A statistical analysis of aviation accidents

Carla Feriche

Contents

Introduction

Source of data

Domain research

Methodology

Results

Conclusions

Introduction

This report provides an analysis of aviation accidents, with a review of the history and an analysis of the different operators of the world. Furthermore, we will see an evaluation of the accidents for over 70 years, divided by the phases of a flight and an analysis of forecasted aviation trends will also provide.

The goal of this report of aviation accidents is to help to understand how safe the industry is. I decided to choose that topic, because I love to travel, I usually take many planes and I wanted to check if it is really a safe transport.

Source of Data

The accidents data has been extracted from the official web "Aviation Safety Network", which aim is to provide everyone, complete and reliable authoritative information on airliner accidents and safety issues.

Domain Research

Apart from trying to answer a few questions, I wanted to see if I found some interesting insights.

It is said that one of the safest transports that exist is the airline flights. Mainly, the reason for the study was to verify this statement.

In order to answer this, I started looking for databases from which to extract the information. When I had a good database, I started to familiarize myself with it, understanding it, cleaning it and therefore manipulating it.

Before and during the whole process of the report, I asked myself some questions that could be summarized in the following:

- Have accidents decreased over the years? (depending on the number of flights)
- What has been the rate of change over the last 10 years?
- Which airlines are the most dangerous? Which ones have had more accidents in the last 10 years, and which ones have had more fatalities? What countries are these airlines from?
- What are the forecast accidents for the next two years?
- Have been more accidents on airlines from less developed countries? And fatalities?
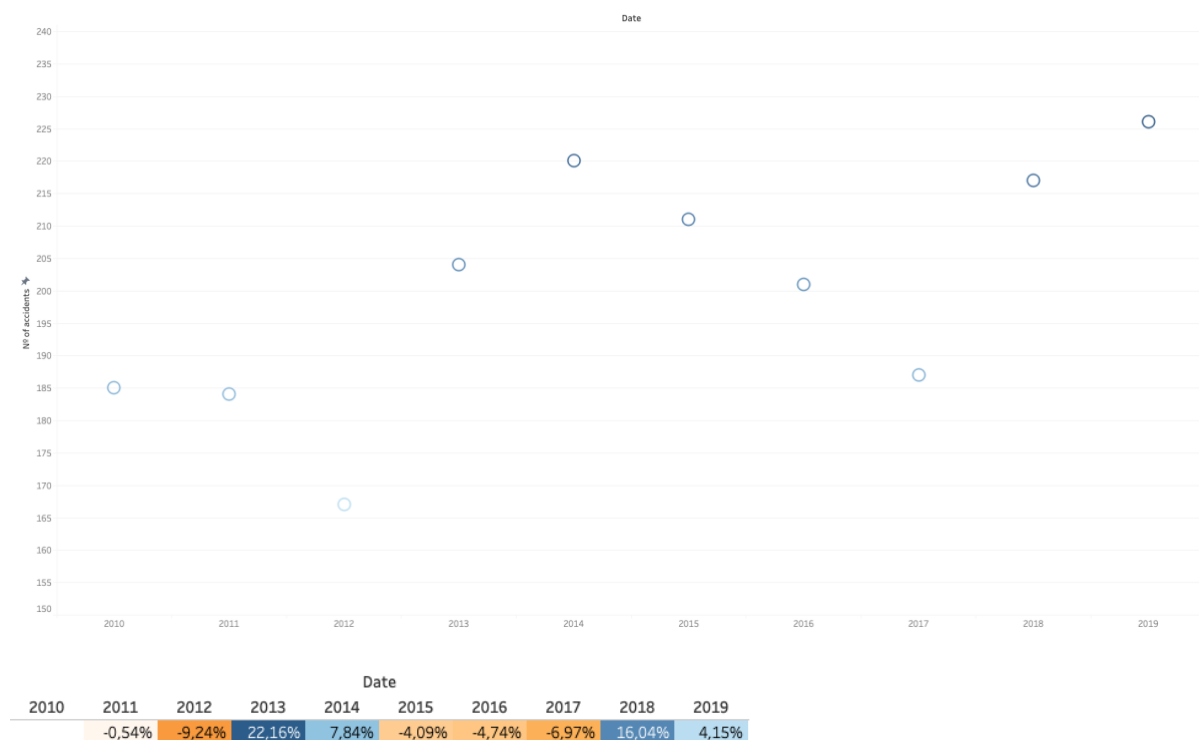- Are there more landing accidents than take-off accidents?

