AIRLINE ACCIDENT RISK ANALYSIS

By: Eddy Wiwatsu

Moringa School

PROJECT BACKGROUND

The Aviation industry is data-rich. Understanding accidents trentscan reveal which aircraft are safest

Using data science techniques, we can identify trends, risk patterns, and contributing factors.

This project focuses on identifying which aircraft models are safest based on accident data.

OBJECTIVE

Load and explore the Airline Accidents dataset.

Perform exploratory data analysis (EDA) to uncover patterns and anomalies.

Identify and visualize the safest aircraft models.

Communicate findings through visuals and an interactive notebook.

Clean and prepare the data for analysis.

Engineer a “Risk Score” metric combining fatality, injury, and damage severity.

```
RangeIndex: 150959 entries, 0 to 150958
Data columns (total 31 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Event Id                150959 non-null  object
 1   Investigation Type      150959 non-null  object
 2   Accident Number         150959 non-null  object
 3   Event Date              150959 non-null  object
 4   Location                150959 non-null  object
 5   Country                 150959 non-null  object
 6   Latitude                150959 non-null  object
 7   Longitude               150959 non-null  object
 8   Airport Code            150959 non-null  object
 9   Airport Name            150959 non-null  object
 10  Injury Severity         150959 non-null  object
 11  Aircraft Damage         150959 non-null  object
 12  Aircraft Category       150959 non-null  object
 13  Registration Number     150959 non-null  object
 14  Make                    150959 non-null  object
 15  Model                   150959 non-null  object
 16  Amateur Built           150959 non-null  object
 17  Number of Engines       150959 non-null  object
 18  Engine Type             150959 non-null  object
 19  FAR Description         150959 non-null  object
 20  Schedule                150959 non-null  object
 21  Purpose of Flight       150959 non-null  object
 22  Air Carrier             150959 non-null  object
 23  Total Fatal Injuries    150959 non-null  object
 24  Total Serious Injuries  150959 non-null  object
 25  Total Minor Injuries    150959 non-null  object
 26  Total Uninjured         150959 non-null  object
 27  Weather Condition       150959 non-null  object
 28  Broad Phase of Flight   150959 non-null  object
 29  Report Publication Date 150959 non-null  object
 30  Unnamed: 30             150959 non-null  object
dtypes: object(31)
```

SUSTAINABILITY GOALS

Dataset Source: Provided by Moringa School for educational use.
File: airline_accidents.csv

Number of Records: *80,000+ rows (each row represents a unique accident record).*

Key Columns:
Event Date – Date of the accident
Make – Aircraft manufacturer (e.g., Boeing, Cessna)
Model – Aircraft model (e.g., 737, 172)
Injury Severity – Degree of damage/injury (fatal, serious, none)
Total Fatal Injuries, Total Serious Injuries, Total Minor Injuries
Aircraft Damage – Degree of aircraft destruction (destroyed, substantial, minor)

1. Loading Data: Imported using Pandas in Jupyter Notebook.
2. Handling Missing Values: Dropped or imputed missing entries.
3. Renamed Columns: Standardized column headers for consistency.
4. Converted Dates: Changed Event Date to proper datetime format.

