Hurricane Activity Case Study Rubric

DS 4002 - Fall 2024 - Instructor: Loreto Alonzi
Individual Assignment

General Description: This rubric outlines what the goals or the project are and the steps to successfully complete this assignment and replicate a case study concerning hurricane activity from a previous DS 4002 student.

Why am I doing this? This case study is an opportunity for the student to potentially gain new experiences with unfamiliar data or analysis tools, or apply their data science skills on real world problems. As the replicator works through this assignment, they will be exposed to replication of projects and result validation, which will be essential in the professional field.

What am I going to do? Access this GitHub repository; nw93929/DS4002-CS3, from a previous DS 4002 student to replicate a project completed during the course. Follow the instructions from this rubric and use the materials from the folder and GitHub repository to replicate the project and learn more about the topic, reflecting on potential future work you can see being done and the viability of the forecasting model.

Tips for Success:

- Read instructions carefully
- If confused on how to conduct analysis, look up package documentation!
- Review supplementary materials like the Data Appendix on the github in DOCUMENTS

How will I know I succeeded? You will meet expectations if you successfully follow the steps in the rubric below.

Spec Category	Spec Details
Data Collection	• Goal: download the dataset necessary for replication of the project • Navigate to the GitHub repository: https://github.com/nw93929/DS4002-CS3 • From the DATA folder, download "Hurricane_Dataset.csv"
Analysis	• Goal: Conduct Granger's Causality Test, Cointegration test, and build a VAR model with assumptions met from initial tests • In the CODE folder, the "Analysis.ipynb" file contains a code outline to successfully complete the analysis portion of the case study. • Gain an initial understanding of the data and justification for the analysis • Conduct Granger's Causality test to check for causality amongst our variables, the null hypothesis is the coefficients are zero, so a significant p-value means there is evidence that past values of X times series do influence the Y time series

	● Conduct the Cointegration test to ensure a statistically significant relationship is present, this is the base assumption of a VAR model so all variables should be significant in order to proceed ● Separate data into training and testing sets and conduct the Augmented-Dickey Fuller test for stationarity o If a series is not stationary, you have to difference until it is, VAR model assumes stationarity so it is needed ● Create a VAR model and select the optimal order based on AIC value
Results	● Goal: Produce visualizations and metrics that explain the viability of our VAR model ● Calculate the RMSE and interpret the result ● Plot forecasted data alongside the actual data and compare the two, reflect on the accuracy of the forecast and if the model is able to reliably predict hurricane frequency
Conclusion	● Goal: Summarize your results and write up your reflections on not just the results but also on your overall project experience ● Questions to consider: ○ What are the key findings from the tests and model results? ○ What do these findings tell you about the viability of using climate change to predict weather patterns? ○ What are some limitations of this case study? ○ What could be improved in future research on the topic? ○ What are some challenges you faced?