Case Study Rubric

DS 4002 – Fall 2024 - Tu-Yen Dang
Due: December 16
Submission format: Upload link to GitHub repository to UVA Canvas

Individual Assignments
General Description: Submit to canvas a link to your case study repository
Preparatory Assignments: Everything in the course

Why am I doing this? This case study offers a chance to apply your technical and conceptual data science skills in a time series analysis, exploring the longevity of music artists' careers. It presents a real-world scenario where you can leverage your existing knowledge while discovering new ways data science is used in practical applications.
- <u>Course Learning Objective</u>: Develop a project plan based on the scientific method principles related to the provided topic

What am I going to do? In this assignment, you will apply what you've been learning to an applicative topic. You now have the opportunity to use these built skills in an independent case study. This will include collecting data from a reliable source, manipulating it/performing an EDA, and analyze it using some form of time series analysis.

The GitHub repository for this case study can be found at https://github.com/tu-yendang/DS4002CS3. This repository is meant to guide you through the process; your task is to provide your own GitHub repository (submitted via link in Canvas) with your final deliverables, along with the process you used to get there. In this assignment, your final deliverables should include:
- GitHub repository – containing all code, data, outputs, and other materials used
    - Well documented source code
    - A data dictionary
    - Some form of time series analysis conclusion
    - Any necessary references used

Tips for success:
- Consider ways to perform a time series analysis before working on it. Certain approaches may be too simple, while others may be too complex for this case study.
- Research research research. This topic has many choices, many of which require prior knowledge/context on subjects unrelated to data science (ie: Disney Channel shows from the time period, what Spotify is, etc.). Familiarize yourself with these topics to make wiser decisions in your study.

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) - To ensure **reproducibility**, the repository will adapt parts of the TIER Protocol 4.0. In a nutshell, the top level page of the repository should contain:

	o README.md file (which auto displays) o LICENSE.md file (use MIT as default) o SCRIPTS folder o DATA folder o 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
	   o The type(s) of software you used for the project.
	   o The names of any packages that need to be installed.
	   o The platform (e.g., Windows, Mac, or Linux) you used.
	● Section 2: A Map of your documentation (outline or tree illustrating the hierarchy of folders and subfolders and the contents in them).
	● 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.

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: This folder contains all the source code for your project. ● Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results. ● All script files should include header comments at the beginning of a script to provide information that anyone working with or executing the script should be aware of. Throughout the scripts should also have comments explaining commands and their purposes.
DATA folder	● Goal: This folder contains all of the data for this project. ● Include initial and final data analyzed. If needed, the code in the SCRIPTS folder should be able to get you from the initial piece of data to the final one. ● If your data fits in github, place all of it here. If your data does not fit in GitHub use a single file explaining the process to obtain the dataset. ● A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics.
OUTPUT folder	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc. Use informative names for your files.
References	● All references should be listed at the end of the document ● Use IEEE Documentation style (link)