Tornado Trend Analysis Case Study Rubric

DS 4002 – Fall 2024 – Instructor: Loreto Alonzi

Submission Format:

- Upload link to GitHub repository on UVA Canvas.
- Hard copy delivered to Prof. Alonzi's office (Room 344).

Why am I doing this?
This case study gives you the opportunity to leverage your data science skills by applying time-series forecasting models to analyze tornado trends across the United States. Through this assignment, you will gain experience using real-world data to develop insights that could potentially inform disaster management and preparedness strategies.

What am I going to do?
The GitHub repository for this case study can be found here. Using the "NCDC Storm Events Database" dataset, you will clean the data, perform exploratory analysis, and develop ARIMA and SARIMA models to forecast tornado metrics, such as frequency, severity (F3 and above), and property damage. You will compare the accuracy of the two models by focusing on Mean Absolute Error (MAE) and provide recommendations for disaster preparedness.

Your final deliverables should include:

- A concise report summarizing your methods and results.
- The dataset used, along with a data dictionary.
- Well-documented Python scripts for data processing, feature engineering, and model development.
- A GitHub repository containing all materials used.

Resources:

- Dataset: "NCDC Storm Events Database"
- Background: Introduction to ARIMA Models [Medium Article Link]

Tips for Success:

- **Be Thorough**: Explain each step of your data preprocessing, model building, and evaluation. Clear documentation is key.
- **Use Clear File and Variable Names**: Avoid ambiguous names like "df1" or "result2". Instead, use descriptive names like "tornado_data_cleaned" or "arima_predictions".
- **Engage in the Real-World Context**: Approach the problem with an understanding of its impact on public safety and resource allocation.

How Will I Know I Have Succeeded?

You will meet expectations for this case study when you successfully complete the following criteria:

Spec Category	Spec Details
Formatting	- **One GitHub Repository** (submitted via link on Canvas) containing all project files. - **Repository Contents**: - README.md - LICENSE.md - DATA folder with all your data - SCRIPTS folder with all your source code - REFERENCES.md
README.md	- **Project Overview**: Brief summary of the project, the purpose of your analysis, and the significance of forecasting tornado trends. - **Instructions**: How to reproduce the results. Provide enough detail for any student to follow.
Data Preparation	- Include a well-documented Jupyter Notebook for data cleaning and preprocessing. - Clearly describe how the dataset was obtained, any preprocessing steps, and the rationale for these steps.
Model Implementation	- Include ARIMA and SARIMA models in separate Jupyter Notebooks, with parameter selection clearly explained. - Provide MAE for each model and discuss model accuracy. - Clearly comment on model parameters and their impact on the results.
Analysis & Results	- Summarize your findings regarding tornado trends in a separate report. - Include z-tests or chi-square tests where relevant to compare trends across years. - Discuss the practical implications of your findings for disaster management.
Visualization	- Include visualizations that effectively communicate tornado frequency, severity trends, and model predictions. - Ensure that plots are labeled, titles are clear, and key insights are highlighted.

REFERENCES.md	- **Citation Requirements**: Include a Markdown file titled "REFERENCES.md" that cites all the resources used, following IEEE documentation style. - **Annotations**: Provide a brief annotation under each citation describing how it contributed to your case study.
Source Code Quality	- Well-documented code with comments explaining each major block of functionality. - Include descriptions for each function, its inputs, and outputs.
Submission of Hard Copy	- Submit the **Hook Document** and **Rubric** in printed form along with two printed references (explainer blog post and technical article).

Acknowledgements
Thank you to Professor Alonzi for providing guidance on the rubric structure and expectations for this assignment.