Analysis and Visualization of Airplane Crashes

Preeti Khatri
1259
Department of Applied Data Science
San Jose State University
San Jose, US

Kaamya Ravikumar
0138
Department of Applied Data Science
San Jose State University
San Jose, US

Jie Dong
0119
Department of Applied Data Science
San Jose State University
San Jose, US

Abstract:* 2017 was considered to be the safest year in aviation with zero deaths from commercial planes and very minimal onboard fatalities attributed to the aviation industry as a whole even when military and cargo flights are factored in. Despite ever increasing demand for flight and the number of passengers choosing to fly accidents have decreased dramatically over the years. Before we talk about how crashes have reduced over the years, the first question to ask ourselves would be "why do crashes occur in the first place"? Aviation industry has undergone several advancements over the decades and owing to the several safety practices by the NTSB and the increased number of trained professional pilots, air travel has become a relatively safer option. According to NTSB, the accident rate for commercial planes is 0.07 per 100,000 flying hours. Though Air travel has improved over the years and the number of fatalities has gone down compared to the initial period, airplane accidents are still a cause of concern. Even if the probability of occurrence is lesser, once an incident occurs it will mostly end as a fatal one. With this project we aim to visualize the several causes and factors causing airplane accidents. The dataset from Kaggle has 2 MB of data with 17 features and more than five thousand records. Since the dataset contains null values, it would be pre-processed to remove those values and format the remaining columns into the desired format. This data cleaning process will be performed using Tableau and Talend. As part of the analysis, we would be using Tableau Software to visualize the prepared data in the form of graphs and tables and integrate all the visualizations into a single dashboard to present it as an entire story.

Keywords -- Tableau, ETL, airplane crashes, visualization

I. Introduction

Going back to the early decades of flying, an era dominated by mail planes and barnstormers, flying was considered a high risk affair. Flying was cheaper and timely delivery was chosen over flight crew and passenger safety. Plane technology was in its initial stages and crashes were more frequent. Statistics say that an airmail pilot had 1 in 4 chances in meeting their end at the control of the aircraft. The history of aircraft is defined by trial and error learning and in the quest to understand and minimize future occurrences, several organizations are constantly on the lookout for statistical data that could aid in gaining a holistic evaluation of situations. The growth and the technological advancements in the aviation industry over the past few decades have led to the subsequent increase in the number of aircrafts being operated. The production of larger and more efficient airplanes which have been integrated with advanced technologies has made this possible. The increase in the operation of aircrafts has also led to the increase in the number of airplane accidents. According to the data from the U.S National Transportation Safety Board(NTSB), 393 people were killed in civil airplane accidents in the US which was a 13% increase compared to the previous year where 347 were killed. The fatal accident rate in general aviation was calculated to be 1.029 per 100,000 flight hours in comparison to the 0.935 per 100,000 flight hours in 2017. The causes and in-depth analysis of what could have led to the fatalities have been researched by several groups over the years.

With this project we aim to analyze, visualize and identify underlying trends and patterns in airplane accidents. Identifying the most common causes of accidents, to understand if there are any specific airline carriers that have been more susceptible to fatal accidents over the years, recovering any additional patterns/trends in the air crashes that could be used to improve air safety and limit the fatalities are some of the goals of this project. The target audience for this kind of analysis and the dashboards developed would benefit independent government agencies like NTSB (National Transportation Safety Board), FAA (Federal Aviation Administration) who investigates every airline crash that happens within the United States and internationally as well and Aviation Industry Experts and Flight Carriers to create insights and awareness.

For this purpose, we have used the dataset from Kaggle. The overall dataset contains data about civil and commercial aviation accidents of scheduled and non-scheduled passenger airliners worldwide, which resulted in a fatality. All cargo, positioning, ferry and test flight fatal accidents, military transport accidents with 10 or more fatalities, commercial and military helicopter accidents with greater than 10 fatalities, civil and military airship accidents involving fatalities, aviation accidents involving the death of famous people, aviation accidents or incidents of noteworthy interest.

II. Related Work

The following research papers were referred to understand the research work conducted previously to analyze several causes and factors influencing airplane crashes. Authors G. Li, S.P. Baker, J.G. Grabowski, G.W. Rebok has performed an analysis on the characteristics of the pilot and the circumstances under which aviation accidents have occurred due to pilot error. Data was integrated from several source files provided by the NTSB that included 329 airline crashes, 1627 commuter and taxi crashes, and 27935 general aviation crashes over the years. Logistic Regression modeling was used to analyze the extent to which individual features influenced the presence of pilot errors. On completion of their research, they concluded that weather conditions and on-airport location increased the probability of pilot errors and the chances of pilot errors decreased with the pilot's certificate rating.

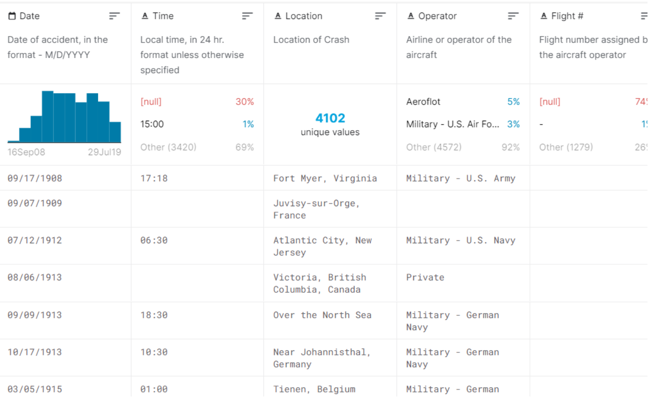
John A. Caldwell performed a study on the impact of pilot fatigue in aviation performance. As mentioned in his study [2] "A report from an NTSB study of major accidents in domestic air carriers stated that "crews comprising captains and first officers whose time since awakening was above the median for their crew position made more errors overall, and significantly more procedural and tactical decision errors" (National Transportation Safety Board, 1994, p. 75)". From the perspective of pilot performance, as fatigue grows, it leads to decreased performance by the pilots, the ability of pilots to put together individual data from separate instruments into an overall picture decreases. It has also been mentioned that sleep deprivation could cause the pilot to have momentary illusions due to involuntary lapses into sleep.

