

# Case Study Rubric

## Predicting Aviation Accidents

### **Purpose:**

You will analyze civil aviation accident data from 2015-2024 to forecast accident trends for 2025 as well as investigate the factors closely correlated with accidents. In this project you will utilize time series forecasting skills and regression analysis techniques using real-world data.

### **Task:**

**Data Preparation:** Understand the dataset, handle missing values, clean and organize data into a time series format (monthly accident counts).

**Exploratory Data Analysis:** Identify trends, seasonality, and patterns in the data. Explore potential correlations with variables such as weather, aircraft make, and purpose of flight.

### **Modeling:**

- After analyzing the model to see if there was autocorrelation or partial autocorrelation, determine whether an ARIMA model is needed.
- If yes, build an ARIMA model and forecast the next year of data. If not, forecast on the original model.
- Conduct regression analysis to identify factors that correlate with accident frequency

### **Evaluation:**

- For ARIMA: report RMSE, MAE, assess stationarity with ADF test and interpret model fit with AIC and BIC
- For regression: report adjusted R-square, p-value and assess significance of predictors

### **Final Deliverable**

- Forecast of 2025 civil aviation accidents
- Analytical report summarizing trends, model findings, and identified risk factors

## **Assessment Criteria:**

<b>Spec Category</b>	<b>Meets Spec</b>
<b>Format</b>	One GitHub Repository (submitted via link on Canvas) <ul style="list-style-type: none"><li>- Make sure your repo includes the following:<ul style="list-style-type: none"><li>- A README.md file</li><li>- A LICENSE.md file</li><li>- A SCRIPTS folder</li><li>- A DATA folder</li><li>- AN OUTPUT folder</li></ul></li></ul>
<b>Data Preparation</b>	Dataset is cleaned appropriately; aggregation and formatting into time series is complete and accurate. Include script in your data folder.
<b>Exploratory Analysis</b>	Trends, seasonality, and external factors are thoughtfully analyzed and visualized. Include script in your data folder
<b>Model</b>	Use a properly fit model (using ACF, ACF/PACF, differencing if needed). Whichever model you use, ensure there are sufficient parameters to backup your model
<b>Regression Analysis</b>	Regression conducted thoughtfully; significant predictors identified with metrics to back up (adjusted r-squared etc.)
<b>Evaluation and Insight</b>	Metrics are well-documented and interpretations show understanding of model limitations and findings.
<b>Report Quality</b>	Clear and organized report summarizing methods, results, and conclusions.