CS3 Rubric – Case Study

DS 4002 – Fall 2024 – Chris Woods
Due: TBD
Submission format: Submit the link to your GitHub Repository
Individual Assignment

Why am I doing this? This project offers a hands on opportunity to test your technical and conceptual data science skills. This case study will simulate a project not unlike something you may encounter in a technical role at your future job. You synthesize a repository full of key insights to help FEMA better prepare for tornadoes. Using real-world data reiterates the importance of the skills you are learning in the classroom.

What am I going to do? Using the materials found in the GitHub repo, you will download and clean the NCEI tornado database, perform exploratory analysis, generate time series forecasting ARIMA and SARIMA models, run statistical tests, and then summarize your work.

Your final deliverables will be:

- Written summary presenting your methodology and findings
- Well documented Python scripts that encompass all the work you did on the project
- The dataset you used
- Any outputs or graphs you choose to make

All of this will be submitted electronically via a link to a GitHub repository built for the case study.

Tips for success:

- Be thorough. You are trying to generate actual, usable insights for a real government organization. Cheating yourself will only lead to less learning.
- Don't overthink it. A clear presentation of fundamentals is more valuable than an unclear presentation of cutting-edge techniques.
- Stay organized. The dataset we are giving you is huge, make sure you keep track of the cleaned version. Also, if you generate output graphs/files, be sure to name them specifically and clearly.

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	**One** GitHub Repository (submitted via link on Canvas) containing all materials used on the project. **GitHub should include:** -A README.md file (which auto displays) -A LICENSE.md file (use MIT as default)

	-A SCRIPTS folder with all your code -A DATA folder with the dataset(s) you used -AN OUTPUT folder with any outputs you generate
README.md	Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. **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. **Section 3:** Instructions for reproducing your results. In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise, but detailed and precise enough, to make it possible for an interested user to reproduce your results without much difficulty.
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.
SCRIPTS folder	Goal: Include materials necessary for a student to follow along to what you have done in the project. Include well documented python scripts detailing exactly what you did and how you did the following: ● Data cleaning and preprocessing ● Conduct exploratory analysis on the data (investigating tornado frequency, severity (f-score), and fatalities) ● Build ARIMA and SARIMA models to analyze tornado frequency over time

	• Evaluate your models using MAE • Generate statistically backed insights to explain changes over time in tornado frequency, severity (f-score), and fatalities
Summary of Analysis and Results (REPORT)	Goal: Present findings in a professional and informative manner. Your report should include analysis of your results in the following categories • Exploratory plots/findings • Effectiveness of time series models and which model you chose • Statistical insights **Visuals:** Your report should include visuals and graphs to aid in presenting the work you did. You should have visuals for: • Exploratory analysis • Time series models Your visuals should: • Effectively communicate information about tornado data trends • Be labeled clearly, titled accurately, and key takeaways are highlighted

Acknowledgements: Special thanks to Professor Alonzi for providing this rubric format and providing aid in completing this assignment.