Authors D.Kerfoot, M. Hoffman has attempted to implement a CRISP-DM methodology for identification and visualization of fatal airplane accidents using a classification model. The attributes used for analysis include attributes of the aircraft, geographical attributes, purpose of flight, to name a few. The CRSIP_DM methodology includes exploration of data, data pre-processing, data modeling and evaluation of the model. Data was modelled using several classifier models like K-NN, Naive-Bayes etc, and the model with the highest accuracy was concluded as Deep-Learning with an accuracy of 63.97%.

III. Methods

A. Airplane Crashes Dataset

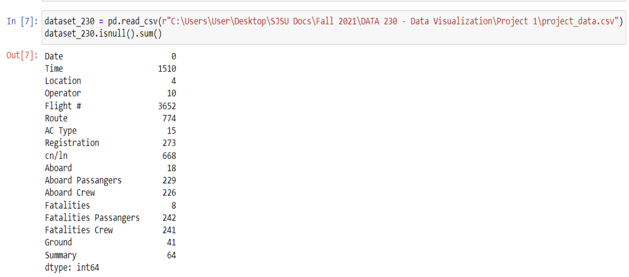
Identifying the right dataset is the preliminary and the most important step in Exploratory data analysis. Once the right dataset has been identified, data preparation, modelling and eventual analysis of data follows. Higher the number of features and size of the dataset, the better the final analysis and visualization. For our project we have chosen the Airplane Crashes Dataset [4] from Kaggle. This data was scraped from planecrashinfo.com and contains 2MB of data with 17 attributes and more than 5k observations. Below is the original dataset from kaggle and the columns that have been used in the analysis along with descriptions:



Date	Time	Location	Operator	Flight #
Date of accident, in the format - M/D/YYYY	Local time, in 24 hr. format unless otherwise specified	Location of Crash	Airline or operator of the aircraft	Flight number assigned b the aircraft operator
16Sep08 29Jul19	[null] 30% 15:00 1% Other (3420) 69%	4102 unique values	Aeroflot 5% Military - U.S. Air Fo... 3% Other (4572) 92%	[null] 74 - 1 Other (1279) 26
09/17/1908	17:18	Fort Myer, Virginia	Military - U.S. Army	
09/07/1909		Juvisy-sur-Orge, France		
07/12/1912	06:30	Atlantic City, New Jersey	Military - U.S. Navy	
08/06/1913		Victoria, British Columbia, Canada	Private	
09/09/1913	18:30	Over the North Sea	Military - German Navy	
10/17/1913	10:30	Near Johannisthal, Germany	Military - German Navy	
03/05/1915	01:00	Tienen, Belgium	Military - German	

	Description
Date	Date of accident, in the format - January 01, 2001
Location	Airline or operator of the aircraft
Operator	Category of flight operated (Military/Cargo/Passenger)
Route	Complete or partial route flown prior to the accident
Type	Aircraft type
Aboard	Total aboard (passengers / crew)
Fatalities	Total fatalities aboard (passengers / crew)
Summary	Brief description of the accident

The actual dataset consists of null values which would be handled in the data preparation stage:



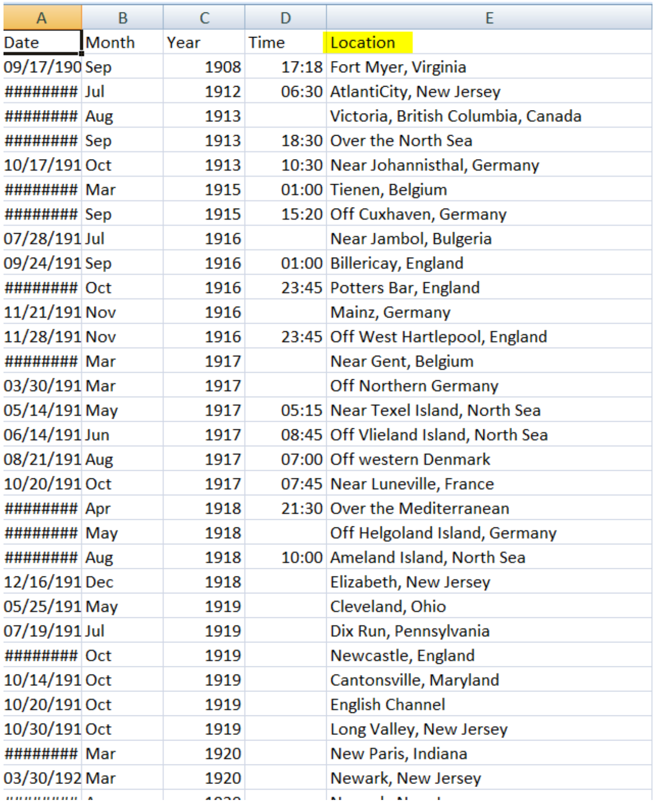
```
In [7]: dataset_230 = pd.read_csv(r"C:\Users\User\Desktop\SJSU Docs\Fall 2021\DATA 230 - Data Visualization\Project 1\project_data.csv")
        dataset_230.isnull().sum()

Out[7]: Date                     0
        Time                  1510
        Location                 4
        Operator                10
        Flight #              3652
        Route                  774
        AC Type                 15
        Registration           273
        cn/ln                  668
        Aboard                  18
        Aboard Passangers      229
        Aboard Crew            226
        Fatalities               8
        Fatalities Passangers  242
        Fatalities Crew        241
        Ground                  41
        Summary                 64
        dtype: int64
```

B. *Data Cleaning*

We used Talend for Data Cleaning and created extra columns for Analysis and Visualization purposes.The first step was to identify and handle the null values present in the dataset. It was decided that the columns that had a higher percentage of null values were handled. Since the distribution of null values in our dataset was skewed, we skipped those rows that had null values. The next step was to format the column values into a specified form. In our dataset, the date column was not in consistent format, so we made it in a single format and extracted Year and Month and included that as additional columns in the dataset. For the purpose of analysis and visualization, based on Aircraft type, we have created an additional column named Operator which we divided into 5 categories:

