☰ ⌾ moruriinnocentlouis-collab / **project_phase_1**   🔍 📥 👤

<> Code   ⌾ Issues   ⑂ Pull requests   ▶ Actions   ▦ Projects   📖 Wiki   ⊘ Security   📈 Insights   ⚙ Set

👁   ⑂   ☆

My first project in my data science journey.

☆ **0** stars   ⑂ **0** forks   👁 **0** watching   ⑂ Branches   ⋀ Activity
                                                        🏷 Tags

🌐 Public repository

⑂   ⑂ **1 Branch**   🏷 **0 Tags**   ⑂   🏷 | 🔍 Go to file   t | | Go to file | | Add file + | | Code | ⋯

👤 **moruriinnocentlouis-collab** **Add tableau workbook**	7dededf · 4 minutes ago	🕐
📄 .gitignore	Add gitignore	3 hours ago
📄 Aircraft_Risk_Summary_202...	Add tableau analysis document	1 hour ago
📄 Aviation Safety Risk Assess...	Add powerpoint presentation	2 hours ago
📄 Aviation_Data.csv	import new file and updated note...	10 hours ago
📄 README.md	Adding readme	3 hours ago
📄 project_phase_1.ipynb	exported the tableau analysis file	50 minutes ago
📄 tableau project phase 1.twb	Add tableau workbook	4 minutes ago
📄 your_data.csv	adding cleaned csv file	2 hours ago
📄 your_data.xlsx	Add excel file cleaned	2 hours ago
📄 ~$your_data.xlsx	exported the tableau analysis file	50 minutes ago

📖 **README**   ✏

Project Phase 1

Overview

This repository contains the data, analysis, and findings for an aviation safety project focused on quantitatively assessing the risk associated with various aircraft models based on historical incident data.

The core objective is to move from reactive incident reporting to a proactive, data-driven system for fleet risk management. The analysis generates a comparable Risk Score for each aircraft model, providing clear direction for resource allocation, maintenance scheduling, and operational training.

Files in this Project

File Name

Description

project_phase_1.ipynb

The main Jupyter Notebook containing all data cleaning, transformation, risk score calculation, and visualization logic. This is the primary analysis document.

Aviation_Data.csv

The raw dataset used for the analysis, containing comprehensive records of aviation incidents and accidents.

README.md (This file)

Project documentation, setup instructions, and key findings.

📊 Methodology: The Aircraft Risk Score

The analysis uses a series of steps to convert raw incident counts into a meaningful, normalized risk metric:

Data Cleaning: Handling missing values, standardizing text fields (Make and Model), and ensuring data types are correct.

Incident Aggregation: Grouping records by Make and Model to calculate the Total Incidents for each specific aircraft type.

Risk Score Calculation: A logarithmic transformation is applied to the Total Incidents to create the final Risk Score.

Formula: $\text{Risk Score} \propto \log(\text{Total Incidents})$

Purpose: This transformation dampens the effect of extreme incident counts, resulting in a more stable and fairer score that is less influenced by simply the volume of flight hours or older data entries.

🛠️ Setup and Requirements

To run the analysis notebook, you will need a Python environment with the following libraries:

pandas: For data manipulation and cleaning.

numpy: For mathematical operations, including the logarithmic transformation.

matplotlib & seaborn: For generating the data visualizations.

Installation

You can install the required packages using pip:

pip install pandas numpy matplotlib seaborn jupyter

How to Run the Analysis

Clone this repository or download the files.

Ensure Aviation_Data.csv is in the same directory as project_phase_1.ipynb.

Open the notebook using Jupyter:

jupyter notebook project_phase_1.ipynb

Run all cells sequentially. The output will display the top/bottom risk tables and generate the three key visualizations.

🔍 Key Findings and Visualizations

The notebook generates three key visualizations, each supporting an actionable business recommendation:

Visualization

Insight Provided

Recommendation

Bar Chart: Top 10 Lowest Risk Models

Identifies models with exemplary safety records.

Benchmark Best Practices from these models and apply them fleet-wide.

Bar Chart: Top 10 Highest Risk Models

Pinpoints the models with the highest calculated risk score.

Prioritize Safety Programs on these high-risk models immediately, focusing on training and enhanced maintenance.

Scatter Plot: Risk Score vs. Total Incidents

Shows the overall distribution and relationship between volume and risk.

Implement Risk Thresholds for proactive, automated monitoring and mandatory review when models cross the alert line.

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

● **Jupyter Notebook** 100.0%