Methodology

To do the analysis, I decided to use Tableau program, as I believed it was the best way to visualize the data and understand it.
For the statistics part, I used the python module, Statistical functions (scipy.stats).

Results

To start familiarizing with the data, let's see how many accidents by year have occurred in the last 10 years around the world and the annual variance by year.


2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
	-0,54%	-9,24%	22,16%	7,84%	-4,09%	-4,74%	-6,97%	16,04%	4,15%

Except for the years 2012 and 2017, accidents seem to increase over the years.

Evolution of the number of flights & accidents

From this graphics we can see that trend actually goes down which was maybe not so obvious from the graphic with number of accidents only. Even so the number of accidents and fatalities is increasing, the number of flights is also increasing. And we could actually see that the ratio of fatalities/total amount of passengers trending down.

Total numbers of air passengers by year

Ratio accidents by 100.000 flights

Total number of accidents by year
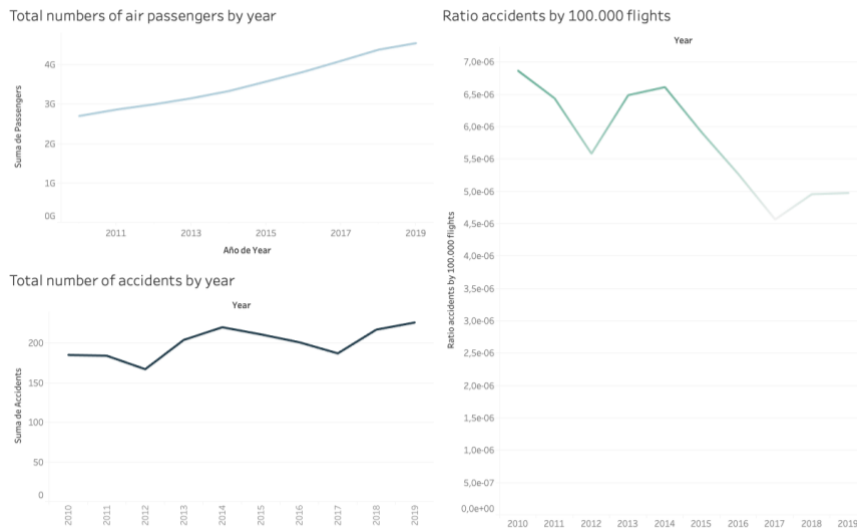
If we calculate a linear regression, we can see there is a correlation between the nº of accidents and the nº of passengers. That means that more flights will mean more accidents unless we work to decrease the accident rate.
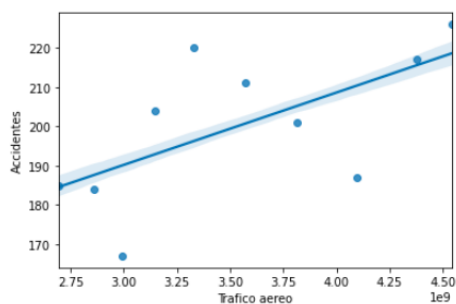
In the last 10 years, the industry-wide accident rate has been decreasing while in the same period, traffic has been increased. This shows that investments in safety bear fruit, safety is enhanced, and accidents are largely prevented from happening.
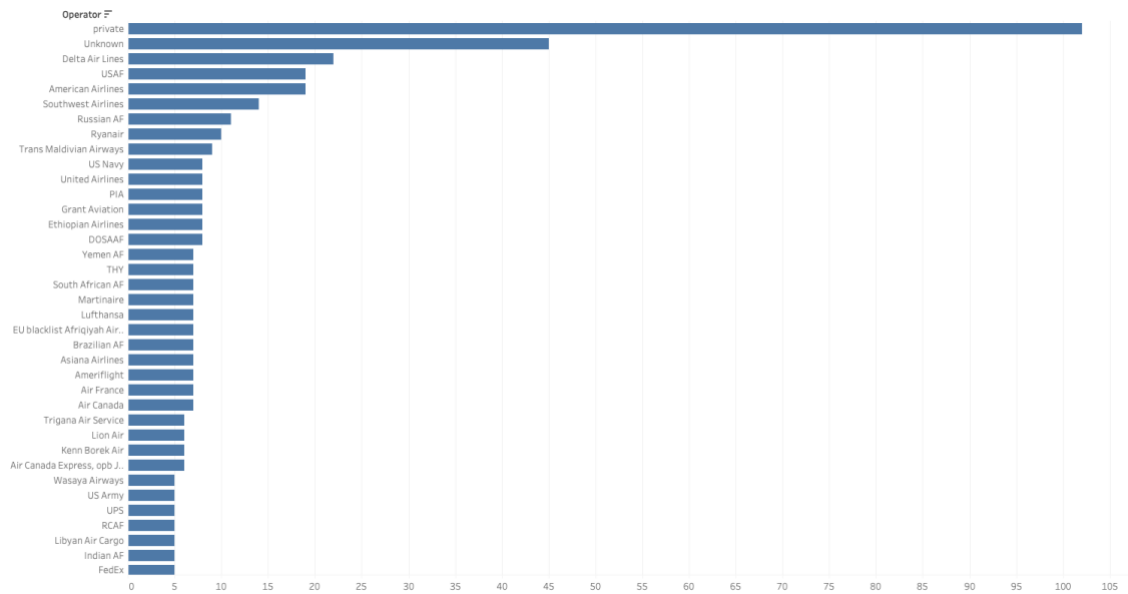
Forecast accidents