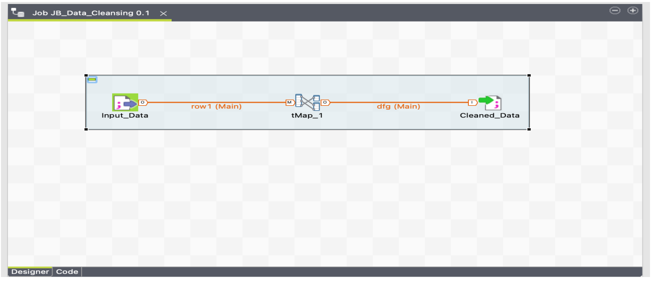
- Military
- Passenger Jets
- Private Jets
- Cargo Jets
- Mail Services Jets

The original data had a column named Summary that had reasons for the crash written in one or two sentences. For analysis purposes, we have also created a column that categorizes the crashes under its respective causes based on the Summary columns of the dataset using some keywords from the original data. The column location had location of the crash presented in an inconsistent format as seen below:



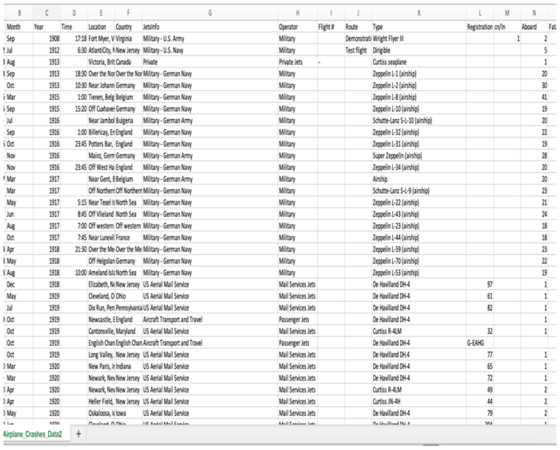
A	B	C	D	E
Date	Month	Year	Time	Location
09/17/190	Sep	1908	17:18	Fort Myer, Virginia
########	Jul	1912	06:30	AtlantiCity, New Jersey
########	Aug	1913		Victoria, British Columbia, Canada
########	Sep	1913	18:30	Over the North Sea
10/17/191	Oct	1913	10:30	Near Johannisthal, Germany
########	Mar	1915	01:00	Tienen, Belgium
########	Sep	1915	15:20	Off Cuxhaven, Germany
07/28/191	Jul	1916		Near Jambol, Bulgeria
09/24/191	Sep	1916	01:00	Billericay, England
########	Oct	1916	23:45	Potters Bar, England
11/21/191	Nov	1916		Mainz, Germany
11/28/191	Nov	1916	23:45	Off West Hartlepool, England
########	Mar	1917		Near Gent, Belgium
03/30/191	Mar	1917		Off Northern Germany
05/14/191	May	1917	05:15	Near Texel Island, North Sea
06/14/191	Jun	1917	08:45	Off Vlieland Island, North Sea
08/21/191	Aug	1917	07:00	Off western Denmark
10/20/191	Oct	1917	07:45	Near Luneville, France
########	Apr	1918	21:30	Over the Mediterranean
########	May	1918		Off Helgoland Island, Germany
########	Aug	1918	10:00	Ameland Island, North Sea
12/16/191	Dec	1918		Elizabeth, New Jersey
05/25/191	May	1919		Cleveland, Ohio
07/19/191	Jul	1919		Dix Run, Pennsylvania
########	Oct	1919		Newcastle, England
10/14/191	Oct	1919		Cantonsville, Maryland
10/20/191	Oct	1919		English Channel
10/30/191	Oct	1919		Long Valley, New Jersey
########	Mar	1920		New Paris, Indiana
03/30/192	Mar	1920		Newark, New Jersey

Some entries had the name of the country, some were present as city, state and country or just as state and country. For the purpose of creating map visualization we converted this column into a uniform format with the country name alone.



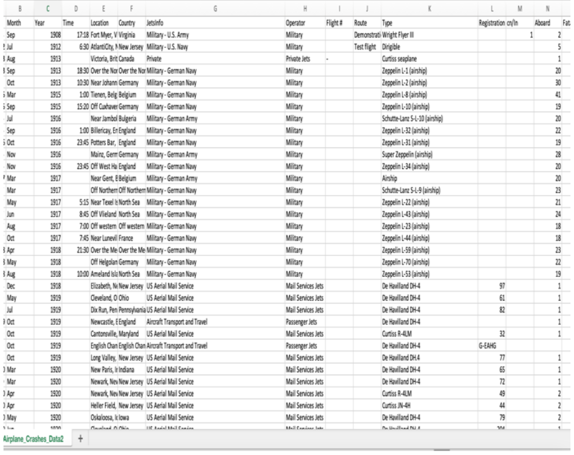
C. *Comparison of datasets before and after cleaning:*

This is an example of how the dataset was originally and after cleaning and preparation along with the addition of new columns for visualization.



