Case Study Rubric

DS 4002 – Fall 2024 – Erin Moulton
Due: TBD
Submission format:

- **Upload link to GitHub repository on UVA Canvas**

Individual Assignment

Why am I doing this?
This case study introduces you to a past project completed by a group of fourth-year data science students. By following their approach– covering data acquisition, exploratory analysis, and the implementation of a time series forecasting model– you will gain valuable insight into project reproducibility and learn an approach to a time series problem.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/erinmoulton/DS4002_CS3/tree/main. To begin, you will download the dataset available in the repository and begin working through the code which first introduces an API to incorporate song length into the dataset. Next, you will perform data cleaning steps such as specifying the years of focus, removing duplicate entries, and refining the dataset for analysis. Once the final dataset is prepared, you will perform an exploratory analysis to familiarize yourself with the data and assess its scope for addressing the goal. You will begin your analysis by performing a T-test to compare average song lengths before and after 2016. This will help verify if there is a significant changepoint in the data. Based on the T-test results, proceed with a Prophet time series forecasting model. Specify 2016 as the changepoint and generate predictions for the next 10 years.

Your final deliverables should include:

- A scatter plot visualizing the trend in average song length over time
- A forecast of future song lengths for the next 10 years, with clear attention to uncertainty intervals
- Well documented, commented source code
- A GitHub repository containing all materials used

How will I know I have succeeded?
You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	● One Github repository (submitted via link on Canvas) ● Create a new Github repository for this assignment titled 'CaseStudy_[insert first and last name]' that contains: ○ README.md ○ LICENSE.md ○ A SCRIPTS folder ○ A DATA folder
README.md	● Goal: This file serves as a summary of what you've produced for the case study and should orient the reader to your repository ● References should be listed at the end of the document ○ Use IEEE Documentation style
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

SCRIPTS folder	● Goal: This folder contains all the source code for your case study. ● Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results. Add comments throughout the scripts.
DATA folder	● Goal: This folder contains all of the data for this case study. ● Include both initial and final dataset