If we calculate the world annual forecast accidents, from the next 2 years, will be something like what we can see in the next graphic. The nº of accidents will be between 200 and 220.
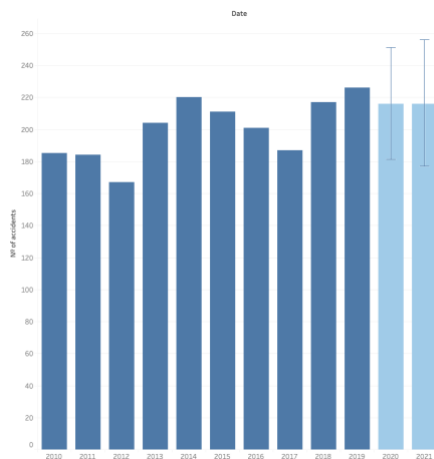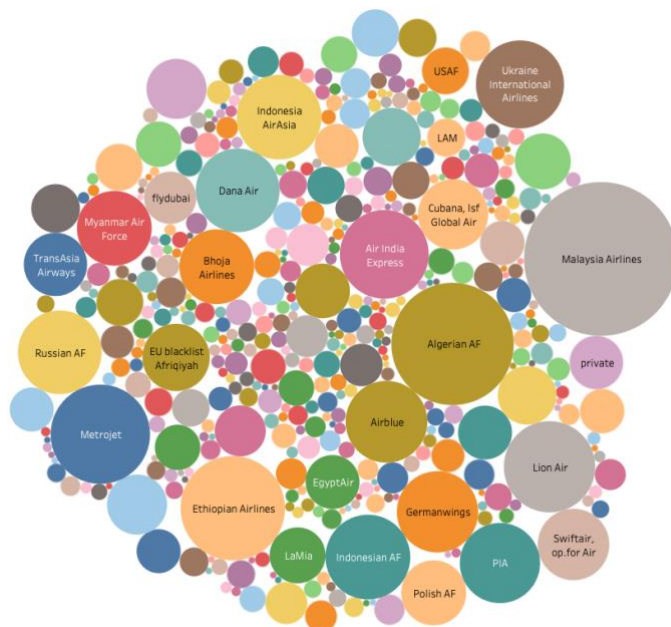
Distribution of accidents by operators

Now, let's find which are the operators with more accidents.

Distribution of fatalities by operators

And here, we can see the operators with more fatalities. Algerian AF and Malasyia Airlines will be the ones that are the top of the list.

On the map, we can visualize from where the airlines with the most fatalities in the world over the last 10 years are from.

Now, let's focus on two commercial airlines. First, we have the commonly known Delta Air Lines and second, Ethiopia Airlines. We can conclude that even though there are more accidents in Delta, there are 0 fatalities. In the other hand, in Ethiopia, has been less accidents, but we have more fatalities.

