National Transportation Safety Board (NTSB) Aviation Accident Data Analysis

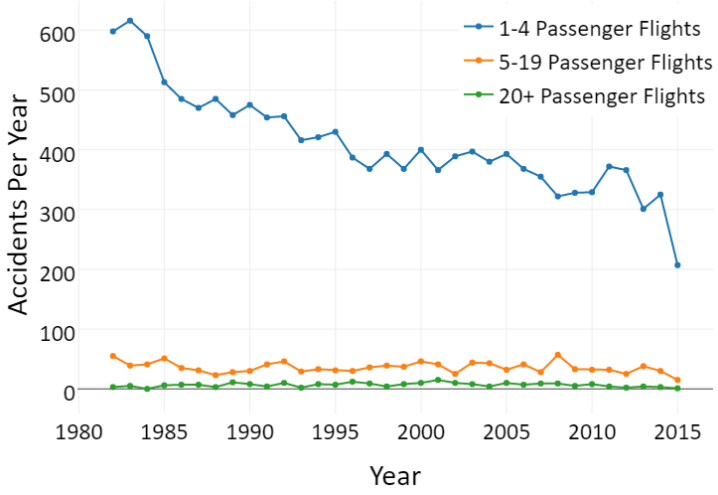
This overview summarizes preliminary findings from the NTSB's publicly available Aviation Accident Database. The database has detailed information and accident findings for 77,257 accidents which occurred from 10/1948 through 9/2015. A customer desiring a reduction in flight accident risk should fly on large commercial aircraft as the majority of accidents (91%) occur on planes carrying only 1-4 passengers (see *Table 1*).

Table 1 - Summary Data from NTSB Accident Database

	Number of Passengers On Plane				
	Small (1-4)	Medium (5-19)	Large (20+)	Unknown	Total
Number of Accidents (all types)	70,148 (91%)	4,022 (5%)	2,395 (3%)	692 (1%)	77,257
Number of Accidents (w/ fatalities)	13,988 (91%)	1,217 (8%)	223 (1%)	0	15,428

Even though large flight volume tripled in the timeframe shown, *Figure 1* shows that the number of accidents in this group has not increased. Additionally, though the number of accidents for small flights have dropped significantly during the period shown, primarily due to fewer hours being flown by small planes, the fatal accident rate over the past decade for small planes still remains more than 60x higher than the rate for large plane flights.

Figure 1 - Aircraft Accidents with Fatalities Over Time by Number of Passengers on Flight (1982-2016)



Accident Cause Analysis

Given the large differences in accident frequency for large and small passenger aircraft, NTSB accident probable cause texts were analyzed for patterns between the groups, as shown in Table 2. Accidents with fatalities tend to be human error related, with small aircraft having pilot errors, while large aircraft have traffic control errors. Accidents without fatalities tend to be mechanical in nature, with small aircraft losing engine power or running out of fuel, and large aircraft having landing gear problems.

Table 2 – Preliminary Probable Cause Text Analysis

Small Aircraft	No Fatalities	91%	Mechanical (Engine & Fuel etc.)
	With Fatalities	94%	Human Error (Pilot)
Large Aircraft	No Fatalities	97%	Mechanical (Landing gear etc.)
	With Fatalities	85%	Human Error (Traffic control)

Preliminary Report Caveats

- Accident rates would be more meaningful if accident-free flight frequency data was also known.
- Detailed verification of accident causes is needed to confirm the unsupervised text analysis groupings.

Python source code and raw data repository: https://github.com/epreble/data-scientist-exercise02