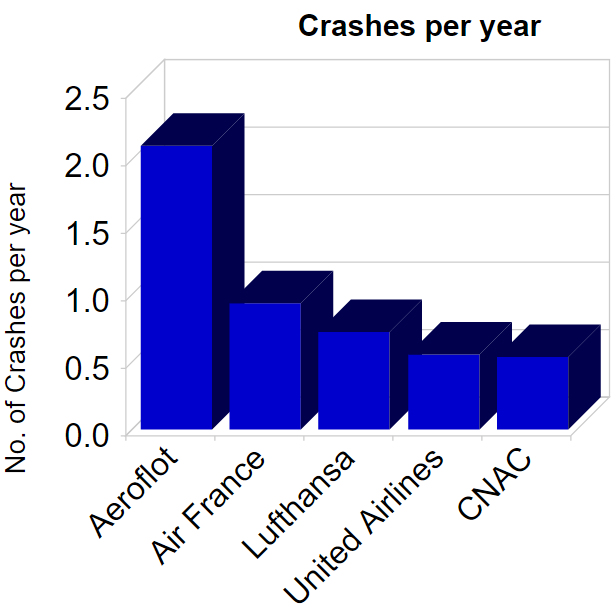
Summary

Our data set contains data about the air crashes from 1908 to 2009. We obtained this data set from Kaggle.com. The dataset is originally hosted on OpenData by Socrata and made available on Kaggle.com by Sauro Grandi. The dataset contains 5268 rows in total and has information including but not limited to the date, time, location, operator, fatalities, and summary of the crash.

The main target audience of our analysis include air travelers, the air crash investigators, the airline companies, and the air-craft manufacturing companies. Based on our research questions decisions can be made on what is the top priority for the aircraft manufacturing companies or the airline companies to be able to reduce air-crashes. Decisions can also be made by the air travelers about the airlines they can avoid to ensure a safe journey. An air crash investigator will not just be interested to know what reason is causing the maximum number of air-crashes but also what type of these crashes lead to the maximum number of fatalities so that decisions can be made and appropriate actions taken to reduce this number.

The most important cleaning activity we did on our data was the classification of crashes based on the reason for the air crash. For example, wherever the summary contained keywords like ‘engine failure’, ‘malfunction’ etc. we have assigned such crashes to the ‘Technical failure’ category and followed the same process for assignment to other categories. We also assigned the air crash time into 2 major time slots of day and night. On the right is one of the plots we came up with which shows the number of crashes on an average for the top 5 operators with the maximum number of air crashes. We considered the time around which these airlines started operating and normalized the data by dividing the total number of cashes per airline with the number of years they have been under operation till 2008. We also explicity excluded the military aircrafts from this analysis because military aircrafts usually have war related reasons for the crash.

Looking at the plot, one can clearly tell that by far ‘Aeroloft’ seems to be the most unsafe operator with an average of more than 2 aircrashes per year followed by Air France and Lufthansa which have less than one crash per year on average.

As per our analysis, since the majority of the aircrashes happen due to technical failure reasons, that should top the priority list for both the operators and the aircraft maufacturing companies who need to ensure that safety of the airplanes is increased. The airtravellers might reconsider travelling by operators including ‘Aeroloft’ ,’Air France’ , ‘Lufthansa’, ‘United Airlines’ or keep travel by these airlines to a minimum to ensure a safer and worry-less travel. Since hijacking(culminating in a crash) has resulted in the maximum fatalities on an average, tighter security norms need to be established at airports, especially in the modern era where terrorist attacks are common and the culprits have come up with novel ways to fool existing procedures.

Word count: 511