

Household Energy Consumption Rubric

DS 4002 – Fall 2024 – Pranav Arora

Due: December 9

Submission format: Upload link to GitHub repo to Canvas

Individual Assignment

Why am I doing this? Through this case study, you will receive hands-on experience applying your data science expertise to a time-series regression task involving household energy consumption. You will have the chance to explore various exploratory data analysis (EDA) techniques to derive insights before learning about and implementing machine learning models that tackle this task, providing an opportunity to leverage data science to solve problems in the community.

What am I going to do? You can access the GitHub repository for this case study at https://github.com/aub5uy/DS4002_CS3. The materials folder contains the dataset, script file, and reference materials. Go ahead and download the dataset (household_power_consumption_dataset.csv) and script file (CS3_Script.ipynb) before opening them in a cloud environment like Google Colab or your local machine. Your task is to walk through the Jupyter Notebook, understand all the code segments, and implement any code areas marked as TODO. Through this process, you will create different graphs/plots to help better understand the data, see how the seasonal naïve and LSTM models work, and implement your own machine learning model! Finally, you will report the model performance results across various metrics to analyze its performance.

Your final deliverables should include:

- A data dictionary
- The final Jupyter notebook containing all implemented code with clear documentation
- A GitHub repository containing all materials and resources
 - Make sure you include the initial dataset and the completed script file
 - The README.md file should have a small written section about your model and a small reflective paragraph

Tips for success:

- Take the time to understand the exploratory data analysis steps and try to extract any value or insights from the plots that could inform model development
- Analyze the two models and see what steps were taken to reach the final results. This can save you a lot of time when developing your own model, as you have a baseline
- Do not worry if your model performance is low; this case study is about learning, exploring, and trying out new techniques to solve a real-world problem instead of just achieving a high accuracy score

How will I know I have Succeeded? You will meet expectations for this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"> ● Repository – A GitHub repo containing all materials <ul style="list-style-type: none"> ○ Submit a link to the repo in Canvas ○ Contents <ul style="list-style-type: none"> ▪ README.md ▪ LICENSE.md ▪ The final CS3_Script.ipynb file ▪ The initial dataset ▪ REFERENCES.md
README.md	<ul style="list-style-type: none"> ● Have a list of the elements you accomplished through this case study ● Regarding the model you developed, write a small paragraph that gives a brief overview of what it is, how it works, and how did it perform ● Write another small paragraph that reflects on your experience as a whole and any key learnings
DATA	<ul style="list-style-type: none"> ● You should have a DATA folder containing the final dataset and the original Data Appendix
SCRIPTS	<ul style="list-style-type: none"> ● The SCRIPTS folder should contain the final Jupyter Notebook file. Make sure it has all the TODO code blocks completed, including the exploratory data analysis plots and the generated model. All the following elements should be visible: <ul style="list-style-type: none"> ○ Bar graphs for the column mean and standard deviation ○ Correlation heatmap ○ Chart and graph for the performance metrics that includes the newly developed model ○ Thoroughly documented code, especially for the model and each step taken
OUTPUT	<ul style="list-style-type: none"> ● The OUTPUT folder should contain any graphs or charts you developed
REFERENCES.md	<ul style="list-style-type: none"> ● This file should include any citations or materials used during the completion of this case study. Please provide the citations in IEEE format

Acknowledgements: Thank you, Professor Alonzi, for providing the rubric template.