

## Data Source and data sets

WorldBank Data:

<https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2020&start=1970&view=chart>

Raw data file name:

1. API\_IS.AIR.PSGR\_DS2\_en\_csv\_v2\_4499051.csv

Power BI Query Name:

1. WorldBankAirTransportPassengers

Aviation Safety Network Data:

<https://www.geckoboard.com/blog/6-data-visualization-techniques-to-display-your-key-metrics/>

Raw data file name:

1. airline-safety.csv
2. table\_02\_14\_102521.xlsx

Power BI Query Name:

1. airline-safety
2. US General Aviation Safety Data

## Data transformations

`airline-safety` dataset did not need much transformation. It contains data for 56 international airlines. These airlines were in the global top 100 list as of 2012. Data shows incidents, fatal accidents and fatalities spanning 30 years starting from 1985 ending 2014. Data for each category is divided to 1985-1999 and 2000 to 2014. We added a new column to the table. The new Column, "Ranking" was calculated based on the value of available seat per week. Those with more than 2 million were ranked top, those between 1 and 2 million were ranked Middle and those below 1 million were ranked bottom.

`US General Aviation Safety Data` was transposed, filtered and transformed into a 10 column table . It has the fatalities and incident counts for all US airlines from 1970 to 2020.

`WorldBankAirTransportPassengers` dataset contained data for number of passengers traveled worldwide from 1960 to 2020. The data set has this data for every country in the world as well as a total for the world. We transformed the table to two columns-Year and Passenger traveled worldwide. We used this transformed table to create a metric for the presentation. The details of the metrics are described below.

## Data Presentation Plan

We plan to present to the executives of several airlines. The goal of the presentation is to alert them of the negative publicity that airline crashes can cause. We will present data showing number of fatalities and incidents in the US. We will also present air traffic data worldwide

## Data Presentation

### Choice of visualization

For the purpose of this presentation, we added a column to the airline-safety data set to rank the airlines into Top, Medium, Bottom. The ranking is based on number of seats available per kilometer per week. The higher number the more kilometers traveled and more passengers. The ranking is color coded in the scatter plot for fatalities and incidents to highlight that popular airlines have fatalities and incidents just as much as the not so popular ones. However, it should be noted that the more an airline travels, the chance of fatal accidents increases. This could be the subject of further study.

We showed number of fatalities and incidents in bold, and a line graph showing the growth of flights worldwide. We indicated the drop in airline travel due to the Covid pandemic.

## Overall Findings

Overall, we did not see a relevance between fatalities and accidents with airline travel and the public's opinion of air travel. It would be worth studying other travel options such as bus, rail, ship, personal and vehicle. It would also be worth looking at the trend of each option with respect to each other counting geopolitical, and other environmental influences.

## Ethical considerations

The dataset for this study does not consider any environmental influence on the data. It is raw. Other sources of data would have more observations and more data showing airlines' maintenance records, rankings in customer service, flight experience, etc.