



geoffreykemboi / final_Airline_Accidents_Phase1_Project



<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security



final_Airline_Accidents_Phase1_Project / README.md



geoffreykemboi Remove subtitle from project overview in README



8393c65 · 18 minutes ago



88 lines (69 loc) · 3 KB

Preview

Code

Blame



Raw



Airline_Accident Data Analysis - Phase 0 project



Project Overview

The company is expanding into the aviation industry and seeks to determine **which aircraft models present the lowest operational risk** for potential acquisition.

This project analyzes historical aircraft accident data to identify models with the **lowest frequency and severity of accidents** based on recorded incidents.



Objectives

1. Assess aircraft safety using available accident data.
2. Identify aircraft models with the **lowest average fatalities per accident**.
3. Provide **actionable insights** and **recommendations** for the Aviation Division to guide purchase decisions.



Key Inferences

1. Use **both frequency and severity metrics**.

Models with frequent accidents but low average fatalities may still be operationally safe; both factors should be considered.

2. Avoid conclusions from small samples.

Models with fewer than five recorded accidents may not provide statistically meaningful results.

3. Normalize with exposure data if available.

Accident rates should ideally be compared per flight-hour or per number of aircraft in service to improve accuracy.

Dataset

File: airline_accidents.csv

Source: Internal or public aviation safety dataset

Rows: ~150,000

Columns: 31

Key fields used:

Column	Description
Event Date	Date of the accident
Make	Aircraft manufacturer
Model	Aircraft model identifier
Total Fatal Injuries	Number of fatalities in the accident
Aircraft Damage	Damage severity (used to group model data)

Methodology

1. Data Cleaning

- Removed null or irrelevant rows.
- Converted date fields to `datetime` format.
- Standardized model and manufacturer names.

2. Feature Engineering

- Derived aggregated risk metrics:
 - `accidents_count`
 - `total_fatalities`
 - `avg_fatalities_per_accident`
 - `percent_accidents_with_fatalities`

3. Analysis

- Grouped accidents by model.
- Filtered models with ≥ 5 accidents for statistical relevance.
- Ranked models by *average fatalities per accident*.

4. Visualization

- Bar chart of the top 10 safest aircraft (by average fatalities).
- Summary tables for:
 - Models with ≥ 5 accidents (low risk)
 - Top 20 most common aircraft by accident frequency



Outputs

Generated artifacts:

File	Description
aircraft_model_risk_summary.csv	Aggregated summary of model-level risk metrics
aircraft_risk_analysis.ipynb	Python notebook for the full analysis
README.md	Project documentation
<i>Visualization plots</i>	Bar chart comparing average fatalities per model



Tools & Libraries

- Python 3.10+
- pandas – Data manipulation and aggregation
- numpy – Numerical computations
- matplotlib – Visualization
- Jupyter Notebook / VSCode – Development environment