DATA ANALYSIS AND STORYTELLING: AIRPLANE AND FATALITIES FROM 1908 – 2009

The likelihood of dying in a plane crash (or even being in one, fatal or otherwise) is basically non-existent. According to statistics from The Economists, the probability of your plane going down is about one in 5.4 million. On the basis, the risk looks pretty small.

PROBLEM STATEMENT

This project is capstone project after concluding the [#30daysoflearning](https://techcommunity.microsoft.com/t5/educator-developer-blog/learning-data-analysis-curriculum-and-resources/ba-p/3497797) organized by [Olanrewaju oyinbooke](https://github.com/theoyinbooke).

The dataset contains data of airplane accidents involving civil, commercial and civil transport worldwide from 1908-09-17 to 2009-06-08. I was interested in the following questions;

1. How many plane crashed yearly? how many were on board? How many survived? How many died?
2. Highest number of crashes by operator?
3. Number of death and fatalities by operator?

PROJECT STRATEGY

The steps using in analyzing this data are listed below;

* Data Collection
* Data cleaning and transformation
* Data analysis and visualization
* Insights

DATA COLLECTION

On the contrary, knowing the purpose of the analysis was my first initial step. The data on Airplane crash and fatalities was collected from [kaggle](https://www.kaggle.com/datasets/saurograndi/airplane-crashes-since-1908) using “Get Data’’ on PowerBi Desktop following the ETL process ( Extract Transform and Load)

DATA CLEANING AND TRANSFORMATION

After extracting the data into power query, the next step was to transform and clean the data in power query for better optimization. “Cleaning data is very important because it helps to prevent errors when analyzing, avoid duplication etc ”. The first thing I did was to insert the appropriate data type into each column because some columns had a wrong data type. Then I looked for duplicates rows using ‘column distribution’ in power query. There were null values in the ‘number of fatalities’ column and I replaced them with 0. I removed unnecessary columns in the data so as to avoid redundancy and created new measures using DAX

DATA ANALYSIS AND VISUALIZATION

After cleaning and transforming the data thoroughly, my goal was to analyze and visualize the data for meaningful insights. I took my time to look for trends within the data such as the number of plane crashes by operators etc, and deep dived into the data for further trends and behaviors.

I was so excited when visualizing and building the dashboard in Power bi because this was my first project after learning Power bi. Therefore, I intended doing this project to improve and showcase my skill in data analysis in Power bi

Project Insights

* The total number of crashes became tragic during the 1970s and 1980s
* Total survivors started trending down by 2002 falling by 14.31% in 7 years
* Only 21% of people survived the general plane crash between 1908 to 2009
* At 179, Aeroflot had the highest total crash which accounted for 33.27% of total crash
* Fatalities dropped from 970 during its steepest decline between 1999 and 2009
* Aeroflot still accounted for 35.57% of total fatalities and recorded more than 7000 death

**However, even with the poor record of the number of death and fatalities per crashes, there is still a positive relationship between total number of crashes and the total number of people boarding plane every year.**

[View reports here](file:///C:\Users\User\Documents\airplane%20project.pdf)