

Case Study Rubric

Due: TBD

Submission format: Upload PDF and link to GitHub repository to Canvas

General Description

Submit to Canvas your written PDF and a link to your GitHub repository.

Why am I doing this?

This study is an opportunity to demonstrate your technical and conceptual skills in data science. You will analyze a real-world energy dataset and produce a forecasting analysis similar to what you may encounter in other professional environments.

What am I going to do?

Using nearly 50 years of U.S. energy data, you will investigate whether renewable energy patterns can help predict fossil fuel production and consumption. You will conduct exploratory analysis and build two forecasting approaches: VARMAX, a multivariate time-series model, and XGBoost, a machine-learning regression model. You will provide a written deliverable that synthesizes your conclusions.

How will I know I have succeeded?

You will meet expectations on this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none">• Submit the written portion as a PDF file.• Submit all code in a GitHub repository.<ul style="list-style-type: none">o Include the cleaned dataset.o The repository should be titled “CS3-[FirstNameLastName]”.• Include references at the end of the written portion in IEEE citation style.

Written Portion	<ul style="list-style-type: none"> ● <u>Goal:</u> Discuss your interpretation of the study and your thought process by answering the following prompts: ● Problem Summary (2-3 sentences) <ul style="list-style-type: none"> ○ Summarize the problem presented in the study and explain why it is important in the context of U.S. energy security and renewable–fossil fuel interactions. ● Analysis Steps (1 paragraph) <ul style="list-style-type: none"> ○ Discuss the steps you took to meet the demands of the deliverable - i.e. reading in the data & how you created each model at a high-level. ● Results & Interpretation (1 paragraph) <ul style="list-style-type: none"> ○ Discuss the results of both forecasting approaches. ○ Explain the significance of these results in the broader context of renewable energy expansion, fossil fuel reliance, and national policy implications. ● Reflection <ul style="list-style-type: none"> ○ Explain the challenges you encountered while completing this study and how you addressed them. ○ Discuss what you might have done differently and how you could improve your approach in the future.
Code	<ul style="list-style-type: none"> ● Your code should include: ● Exploratory Data Analysis (EDA) <ul style="list-style-type: none"> ○ Examine long-term trends in fossil fuel and renewable energy production and consumption through at least 2 meaningful plots, summaries, and observations. ● Two Forecasting Approaches <ul style="list-style-type: none"> ○ You must implement both of the following: ○ 1. VARMAX- A time-series forecasting method <ul style="list-style-type: none"> ▪ Train and test the model on the provided cleaned dataset. Use renewable energy variables to forecast fossil fuel production and consumption. ○ 2. XGBoost: A machine-learning forecasting method

	<ul style="list-style-type: none"> ▪ Use renewable energy features (and any engineered lagged features, if needed) to predict fossil fuel values. ▪ Train/test using an appropriate chronological split. ● Model Evaluation <ul style="list-style-type: none"> ○ Evaluate both forecasting approaches using error metrics such as MAE and RMSE. ● Code Readability <ul style="list-style-type: none"> ○ Include comments so an outside viewer can understand your process clearly.
References	<ul style="list-style-type: none"> ● At the end of the written portion, include a list of references in IEEE citation style. ● These should be sources not already included in the provided reference materials.