramatically since 1982, as seen in Figure 1. However, there are certain factors that can be avoided to further reduce the risk of being involved in an aviation accident. The most critical factors that increase risk are: ***1) higher traffic flight times and areas, 2) small and/or amateur built planes, and 3) poor weather conditions.***

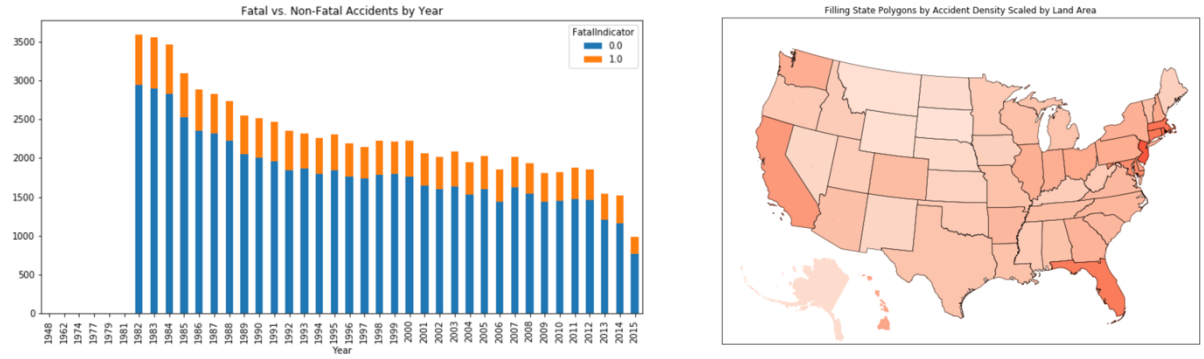

Figure 1. Descriptive Plots of Accident Trends Over Time and Across the United States

A higher proportion of aviation accidents occur during summer months and on weekends when there are more flights. The map in Figure 1 above displays all accidents in each state of the United States scaled by land mass. There is a higher density of accidents in higher traffic areas such as Florida, California, and the Northeast. The most common airplane makes and models of all accidents as well as fatal accidents are the ***Cessna 152, Cessna 172, and Piper Pa-28-140***. These are all two-seater, single-engine planes.

The significant effect of poor weather conditions was largely observed trough text data. Accident narrative and probable cause descriptions provide insight into the external and internal factors that caused an accident, thus enabling greater knowledge of the most common causes of aviation accidents.

Unsurprisingly, the most common terms in the narrative text are ***airplane/aircraft, pilot, and accident***. Accident narratives before 2000 tend to focus more on the investigation and source whereas after 2000, there is more descriptive language about the cause of the accident such as engine and landing.

The narrative text for each document was grouped into one of four different clusters based on the dominant topic: ***1) weather conditions, 2) under investigation, 3) engine failure, 4) landing.*** Figure 2 shows the accidents clustered by structured data including weather condition, broad phase of flight, and engine versus the same accidents clustered by narrative topic. The boundaries of the narrative clusters are much less sharp and the weather cluster is at the very center.

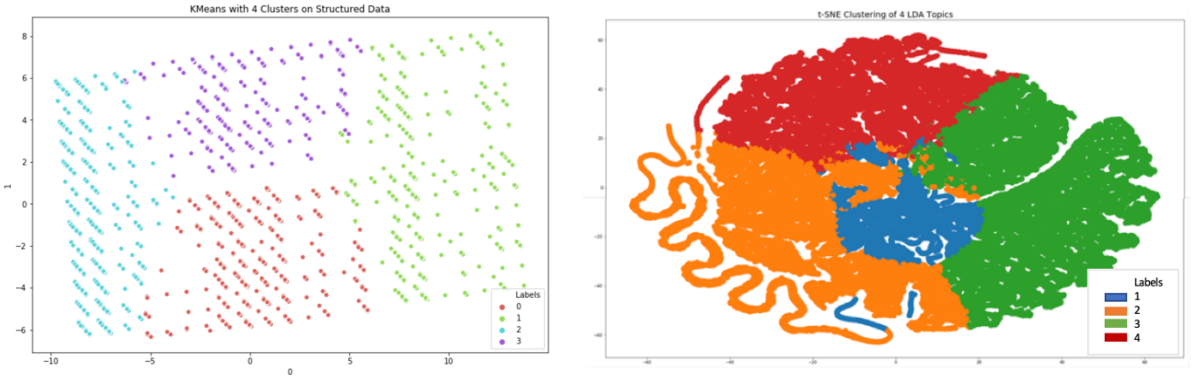

Figure 2. Comparison of Clustering on Structured Data versus Text

In summary, it is recommended to avoid flying during high traffic times in high traffic areas, in small or amateur built airplanes, or in poor weather conditions to reduce the risk of being involved in an aviation accident based on this analysis.