Month	Year	Time	Location	Country	JetsInfo	Operator	Flight #	Route	Type	Registration	cn/In	Aboard	Fat
Sep	1908	17:18	Fort Myer, V	Virginia	Military - U.S. Army	Military		Demonstrati	Wright Flyer III		1	2	
Jul	1912	6:30	AtlantiCity, N	New Jersey	Military - U.S. Navy	Military		Test flight	Dirigible			5	
Aug	1913		Victoria, Brit	Canada	Private	Private Jets	-		Curtiss seaplane			1	
Sep	1913	18:30	Over the Nor	Over the Nor	Military - German Navy	Military			Zeppelin L-1 (airship)			20	
Oct	1913	10:30	Near Johann	Germany	Military - German Navy	Military			Zeppelin L-2 (airship)			30	
Mar	1915	1:00	Tienen, Belg	Belgium	Military - German Navy	Military			Zeppelin L-8 (airship)			41	
Sep	1915	15:20	Off Cuxhaver	Germany	Military - German Navy	Military			Zeppelin L-10 (airship)			19	
Jul	1916		Near Jambol	Bulgeria	Military - German Army	Military			Schutte-Lanz S-L-10 (airship)			20	
Sep	1916	1:00	Billericay, En	England	Military - German Navy	Military			Zeppelin L-32 (airship)			22	
Oct	1916	23:45	Potters Bar,	England	Military - German Navy	Military			Zeppelin L-31 (airship)			19	
Nov	1916		Mainz, Germ	Germany	Military - German Army	Military			Super Zeppelin (airship)			28	
Nov	1916	23:45	Off West Ha	England	Military - German Navy	Military			Zeppelin L-34 (airship)			20	
Mar	1917		Near Gent, B	Belgium	Military - German Army	Military			Airship			20	
Mar	1917		Off Northern	Off Northern	Military - German Navy	Military			Schutte-Lanz S-L-9 (airship)			23	
May	1917	5:15	Near Texel Is	North Sea	Military - German Navy	Military			Zeppelin L-22 (airship)			21	
Jun	1917	8:45	Off Vlieland	North Sea	Military - German Navy	Military			Zeppelin L-43 (airship)			24	
Aug	1917	7:00	Off western	Off western	Military - German Navy	Military			Zeppelin L-23 (airship)			18	
Oct	1917	7:45	Near Lunevil	France	Military - German Navy	Military			Zeppelin L-44 (airship)			18	
Apr	1918	21:30	Over the Me	Over the Me	Military - German Navy	Military			Zeppelin L-59 (airship)			23	
May	1918		Off Helgolan	Germany	Military - German Navy	Military			Zeppelin L-70 (airship)			22	
Aug	1918	10:00	Ameland Isla	North Sea	Military - German Navy	Military			Zeppelin L-53 (airship)			19	
Dec	1918		Elizabeth, Ne	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	97		1	
May	1919		Cleveland, O	Ohio	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	61		1	
Jul	1919		Dix Run, Pen	Pennsylvania	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	82		1	
Oct	1919		Newcastle, E	England	Aircraft Transport and Travel	Passenger Jets			De Havilland DH-4			1	
Oct	1919		Cantonsville,	Maryland	US Aerial Mail Service	Mail Services Jets			Curtiss R-4LM	32		1	
Oct	1919		English Chan	English Chan	Aircraft Transport and Travel	Passenger Jets			De Havilland DH-4	G-EAHG			
Oct	1919		Long Valley,	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH.4	77		1	
Mar	1920		New Paris, I	Indiana	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	65		1	
Mar	1920		Newark, Nev	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	72		1	
Apr	1920		Newark, Nev	New Jersey	US Aerial Mail Service	Mail Services Jets			Curtiss R-4LM	49		2	
Apr	1920		Heller Field,	New Jersey	US Aerial Mail Service	Mail Services Jets			Curtiss JN-4H	44		2	
May	1920		Oskaloosa, Ic	Iowa	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	79		2	

Airplane_Crashes_Data2 +



Month	Year	Time	Location	Country	JetsInfo	Operator	Flight #	Route	Type	Registration	cn/In	Aboard	Fat
Sep	1908	17:18	Fort Myer, V	Virginia	Military - U.S. Army	Military		Demonstrati	Wright Flyer III		1	2	
Jul	1912	6:30	AtlantiCity, N	New Jersey	Military - U.S. Navy	Military		Test flight	Dirigible			5	
Aug	1913		Victoria, Brit	Canada	Private	Private Jets	-		Curtiss seaplane			1	
Sep	1913	18:30	Over the Nor	Over the Nor	Military - German Navy	Military			Zeppelin L-1 (airship)			20	
Oct	1913	10:30	Near Johann	Germany	Military - German Navy	Military			Zeppelin L-2 (airship)			30	
Mar	1915	1:00	Tienen, Belg	Belgium	Military - German Navy	Military			Zeppelin L-8 (airship)			41	
Sep	1915	15:20	Off Cuxhaver	Germany	Military - German Navy	Military			Zeppelin L-10 (airship)			19	
Jul	1916		Near Jambol	Bulgeria	Military - German Army	Military			Schutte-Lanz S-L-10 (airship)			20	
Sep	1916	1:00	Billericay, En	England	Military - German Navy	Military			Zeppelin L-32 (airship)			22	
Oct	1916	23:45	Potters Bar,	England	Military - German Navy	Military			Zeppelin L-31 (airship)			19	
Nov	1916		Mainz, Germ	Germany	Military - German Army	Military			Super Zeppelin (airship)			28	
Nov	1916	23:45	Off West Ha	England	Military - German Navy	Military			Zeppelin L-34 (airship)			20	
Mar	1917		Near Gent, B	Belgium	Military - German Army	Military			Airship			20	
Mar	1917		Off Northern	Off Northern	Military - German Navy	Military			Schutte-Lanz S-L-9 (airship)			23	
May	1917	5:15	Near Texel Is	North Sea	Military - German Navy	Military			Zeppelin L-22 (airship)			21	
Jun	1917	8:45	Off Vlieland	North Sea	Military - German Navy	Military			Zeppelin L-43 (airship)			24	
Aug	1917	7:00	Off western	Off western	Military - German Navy	Military			Zeppelin L-23 (airship)			18	
Oct	1917	7:45	Near Lunevil	France	Military - German Navy	Military			Zeppelin L-44 (airship)			18	
Apr	1918	21:30	Over the Me	Over the Me	Military - German Navy	Military			Zeppelin L-59 (airship)			23	
May	1918		Off Helgolan	Germany	Military - German Navy	Military			Zeppelin L-70 (airship)			22	
Aug	1918	10:00	Ameland Isla	North Sea	Military - German Navy	Military			Zeppelin L-53 (airship)			19	
Dec	1918		Elizabeth, Ne	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	97		1	
May	1919		Cleveland, O	Ohio	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	61		1	
Jul	1919		Dix Run, Pen	Pennsylvania	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	82		1	
Oct	1919		Newcastle, E	England	Aircraft Transport and Travel	Passenger Jets			De Havilland DH-4			1	
Oct	1919		Cantonsville,	Maryland	US Aerial Mail Service	Mail Services Jets			Curtiss R-4LM	32		1	
Oct	1919		English Chan	English Chan	Aircraft Transport and Travel	Passenger Jets			De Havilland DH-4	G-EAHG			
Oct	1919		Long Valley,	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH.4	77		1	
Mar	1920		New Paris, I	Indiana	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	65		1	
Mar	1920		Newark, Nev	New Jersey	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	72		1	
Apr	1920		Newark, Nev	New Jersey	US Aerial Mail Service	Mail Services Jets			Curtiss R-4LM	49		2	
Apr	1920		Heller Field,	New Jersey	US Aerial Mail Service	Mail Services Jets			Curtiss JN-4H	44		2	
May	1920		Oskaloosa, Ic	Iowa	US Aerial Mail Service	Mail Services Jets			De Havilland DH-4	79		2	

