

# PHASE\_1.PROJECT

Data Analysis on the Aviation Business for both Private and Commercial Enterprise

## OVERVIEW

This repo looks intently on the Aircraft Risk Evaluation by accessing the various aircraft models for either commercial or private enterprises. For instance, some of the key risk factors would include: safety records, maintainance complexity and the historical performance based on the accidents recorded over the years from 1962 - 2023.

## BUSINESS PROBLEM

Your company is expanding in to new industries to diversify its portfolio. Specifically, they are interested in purchasing and operating airplanes for commercial and private enterprises, but do not know anything about the potential risks of aircraft. You are charged with determining which aircraft are the lowest risk for the company to start this new business endeavor. You must then translate your findings into actionable insights that the head of the new aviation division can use to help decide which aircraft to purchase

## THE DATA

The data we were given from the following sources:

1. From the National Transportation Safety Board that includes aviation accident data from 1962 to 2023 about civil aviation accidents and selected incidents in the United States and international waters.
2. Pulled from Kaggle dataset, here is the link: <https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses>

## SAFETY METRICS

This project will be a success if i'm able to identify:

1. Which Engine Number will provide safety for my Business Clients.
2. Which Purpose of Flight will be best suited for the new business venture.
3. If there's a correlation between Safety\_Score against Number.of.Engines.
4. Whats the distribution for the Number of Engines against the Make\_Model for cost efficiency.

## DESCRIPTION OF DATA

import Pandas as pd ... for data analysis and manipulation. import Numpy as np ... for mathematical functions especially when dealing with arrays. import seaborn as sns ... for data visualization import matplotlib.pyplot as plt ... for data visualization %matplotlib inline ... allows for the automatic rendering of Matplotlib plots directly in the notebook rather than in a separate window. Data Visualization in Tableau Public

## LOADING THE DATA

## USState\_codes.csv

---

```
df = pd.read_csv("USState_codes.csv", index_col = 0)
```

## AviationData.csv

---

```
df_1 = pd.read_csv('AviationData.csv', encoding='ISO-8859-1', dtype={'column_6_name': str, 'column_7_name': float, 'column_28_name': int}) df_1 = pd.read_csv('AviationData.csv', encoding='ISO-8859-1', low_memory=False)
```

## CLEANING THE DATA

```
cleaned_df6 cleaned_df6.to_csv('cleaned_data4.csv', index = False)
```

## MERGING THE DATAFRAMES

```
.merge() .. joins the two DataFrames within their index df6 = df5.merge(df, on = "Abbreviation")
```

## VISUALIZATIONS

1. Safety\_Score for the Top 10 Make\_Model
2. Distribution of safety score by purpose

## .merge() .. joins the two DataFrame within their index

---

```
df6 = df5.merge(df, on = "Abbreviation")
```

## CREATING CLEANED CSV FILE

```
cleaned_df6 = df6 cleaned_df6.to_csv('cleaned_data4.csv', index = False)
```