DATA CLEANING & PREPROCESSING

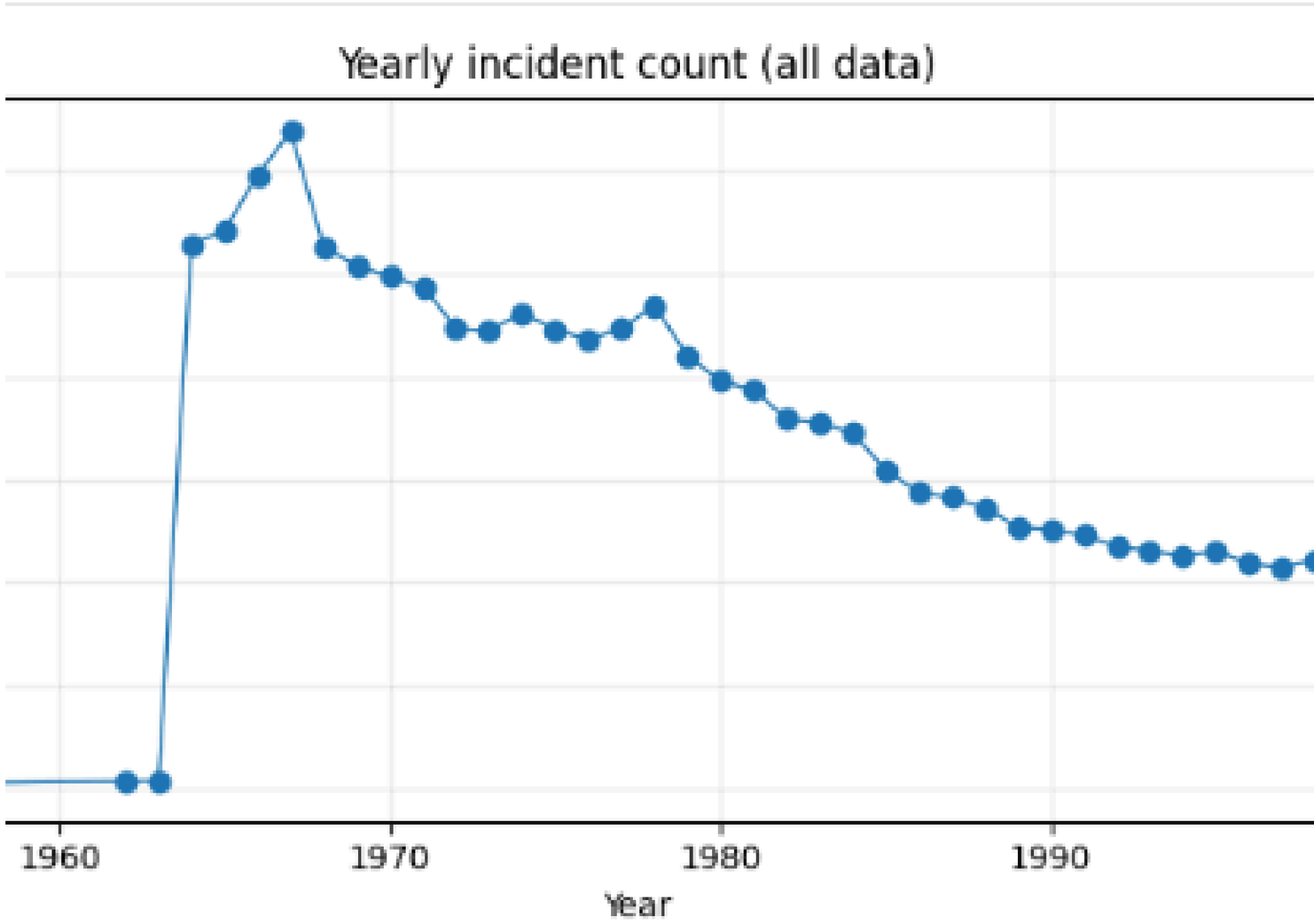
5. Extracted Year: Created a Year column for time-based analysis
6. Cleaned Text Data: Trimmed spaces and converted all text to lowercase.
7. Filtered Columns: Kept relevant columns for risk calculation.
8. Checked Data Types: Ensured numerical fields are integers/floats..

```
# parse the date column

df['event_date'] = pd.to_datetime(df['Event Date'], errors= 'coerce')

# creating year column

df['year'] = df['event_date'].dt.year

# cleaning make and model column names

df['Aircraft_Model'] = df['Model'].fillna('unknown').str.strip().str.upper

df['Aircraft_Make'] = df['Make'].fillna('unknown').str.strip().str.upper


# checking on missing values

df.isna().sum().sort_values(ascending = False).head(10)
```

Protecting Nations and Systems

EDA

```
First 10 rows of the cleaned numeric columns:
   Total Fatal Injuries  Total Serious Injuries  Total Minor Injuries  \
0                   0.0                     0.0                   0.0
1                   0.0                     0.0                   0.0
2                   0.0                     0.0                   0.0
3                   1.0                     0.0                   0.0
4                   3.0                     0.0                   0.0
5                   0.0                     0.0                   0.0
6                   0.0                     0.0                   0.0
7                   0.0                     0.0                   0.0
8                   1.0                     1.0                   2.0
9                   2.0                     0.0                   0.0

   Aircraft Damage
0              0.0
1              0.0
2              0.0
3              0.0
4              0.0
5              0.0
6              0.0
7              0.0
8              0.0
9              0.0
```



Accidents Over Time

Accidents have reduced significantly since the 1970s. Peak accident periods often align with older aircraft generations.

Correlation Analysis

Strong relationship between fatal injuries and total damage.

INSIGHTS & INTERPRETATION

Travel Insights

Accident rates and risk scores decrease over decades.

Safety

Modern aircraft show lower risk due to improved technology.

Data Quality

Manufacturers with fewer than 20 accidents were excluded to maintain reliability.

Model Behaviour

Single-engine light aircraft (like Piper and Cessna) dominate the dataset due to general aviation frequency.

LIMITATION

Missing or inconsistent records in historical data.

Lack of flight count data (cannot compute accident rate per flight).

Damage levels may be subjective or inconsistently reported.

Risk formula uses weighted assumptions; actual safety analysis needs engineering expertise.

Thank You