Airplane_Crashes_Data2 +

D. *Tableau Imports*

This is the screenshot of the final data that we imported into Tableau for visualization. Not all the columns mentioned in data description have been included in the final import as we have kept only those columns that were used for creating the final narrative.

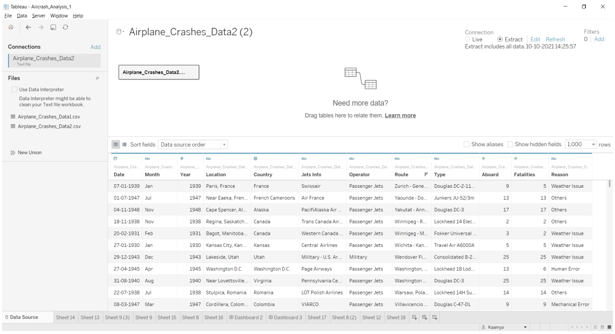


Tableau - Aircrash_Analysis_1
File Data Server Window Help
Connections Add
Airplane_Crashes_Data2
Text file
Files
Airplane_Crashes_Data2 (2)
Need more data?
Drag tables here to relate them. Learn more
New Union
Data Source
Sheet 14 Sheet 13 Sheet 9 (3) Sheet 9 Sheet 15 Sheet 8 Sheet 16 Dashboard 2 Dashboard 3 Sheet 17 Sheet 8 (2) Sheet 12 Sheet 18
Kaamya

□ Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

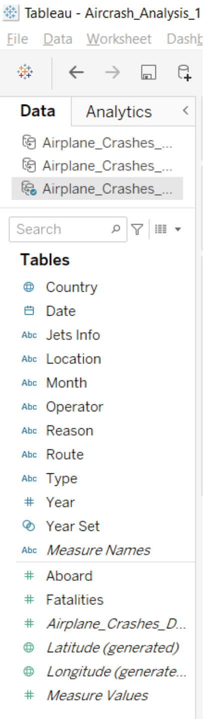
Airplane_Crashes_Data1.csv
Airplane_Crashes_Data2.csv

Airplane_Crashes_Data2....

Connection ○ Live ● Extract Edit Refresh
Extract includes all data. 10-10-2021 14:25:57
Filters 0 Add

Sort fields Data source order
□ Show aliases □ Show hidden fields 1,000 rows

Date	Month	Year	Location	Country	Jets Info	Operator	Route	Type	Aboard	Fatalities	Reason
07-01-1939	Jan	1939	Paris, France	France	Swissair	Passenger Jets	Zurich - Gene...	Douglas DC-2-11...	9	5	Weather Issue
01-07-1947	Jul	1947	Near Eaeka, Fren...	French Cameroons	Air France	Passenger Jets	Yaounde - Do...	Junkers JU-52/3m	13	13	Others
04-11-1948	Nov	1948	Cape Spencer, Al...	Alaska	PacifiAlaska Air ...	Passenger Jets	Yakutat - Ann...	Douglas DC-3	17	17	Others
18-11-1938	Nov	1938	Regina, Saskatch...	Canada	Trans Canada Air...	Passenger Jets	Winnipeg - R...	Lockheed 14 Elec...	2	2	Others
20-02-1931	Feb	1931	Bagot, Manitoba...	Canada	Western Canada ...	Passenger Jets	Winnigeg - M...	Fokker Universal ...	3	2	Weather Issue
27-01-1930	Jan	1930	Kansas City, Kan...	Kansas	Central Airlines	Passenger Jets	Wichita - Kan...	Travel Air A6000A	5	5	Weather Issue
29-12-1943	Dec	1943	Lakeside, Utah	Utah	Military - U.S. Ar...	Military	Wendover Fi...	Consolidated B-2...	25	25	Weather Issue
27-04-1945	Apr	1945	Washington D.C.	Washington D.C.	Page Airways	Passenger Jets	Washington, ...	Lockheed 18 Lod...	13	6	Human Error
31-08-1940	Aug	1940	Near Lovettsville...	Virginia	Pennsylvania Ce...	Passenger Jets	Washington ...	Douglas DC-3	25	25	Weather Issue
22-07-1938	Jul	1938	Stulpica, Romania	Romania	LOT Polish Airlines	Passenger Jets	Warsaw, Pola...	Lockheed 14H Su...	14	14	Others
08-03-1947	Mar	1947	Cordillera, Colom...	Colombia	VIARCO	Passenger Jets	Villavicencio ...	Douglas C-47-DL	9	9	Mechanical Error



Airplane_Crashes_...
Airplane_Crashes_...
Airplane_Crashes_...

Tables

- Country
- Date
- Abc Jets Info
- Abc Location
- Abc Month
- Abc Operator
- Abc Reason
- Abc Route
- Abc Type
- # Year
- Year Set
- Abc *Measure Names*
- # Aboard
- # Fatalities
- # *Airplane_Crashes_D...*
- *Latitude (generated)*
- *Longitude (generate...*
- # *Measure Values*

IV. Results

Our end result is a dashboard that analyzes and narrates the overall trend and factors involved in airplane accidents over the years.