However, we cannot make decisions about any Operator like "which airline is much safer to flight with" without knowledge of total amount flights. If Delta has the largest number of accidents this doesn't mean that it is not worse to flight with because it might have the largest number of flights.

| | Operator | |
Year of Date	Delta Air Lines	Ethiopia n Airlin..
2010	0	90
2011	0	
2012	0	
2013	0	0
2014	0	0
2015	0	
2016	0	0
2017	0	
2018	0	0
2019	0	157
Grand Total	0	247

| | Operator | |
Year of Date	Delta Air Lines	Ethiopia n Airlin..
2010	1	1
2011	4	
2012	1	
2013	3	1
2014	2	1
2015	2	
2016	1	1
2017	3	
2018	3	1
2019	2	2
Grand Total	22	7

Most of the accidents over the past 20 years happened during "En route" and "Landing" phases.

It is not a surprise that the largest percentages accidents are seen to occur during en route and landing, as are highly complex flight phases which place significant demands on the crew that can lead to accidents.

To calculate this, we use the number of accidents per year. The data include corporate jet and military transport accidents, but just fatal accidents and fatalities per year for airliner (14+ passengers) accidents.
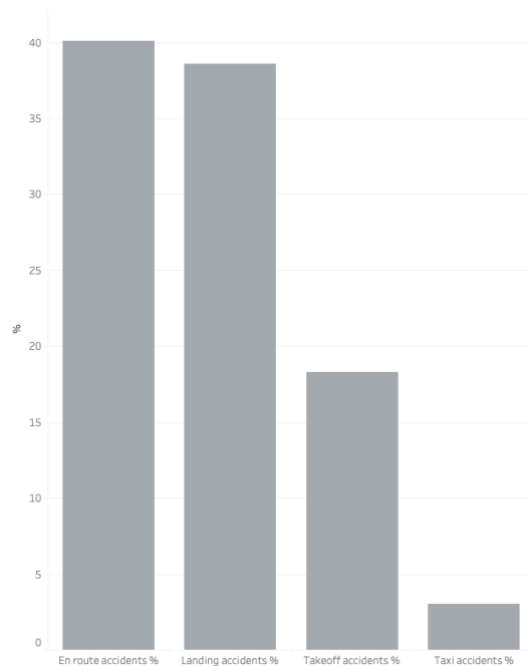
Definitions of flight phases

Taxi: this phase includes both taxi-out and taxi-in. Taxi-out starts when the aircraft begins moving forward under its own power and ends when it reaches the takeoff position. Taxi-in normally starts after the landing roll-out, when the aircraft taxis to the parking area. It may, in some cases, follow a taxi-out.

Landing: this phase begins when the aircraft is in the landing configuration and the crew is dedicated to land on a particular runway. It ends when the aircraft's speed is decreased to taxi speed.

Takeoff: this phase begins when the crew increases thrust for the purpose of lift-off.

En route: this phase begins when the aircraft reaches the initial cruise altitude. It ends when the crew initiates a descent for the purpose of landing.

In order to see if there are significance differences between the fight phases over the years, we are going to calculate an inferential statistic, t-test, to determine if there is a significant difference between the means. A paired simple t-test is the method we are going to use as we are comparing dependent samples from the same group at different times.

After running the *stats.ttest_rel* in Python, we can conclude that, with all the possible combination between phases, our p-value returned has been lower than 0.05 except for the "En route" vs "landing". So, we reject the null hypothesis in all the case except the last one, and we argue that are significantly different the means between each other.
In the case of "En route" vs "Landing", we state that we cannot reject the hypothesis. The phases are not significantly different.
As we have seen before, "En route" and "landing" are the phases where occur more accidents from the last 20 years.

We also wanted to know if there was significance difference between the year 2000 and the year 2019. To do so, we calculate a t-statistic for the independents sample, and we come to de conclusion that there is significance difference between both years as the p-value was lower than 0.05.

Conclusions

This analysis clearly shows that accidents have decreased considerably over the last decade, probably due to technological improvements. Furthermore, we can conclude, that the nº of accidents compare with the nº of flights in a year is very small. So, that means it's a very safe industry, which on top of that, is getting better every year.

Likewise, we can also reach the conclusion that the airlines that have had more fatalities are those from less developed countries, which may not have the most advanced technologies.

On the one hand, in relation to the phases of the accidents, most accidents occur in the "En route" and "Landing" phases, which, through statistical analysis, we cannot conclude that the phases are significantly different.

On the other hand, we have seen that between the year 2000 and 2019, there have been no changes between them, that is, at present, there is still more or less the same percentage of accidents in each phase as in the year 2000.

In the next steps of the work, I would like to find the number of flights per airline and be able to calculate the accident ratio taking into account the number of flights. Also, I would like to study the different transports that exist, to be able to affirm that the airplane is the safest transport that we have.