Airplane Crashes (1908-2010) – Power BI Report

Dataset Link: [airplane\_crashes\_and\_fatalities\_since\_1908](https://www.kaggle.com/datasets/saurograndi/airplane-crashes-since-1908)

# Raw Data:

Rows: 5269

Columns: 13

# Process:

Pre-process the data and perform following tasks with several other steps:

In “time” column remove ambiguous values and also remove unwanted rows w.r.t. this “time” column.

Create a duplicate column of “date” column, split that column by delimiter and then extract year (gave column name “Year”) of the crash as well as month (gave column name “Month”) of the crash. Also changed their data type to *Whole Number* from *Text*.

Create a duplicate of “time” column, split it by delimiter and the extract the Hour of the incident (gave column name “Hour”). Also changed their data type to *Whole Number* from *Text*.

Create a new custom column “YearBracket”, this column is created to break down the “Year” data into further decades, which defiantly makes Data Analysis more easy and understandable. Applied following conditional query:

= if([Year]>1900 and [Year]<=1910) then "1900" else if([Year]>1910 and [Year]<=1920) then "1910" else if([Year]>1920 and [Year]<=1930) then "1920" else if([Year]>1930 and [Year]<=1940) then "1930" else if([Year]>1940 and [Year]<=1950) then "1940" else if([Year]>1950 and [Year]<=1960) then "1950" else if([Year]>1960 and [Year]<=1970) then "1960" else if([Year]>1970 and [Year]<=1980) then "1970" else if([Year]>1980 and [Year]<=1990) then "1980" else if([Year]>1990 and [Year]<=2000) then "1990" else if([Year]>2000 and [Year]<=2010) then "2000" else "Others"

Also changed their data type to *Whole Number* from *Text*.

Create a new custom column “DayOrNight”, this column is created to break down the “Hour” data into further Time Brackets to make Data Analysis more easy and understandable. Applied following conditional query:

if([Hour] > 6 and [Hour] < 19) then "Day" else if(([Hour] >= 0 and [Hour] <= 6) or ([Hour] >= 19 and [Hour] <= 23)) then "Night" else if([Hour] = Null.Type) then "Unknown" else "")

Then replace the “Error” values with “Unknown”.

# Processed Data:

Rows: 5198

Columns: 18

# Visualization:

# Insights:

No of people on aboard was 144K

Total Fatalities happened 105K

Total Fatalities happened on Ground due crashes 8436

In all other decades there are low fatalities on ground but in last decade (2000-2010) we can clearly see that it increases suddenly.

In Months data there we can see in January and December there are more plane crashes and in April, May and June there very low rate of accidents and surprisingly in February there are also very few crashes.

In Hours data the statistics shows that there are low rate of incidents in night/low light hours (0, 1, 2, 3, 4, 5, 6, and 21, 22, 23) and there high rate of incidents in day/bright light hours (7 to 20).

If we see decade’s data then 1970’s have most number of crashes.

In day brackets there are more incidents happened in day time and less in night time, but there is a huge number of data (almost 41%) which have missing data.

There almost 30K people who survived those crashes. The people who were on aboard but survived the accident.

More than 8K people died when plane crashed on ground they were present at the ground place.

# Conclusion:

In conclusion of this whole analysis we can say that in day time more crashes happened, its confusing may be there are more flights take off in day time in early days due to less technology advancement is a major reason for that. In winter seasons there is a high rate of incidents, maybe it’s due to low see range in winter and also bad weather but on the other side in summer when mostly winter is clear less crashes occurred. In 2000’s there are more people died on ground may be 9/11 is the major reason of that. In this data wars are also included so, all those factors also have huge impact on the data, like after 1940 there is a high curve of incidents and also mostly less 5 fatalities occurred. And in the end almost 30K people survived those crashes.