A. *Visualization 1: Accident Trend in the Aviation Industry over the years*

The figure below gives an overall insight on how the airplane industry as a whole has evolved. We could imply that with the technological advancements in aviation in addition to better quality of pilot training, accidents have seen a downward trend over the years. This was created by using a line graph which takes the sum of fatalities along the y-axis and the year along the x-axis.

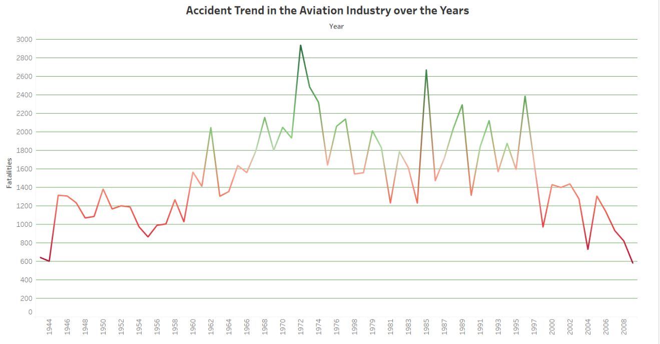
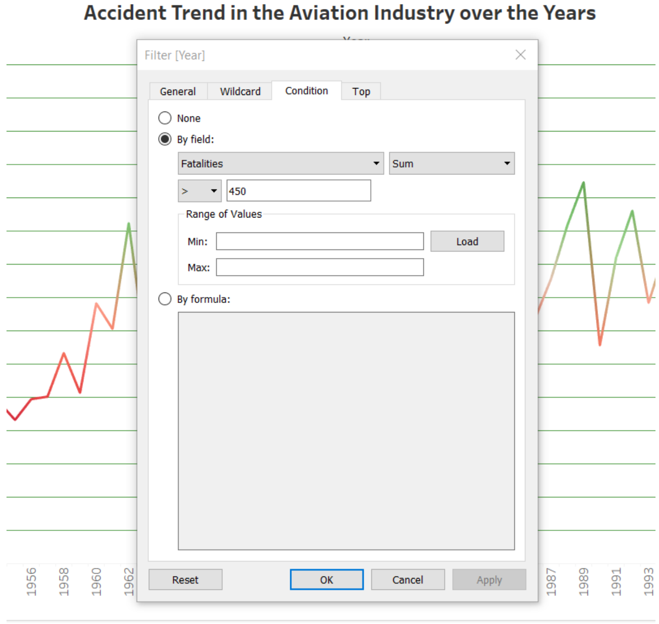


Figure 1 Accident Trend in the Aviation Industry over the years

The goal of this visualization is to show the audience the overall trend of accidents in the aviation industry. The initial data points in our dataset (years 1908-1930) had relatively less accidents which could possibly be due to less popularity of air traveland the more recent data (2009-present) had decreased accidents due to better technologies in aviation. Since these points were not contributing in creating a view of producing the growth and fall of fatalities we have excluded these by using the filter feature in Tableau.



Filter [Year]

General | Wildcard | Condition | Top

○ None
● By field:
Fatalities ▾ Sum ▾
> ▾ 450
Range of Values
Min:
Max:
○ By formula:

B. *Visualization 2: Leading causes of Air Crash*

This visualization was created to show those that were the most common causes for the accidents to occur. As an obvious result, human errors are seen as the most common cause for accidents followed by weather issues, mechanical errors and other categories. The other category includes accidents that were caused during situations like practice flights of military carriers, etc. Since the target group for our dashboard are organizations like NTSB and FAA we have not limited our analysis to only passenger carriers. The column has values of cause and year and the row has the fatality count.

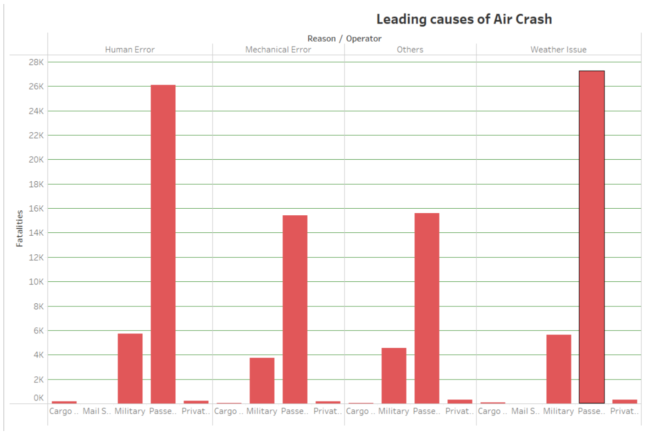


Figure 2 Leading causes of Air Crash

C. *Visualization 3: Aboard vs Fatality Ratio*

What are the odds of passengers surviving a plane crash? There could be many circumstances and situations that could possibly answer this question. The dataset that we have shows the odds of surviving a plane crash is relatively less. For this, a graph that shows the aboard vs fatality ratio has been designed.

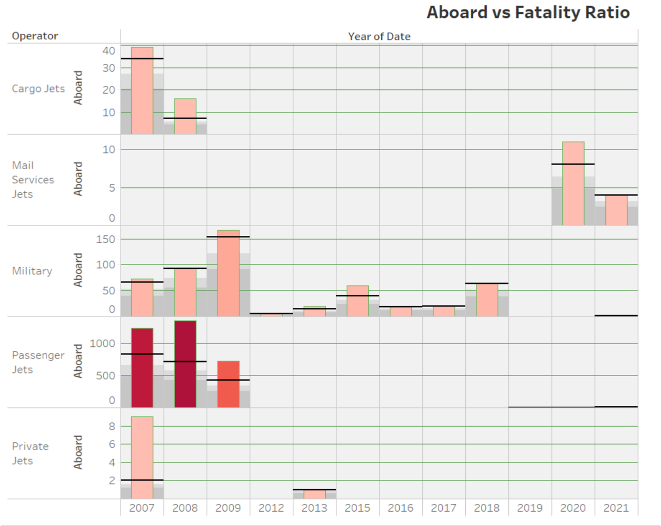


Figure 3 Aboard vs Fatality Ratio

D. *Visualization 4: Fatalities by Operator Type*

Our dataset has classified every airline into 5 categories as mentioned in the previous section. Another question that could be answered is the category of the flight that has been affected the most. The visualization provides a comparison of the fatalities caused by each category over the years. As expected, passenger flights are the most impacted. This could be attributed to the fact that the ratio of passenger flights is higher than compared to other categories present. Hence more the number of fatalities. This has again decreased over the years due to advancements as we saw earlier. This has the year and sum of fatalities as the column and row value. The trend has been differentiated by operator type.

