Air Quality Analysis Case Study Rubric
DS 4002 – Spring 2025 – Neha Dacherla
Due Date: May 10, 2025

General Description: Submit a link to your GitHub repository on UVA Canvas.

Why am I doing this?
This case study provides you, a data science analyst, with the opportunity to leverage your data science skills by applying time series forecasting models to analyze air quality trends across the United States. Through this assignment, you will research real-world air quality data; develop a data and analysis plan; and perform the analysis plan to achieve the mission laid out to you.

What am I going to do?
The Github repository for this case study can be found using the following link: https://github.com/nehadacherla/DS4002-CaseStudy. Utilizing the "maine.csv" and "utah.csv" datasets, you will visualize the time series data; perform exploratory data analysis; and develop Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) models, 6 month air quality forecasting models for each state using ARIMA, and periodograms with Exponential Smoothing (ETS). You will then evaluate your models with the following metrics: RMSE (Root Mean Square Error); MAE (Mean Absolute Error); AIC (Akaike Information Criterion); and BIC (Bayesian Information Criterion).

Your final deliverables should include:
- The dataset you used along with a data dictionary
- Scripts for data processing, feature engineering, and model development
- A concise report summarizing methods and results
- A GitHub repository with every step of the project incorporating all materials

Tips for Success:
- **Be descriptive:** Explain each data processing, model building, and evaluation step thoroughly. Detailed documentation is critical.
- **Think real-world:** Understand that this project is very applicable to public health and environmental policy in our country and approach it like a real-world problem.

How will I know I have succeeded?
You meet expectations for the case study when you successfully complete the following criteria:

Spec Category	Spec Details
Formatting	• GitHub Repository submitted via link

	on Canvas containing all project files
	○ Contents:
	    ■ README.md
	    ■ LICENSE.md
	    ■ DATA folder with all data from project
	    ■ SCRIPTS folder with all source code from project
	    ■ REFERENCES.md

README.md	● **Project Overview:** Brief summary of the project, the purpose/motivation of the analysis, and the significance of forecasting air quality. ● **Instructions:** Provide instructions on how to reproduce results with enough detail for any data science analyst to follow.
Source Code File	● Well-documented R markdown (Rmd) file for the data cleaning and preprocessing steps ● Describe how the dataset was obtained and the preprocessing steps ● Include Exploratory Data Analysis script ● Include ACF and PACF models, ARIMA models, and ETS periodograms script in separate Rmd ● Provide RMSE, MAE, AIC, and BIC for models and discuss accuracy ● Comment on model parameters and their impact on results
Analysis & Results	● Summarize findings on air quality trend differences between Utah and Maine in a separate report ● Discuss the implications of findings for scientists and policymakers to assess
Visualization	● Include visualizations to understand correlation between the time series analysis; frequencies and periodic patterns; and air quality forecasting

	between both states ● Ensure visualizations are labeled with clear titles and key insights displayed
REFERENCES.md	● Citations: include a Rmd file titled "REFERENCES.md" citing all resources used and following IEEE documentation style

Acknowledgements