Airlines Safety Stats

Thip Rattanavilay, Master of Data Science

Bellevue University

DSC640

Author Note

The topic detail has been selected on personal interest and the data has been collected from online resources mentioned in the reference section.

**Project Task 1: Dashboard**

*Your first task is to create an internal dashboard for your peers and data science management team that outlines the facts – what are the stats and what are the trends? Is there any supplemental data that you can use to support that air travel is still in fact the safest? Is there anything politically going on that would cause this type of media attention to be at a peak – remember, this is for an internal review by your peers and management – and will likely spark a lot of discussion for how you approach the next level of discussion with your executive leadership team. Is there anything to show sales are down or are headed that way? Do the safety incidents appear to be in a specific geographic area or by a specific airline every time? Do some analysis of the data you have and look for other sources to see what you can find to help inform your internal team.*

**Source Datasets:**

1. Main data: Airline Safety

Data for 56 airlines that were in the global top 100 as of December 2012 and which have operated continuously since Jan. 1, 1985.

Codebook:

|  |  |
| --- | --- |
| **Header** | **Definition** |
| airline | Airline (asterisk indicates that regional subsidiaries are included) |
| avail\_seat\_km\_per\_week | Available seat kilometers flown every week |
| incidents\_85\_99 | Total number of incidents, 1985–1999 |
| fatal\_accidents\_85\_99 | Total number of fatal accidents, 1985–1999 |
| fatalities\_85\_99 | Total number of fatalities, 1985–1999 |
| incidents\_00\_14 | Total number of incidents, 2000–2014 |
| fatal\_accidents\_00\_14 | Total number of fatal accidents, 2000–2014 |
| fatalities\_00\_14 | Total number of fatalities, 2000–2014 |

1. Supplemental data: Auto Fatalities

Data showing number of people killed and injured in fatal collisions.

Codebook:

|  |  |
| --- | --- |
| **Header** | **Definition** |
| Year | Year of incident |
| Deaths | Number of fatalities |
| VMT\_Vehicle\_Miles\_Travelled\_bn | Vehicle travelled miles in billion |
| Fatalities\_Per\_100\_Million\_VMT | Number of fatalities per 100 million Vehicle Miles Travelled |
| Population | US population |
| Fatalities\_Per\_100000\_population | Fatalities per 100k population |

Table

Description automatically generated

Table

Description automatically generated

**Transformed Data Set for Power BI:**

R is used for transformation of data and Tablu or Power BI is used to build the dashboard. Transformed datasets are as follows:

AirlineSafety – Weighted incident, fatal accident and fatalities are calculated based on per billion ‘*available seat km per week*’. The table is in wide format.

AirlineSafetyLong – AirlineSafety data transposed to a long format for easy reporting.

AutoStats – Summary of auto fatalities for the given timeframe.

Fatalities – Fatalities for the same period for airlines and auto. This for comparison in the visual.

**Considerations:**

Following airlines are flagged as first world countries based on their coutry of core operations: *Air Canada, Air France, Air New Zealand, Alaska Airlines, Alitalia, All Nippon Airways, American, Austrian Airlines, British Airways, Delta / Northwest, Finnair, Hawaiian Airlines, Japan Airlines, KLM, Lufthansa, Qantas, Singapore Airlines, SWISS, TAP - Air Portugal, United / Continental, US Airways / America West, Virgin Atlantic*.

**Dashboard:**

Intention of the dashboard is to expose the truth behind the current negative buzz on air travel through historical facts and stats. From a design perspective, colorblind safe pallet has been used and simple high impact chart types have been used to maintain clarity.

1. **Fatalities by Airlines**

This scatter plot plots fatalities by airlines (color coded) for the two time periods (1985-1999, 2000-2014). The intention is to find correlation between the fatalities across the two periods.

There is no predictable nature of the number of incidents for the airlines between the two periods. Air Kenya had no fatalities in ’85-’99 period but has very high rate in the other period (’00-’14). Avianca on the other hand had high fatality during ’85-’99 but got better in the second half. China and Pakistan airlines has high rate in both halves. But basically, there is no pattern.

1. **Incidents by Airlines**

The intention is to find correlation between the incidents across the two periods.

We can see that there is a modest correlation between the two periods.

The two major outliers in the chart are Pakistan International Airlines and Ethiopian Airlines, which have had a persistently high rate of incidents. A third outlier, Russia’s Aeroflot, had an extraordinarily high number of reported incidents in ’85-’99. But many of these incidents are attempted hijackings around the time of the breakup of the Soviet Union. Contrastingly it has relatively low number in the ’00-’14 period. Even if we exclude these the outliers, there is still a positive correlation.

The statistical inference from this chart is, some airlines are slightly safer to fly than others.

1. **Top 5 airlines by Total Incidents**

This horizontal bar chart lists the top 5 airlines in terms of overall incidents. Noticeably, none of the 5 airlines are from the first world countries. Therefore, safety measure and technical deficit could be a reason behind the incidents.

The takeaway from this graph is, it cannot be generalized that air travel is unsafe. It might very well have a relationship between airlines and country of operations.

1. **Fatalities Trend: First World vs Rest**

This clustered graph compares the count of fatalities between first world countries and the rest of countries over the two periods. Although in both cases, the overall count decreased in ’00-’14 compared to ’85-’99, first world numbers are considerably lower than the rest in both cases. This chart re-established the fact that there is stark difference between air travel quality of countries. Air travel is not at all unsafe for some countries.

1. **Incidents Trend: First World vs Rest**

This chart is almost identical to the prior graph. The inference is also very similar. The only noteworthy point is, first world countries did not improve much in terms of incidents. The high air traffic could be a reason behind the high number of incidents in first world countries. But in comparison with the rest, the numbers are still pretty low.

1. **Motor Vehicle Deaths (US) vs Airlines Deaths (First World)**

This visual use supplemental data of auto fatalities in the USA for the same two periods (

’85 - ’99, ’00 - ‘14). The airlines fatalities are filtered for US airlines and compared to the auto numbers. Although there are decreasing trend of incidents in both travels, auto casualties are consistently far greater than air travel. If the facts in the data is to be considered as truth then this visual clearly shows that just by considering number of fatalities, travelling by road is far more dangerous than travelling by air.

Reference : Data Set

FiveThirtyEight, Airline Safety, Retrieved from

<https://github.com/fivethirtyeight/data/tree/master/airline-safety>

Motor vehicle fatality rate in U.S. by year, Retrieved from

<https://en.wikipedia.org/wiki/Motor_vehicle_fatality_rate_in_U.S._by_year>