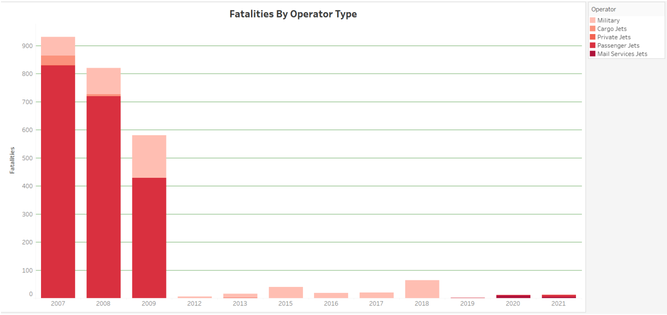


Figure 4 Fatalities by Operator Type

E. Visualization 5: Fatalities by Airline

Every airline has its own safety practices and training methods which they have adopted over the years to improve performance and build customer's trust. As a committee and for individual airlines, an analysis that could give a collective view of the airline that has had the most fatalities would be beneficial for their self-improvement. The target point of the below visualization is to address that and shows the airline that has been accounted for maximum fatalities. This is designed using a text cloud visualization and the count of fatalities is seen by hovering over the airline names.

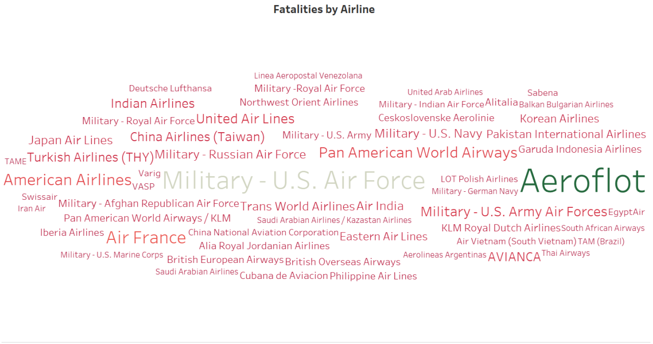


Figure 5 Fatalities by Airline

F. Visualization 6: Fatalities by Country

Let's now move on to look at the countries that have seen the highest fatalities over the years. A map visualization was used to implement this that comes along with a filter option to see the fatalities each year.

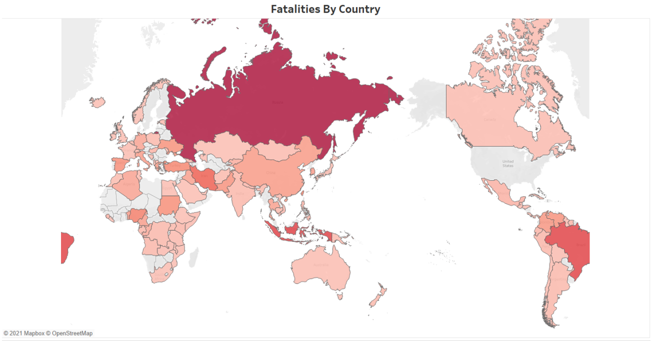


Figure 6 Fatalities by Country

G. Visualization 7: Fatalities by Route

Another question that could complement the above visualization would be to identify the routes that have the maximum fatalities. A tree map analysis has been created for this purpose to view the route and number of fatalities that have been recorded as well.

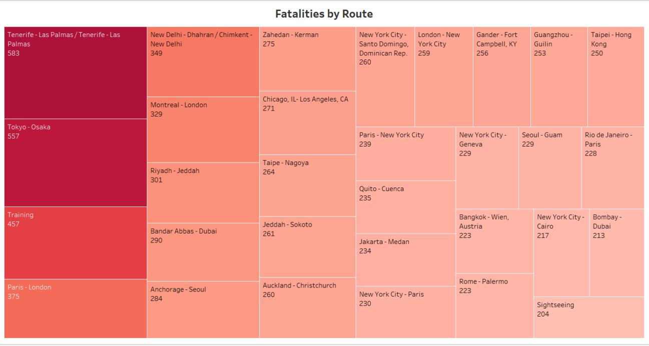


Figure 7 Fatalities by Route

H. Dashboard View:

The final dashboard has been created as a summary of all the visualizations that have been presented above. The initial part of the dashboard has been grouped with an aim of addressing the initial trend observed in the aviation industry.

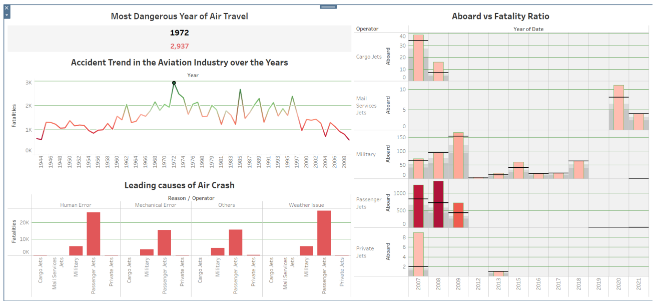


Figure 8 Trend observed in the aviation industry

The dashboards have been built in an interactive format where applying filters for one chart would format the entire dashboard to provide corresponding results. The below image shows the trend analysis for passenger jets in all the components of the dashboard.

