Weather Forecasting Model Case Study Rubric

DS 4002 – Bharathi Thambidurai
Due: TBD

Submission format:

- **Upload link to GitHub repo to Canvas**

Individual Assignment

Why am I doing this? This case study gives you the opportunity to apply foundational data science skills to a real-world question that affects local planning and decision making. You will explore how historical temperature patterns in Charlottesville can be used to build a simple forecasting model and evaluate how accurately that model predicts a recent year. Through this case study, you will practice data exploration, time-series modeling, visualization, and communicating findings clearly to a non-technical audience.

What am I going to do? You will work with a dataset containing Charlottesville's monthly temperature measurements from 2000–2024. Your goal is to analyze the historical period (2000–2023), develop a time-series forecasting model, generate a forecast for 2024, and compare your predictions to observed 2024 data. You will produce visualizations, a written summary, and reproducible code that communicates your approach and results. All necessary files, sample scripts, and supplemental readings are provided in: github.com/gbharathit/DS-4002-CS3-Case-Study

Deliverables include:

- A data dictionary of important variables, data types, and their definitions
- An R Markdown file used for EDA and analysis (well-documented)
- A PDF/document of plots and statistical model results
- A GitHub repository containing all data files, scripts, and documents used

Tips for success:

- Don't overthink it. Focus on clear, fundamental analysis, not advanced forecasting.
- Use the provided materials. The dataset, example scripts, and R tutorial guide you step-by-step.
- Ask for help. Instructors and peers can support you, especially if time-series concepts are new.
- Follow the workflow: EDA → Model → Forecast → Evaluate. Staying organized makes the project easier.
- Be creative. You're not limited to the example. Exploring other years or patterns is encouraged.
- Keep your code clean. Comment your steps so your process is clear and reproducible.
- Prioritize clarity. Strong explanations matter just as much as the model itself

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	● Repository – A GitHub repo (and cloud storage folder if necessary) containing all materials ○ Submit a link to the final repo ○ These are the top-level items in the repo ▪ README.md (auto displays) ▪ LICENSE.md (use MIT as default) ▪ A DATA folder – May either be included in the repo or linked as appropriate (e.g. GitHub won't let you store 2 TB) ▪ A SCRIPTS/OUTPUT folder
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ○ Use markdown headers to divide content. ○ Make an H2 (##) section explaining the contents of the repository ● Section 1: Software and platform section ○ The type(s) of software you used for the project. ○ The names of any add-on packages that need to be installed with the software. ○ The platform (e.g., Windows, Mac, or Linux) you used. ● Section 2: A Map of your documentation. In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder.
LICENSE.md	● Goal: This file explains to a visitor the terms under which they may use and cite your repository. ● Select an appropriate license from the GitHub options list on repository creation. ● Usually, the MIT license is appropriate.
DATA folder	● Goal: This folder contains all of the data for this project. ● Data ○ You should include the initial data, and the final data analyzed. ▪ The code in the SCRIPTS folder should be able to get you from the initial data to the final one ○ *Initial Data*: average monthly minimum and maximum temperatures (two separate .csv files) ○ *Final Data*: average monthly temperatures, removing unnecessary data and combined into one .csv file

SCRIPTS/OUTPUT folder	● Goal: This folder contains all of source code for your project and the output generated by your project, e.g. figures, tables, etc. ● This should include ○ *Source code file*: all code, including data cleaning, EDA, and model building in an R Markdown file called 'CS3_script' ○ *Data dictionary*: can be .pdf or .docx ○ *All EDA plots and statistical testing results*: can be screenshots on a .docx or can include the exported R Markdown file in a .html or .pdf format

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020).