Aviation Accident Analysis

- Uncovering Patterns in US Flight Incidents
- Presented by: Tanveer Mbitiru Chege
- Moringa – Phase 1 Project

Project Overview

- Goal: Analyze aviation accident data to identify trends, potential causes, and safety patterns.
- Dataset: NTSB Aviation Data
- Scope: Aircraft type, flight phase, injuries, and fatality patterns.

Data Cleaning Steps

- • Removed columns with >50% missing values
- • Standardized inconsistent text values
- • Filled missing categorical fields with 'Unknown'
- • Converted dates and injury counts to correct types eg we had categorical data having unk, UNK, Unknown and Unknown in the same data.
- • Removed duplicates

Feature Engineering

- • Extracted Year and Month from Event Date
- • Created new columns:
- – Total.Occupants
- – Was.Fatal (1 if any fatalities)
- – Fatality.Rate (%)

Key Findings

- • Most common accident phase: Landing and Takeoff
- • Top manufacturers in accidents: Cessna, Piper, Beech
- • Most fatalities occurred during Cruise phase
- • Fatal accident rates have decreased over time

Visual Insights

- • Accidents Per Year
- • Accidents by Aircraft Model
- • Fatality at each Broad Phase of Flight
- • Fatality Rate by Year
- • Flight Phase vs Damage
- • Overlook of my data

Limitations

- • Some fields had many missing values (e.g., Weather)
- • Missing fatalities assumed as 0 may undercount true impact
- • Only reported NTSB accidents – no near-misses

Conclusion

- • Cleaned and analyzed real-world aviation data
- • Identified trends and created visuals
- • Built a complete data science pipeline from start to finish. I was given data that was mangled up, I had to explore it clean it, EDA on it, have a tableau visualization and come up with conclusions.