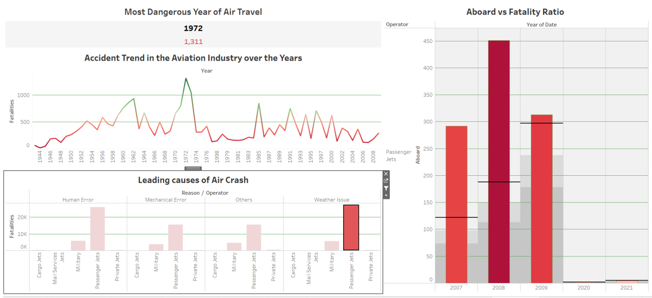


Figure 9 Airline specific and geographic analysis

The second part of the dashboard has been grouped with an aim of answering the airline specific and geographic analysis questions. The upper part of the dashboard starts with identifying the operator type that has been prone to fatalities which is further broken down to identifying the actual airline company which has seen the accidents. The lower part has been grouped to identify the country first and further break it down to specific routes.

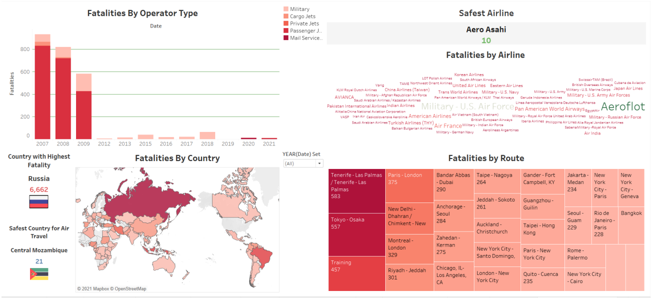


Figure 10 Airline specific and geographic analysis

V. Discussion

Based on some insights from our analysis, we expect local transport departments will be aware of current air crash statistics in the world. Also, we hope our findings could help reduce human errors which is the most common airplane accident reason under extreme weather conditions.

Fig.1 depicts the air accident trend from 1944 to 2008, the number of accidents decreasing after the peak year 1972. One possible reason for the decline in the trend could be attributed to the increase in the popularity of flight travel that led to the inevitable advancements in the technology and higher quality of training provided to the pilots and the flight crew .

Fig.2 shows the common causes of accidents and the human errors have accounted for more fatalities than any other category. This could be due to a number of reasons including pilots having lack of training, fatigue, negligence of flight crew members, etc. According to the Interstate Aviation Committee report by 2018, 75 percent of plane crashes were caused by pilots' errors in Russia.

Fig.3 gives an insight into Aboard vs Fatality ratio. This would be used in our future work where we aim to analyze deeper into this question where factors like position of the passengers inside an aircraft, the part of the flight during which the accident occured, etc to gain more knowledge on this.

Fig.4 describes the type of flights that have been prone to accidents. Passenger flights have seen the highest fatalities mostly because of their frequency of operation and increase in popularity over the years.

Fig.5 identifies the airline companies that have been impacted by fatal crashes in the form of a text cloud. The size of the text corresponds to the sum of fatalities. Aeroloft which is a russian airline company has the highest crashes and fatalities. This supports the graph that shows Russia as the most dangerous country.

Fig.6. is an interactive dashboard that shows a collective view of countries with the number of fatalities that occurred. A filter has been provided to see year-specific values.

Fig.7 is a heat map visualization of the routes that have seen the highest fatalities over the years. Tenerife - Las Palmas have the highest record of crashes.

Fig.8 and Fig.9 are a consolidated view of the dashboards developed for the final view. The first dashboard describes the overall trend in the aviation industry whereas the second dashboard shows the breakdown in terms of airline and geographic information.

Overall we could conclude that the period of 1960-1980 have seen the highest fatalities with Russia being the worst hit. Aeroflot , a Russian airline company, has contributed to the maximum number of crashes with pilot/human error being the leading cause of accidents. This has however gone down over the years possibly due to more advancements in flight technology. Central Mozambique has faced the least number of fatalities

VI. Future Work

The future goal of the project would be to improvise on Fig.3 which shows the aboard vs fatality ratio. We aim to collect data that describes the position of passengers within an aircraft who have lost their lives in a crash and the degree of impact that it might have had on the fatality. We would analyze further on the part of the journey during which the crash occurred. eg: during take-off, landing, taxi, mid-air, etc. In this project, we used Tableau to briefly analyze and visualize internationally aviation crashes dataset. The dataset that we have describes the time the plane crashed. We would collect data that includes the take off time as well and identify a pattern during which the accident occurs. For eg: 10 minutes prior to landing or 10 mins after take off.

VII. What did we learn

We were able to identify and work with the most suitable dataset for our analysis. We got familiar with Talend, an ETL tool for data cleaning purposes. We had the chance to understand the Tableau software and the various features it offers to create an interactive dashboard for effective communication. We have chosen the best suited visualization types that would assist the end user to get a visual narrative of the problem statement. We were able to integrate the set of charts created into a single dashboard with additional supporting statistical components included as part of it.

VIII. References

[1] G. Li, S.P. Baker, J.G. Grabowski, G.W. Rebok, "Factors Associated with Pilot Errors is Aviation Crashes" , February 2001, Aviation Space and Environmental Medicine 72(1):52-8, https://www.researchgate.net/publication/12136068_Factors_associated_with_pilot_error_in_aviation_crashes.

[2] John A. Caldwell, “Crew Schedules, Sleep Deprivation, and Aviation Performance”, First Published March 20, 2012, https://doi.org/10.1177/0963721411435842.

[3] D. Kerfoot, M. Hoffman, “Analysis of Aviation Accidents Data”, October 2018, Conference: Civil Engineering Research in Ireland (CERI), Dublin, Ireland, https://www.researchgate.net/publication/331742698_ANALYSIS_OF_AVIATION_ACCIDENTS_DATA

[4] Dataset : https://www.kaggle.com/cgurkan/airplane-crash-data-since-1908

[5] Aljazeera News: https://www.aljazeera.com/news/2021/7/8/plane-crash-emphasises-russian-poor-safety-record-regional-woes