# Project - Airplane Crashes till 2023

## 🧮 Dataset Overview: Variable Breakdown by Type

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Description** | **Type of Data** | **Measurement Scale** |
| Date | Date of the crash | Categorical (temporal) | **Interval** (time-based) |
| Time | Time of the crash | Categorical (temporal) | **Interval** (but incomplete) |
| Location | City, region, country | Categorical | **Nominal** |
| Operator | Who operated the aircraft (e.g., Military - U.S. Navy) | Categorical | **Nominal** |
| Flight # | Flight number (often missing) | Categorical / Missing | **Nominal** |
| Route | Route taken (e.g., "Paris - London") | Categorical | **Nominal** |
| AC Type | Aircraft type | Categorical | **Nominal** |
| Registration | Aircraft registration number (optional ID) | Categorical | **Nominal** |
| cn/ln | Construction/line number (aircraft production serial) | Categorical / Missing | **Nominal** |
| Aboard | Total number of people aboard | Numerical | **Ratio** |
| Aboard Passangers | Number of passengers aboard | Numerical / Missing | **Ratio** |
| Aboard Crew | Number of crew aboard | Numerical / Missing | **Ratio** |
| Fatalities | Total fatalities | Numerical | **Ratio** |
| Fatalities Passangers | Passenger fatalities | Numerical / Missing | **Ratio** |
| Fatalities Crew | Crew fatalities | Numerical / Missing | **Ratio** |
| Ground | Number of people killed on ground | Numerical | **Ratio** |
| Summary | Description of the crash | Categorical (text) | **Nominal** (free text) |

## 🔮 Feature Engineering

**Feature Engineering** is like seasoning your dish — raw ingredients (data) are fine, but features? That’s the *flavor bomb*. 💣✨

Feature engineering = creating new, more useful variables (features) from your existing raw data

### 🚀 Top 6 Feature Engineered Variables for Airplane Crash Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Name** | **Type** | **How It’s Created** | **Why It’s Useful** |
| Year | Numerical | Extracted from Date column | Helps analyze trends over time, group crashes by decades or war periods |
| is\_Military | Binary (0/1) | If 'military' appears in Operator column (case-insensitive) | Allows comparison of military vs civilian crash rates |
| Fatality\_Rate | Ratio (0–1) | Fatalities / Aboard | Indicates severity of crash; essential for comparing across incidents |
| Is\_Fatal | Binary (0/1) | 1 if Fatalities > 0, else 0 | Useful as a target variable for classification or just segmentation |
| Aircraft\_Type\_Simple | Categorical | Simplified from AC Type (e.g. 'Zeppelin', 'Seaplane', 'Jet', etc.) | Helps group by broad aircraft categories |
| Crash\_Location\_Type | Categorical | Extracted from Location (e.g. contains "Sea", "Island", "Mountain") | Useful for analyzing crash patterns by geography |

## EDA

### 📂 BASIC QUESTIONS – "What's here?"

|  |  |  |
| --- | --- | --- |
| **#** | **Question** | **Answers** |
| 1 | How many total crashes are there? |  |
| 2 | How many rows (crashes) are missing data? |  |
| 3 | What are the most common aircraft types? |  |
| 4 | What are the most common crash locations (top 10)? |  |
| 5 | How many unique operators are there? |  |
| 6 | How many crashes involve military vs. civilian operators? |  |
| 7 | What's the distribution of total people aboard? |  |
| 8 | What % of crashes involved fatalities? |  |
| 9 | What % of crashes killed people on the ground? |  |

### 📈 TIME-BASED TRENDS – "When did things happen?"

|  |  |
| --- | --- |
| **#** | **Question** |
| 10 | How do crash counts trend over years? (line chart) |
| 11 | Is there any decade with notably higher crash rates? |
| 12 | Are there more crashes in certain months/seasons? |
| 13 | Are crashes more likely at certain times of day? (morning vs. night) |
| 14 | How do total fatalities vary by year or decade? |
| 15 | Do military crashes peak during world wars? |

### 🧮 FATALITY INSIGHTS – "Who died and how bad was it?"

|  |  |
| --- | --- |
| **#** | **Question** |
| 16 | What’s the average number of people aboard per crash? |
| 17 | What’s the average number of fatalities per crash? |
| 18 | What’s the deadliest crash in the dataset? |
| 19 | How often do crashes kill all passengers/crew? |
| 20 | How many crashes had zero fatalities? |
| 21 | What's the ratio of fatalities:aboard per year? |
| 22 | Are passenger or crew fatalities more common? |
| 23 | How many crashes killed people on the ground? Top 5 worst? |

### 🧰 AIRCRAFT & OPERATOR ANALYSIS – "What kinds of planes crash?"

|  |  |
| --- | --- |
| **#** | **Question** |
| 24 | Which aircraft types have the most crashes? |
| 25 | Which aircraft types have had the most crashes since 2000? |
| 26 | Which commercial aircraft types have the highest fatality averages in recent years? |
| 27 | Which current operators have the most fatal crashes since 2000? |
| 28 | Are modern ground fatalities still a concern, and which commercial operators are involved? |
|  |  |

### 🌎 LOCATION-BASED – "Where are the crashes happening?"

|  |  |
| --- | --- |
| **#** | **Question** |
| 29 | Top 15 countries with most aircraft crashes since 2000 |
| 30 | Region-wise distribution since 2000 (using 'Crash\_Location\_Type') |
| 31 | Top Crash Locations (City/Region) Since 2000: |

### 🔤 TEXT SUMMARY ANALYSIS – "What are the summaries saying?"

|  |  |
| --- | --- |
| **#** | **Question** |
| 32 | Generate word cloud of most frequent words in Summary |
| 33 | What keywords appear most often in deadly crashes? |
| 34 | Common causes (manual tagging or keyword search like “fire”, “shot down”, “engine failure”) |
| 35 | Frequency of known vs. unknown causes |
| 36 | What keywords dominate crash summaries during the World War I and II periods compared to the Cold War and post-Cold War eras? |
| 37 | Compare use of "unknown" or vague causes in summaries pre vs post-2000. |

### 🧠 DERIVED METRICS – Smart Feature Ideas

|  |  |
| --- | --- |
| **#** | **Metric/Question** |
| 37 | Fatality Rate = Fatalities / Aboard |
| 38 | Is it a high-fatality crash? (e.g., >10 deaths) |
| 39 | Was it during war time? (WWI, WWII era tags) |
| 40 | Type of crash: Military / Civilian / Mail / Private |

### 🔥 TOP 10 MOST RELEVANT & IMPACTFUL

portfolio project, **these 10 are gold**:

1. **Crash trend over years**
2. **Top 10 deadliest crashes**
3. **Military vs. Civilian crash analysis**
4. **Top crash locations (bar chart or map)**
5. **Fatality rate per crash**
6. **Aircraft type vs. fatality average**
7. **Operator with most crashes**
8. **Total fatalities per year**
9. **Word cloud of common crash causes**
10. **Passenger vs. crew vs. ground fatality breakdown**

predictive modeling dataset