

# Music Genre Case Study Rubric

**DS 4002 – Spring 2025 - Sophie Phillips**

**Due:** May 9

**Submission format:** Upload link to your GitHub repository to UVA Canvas

**General Description:** Submit to canvas a link to your GitHub repository for this case study

**Why am I doing this?** This is your opportunity to demonstrate your data science and analysis skills to better understand music genre trends. You will conduct a time series analysis on music trends from 1896-2017 and predict the most popular genre for the upcoming year. In conducting your analysis, you will exercise drawing insights into what the results may mean from a sociocultural standpoint.

**What am I going to do?** You will provide a GitHub repository that meets the requirements of this case study rubric. All instructions and resources for this project can be found here:

<https://github.com/phillipsophie/Music-Genre-Case-Study>

You will conduct a time-series analysis on music data sourced from the UC Irvine Machine Learning Repository.<sup>1</sup> You will clean and merge data tables, perform exploratory data analysis (EDA), use Facebook's Prophet to forecast genre popularity over time, conduct validation on the model's accuracy, and analyze the results to draw meaningful conclusions from your findings.<sup>2</sup> You will have freedom on how you choose to clean the data and the scope of variables you choose to analyze within the dataset.

**Your final deliverables should include:**

- A GitHub repository containing your code files, data files, all visualizations produced, a data dictionary, and a detailed README file to ensure reproducibility
- Analysis consisting EDA of trends in music genres over time, a prediction of the most popular music genre the following year, and validation of the Prophet model's predictive accuracy
- A meaningful insight or conclusion drawn from your results

**Tips for success:**

- Remain open-minded and curious in you EDA to creatively think about possible insights of your results
- Link trends to real-world events to explain why genres rose or fell during specific periods.
- Use clear, intentional visuals that highlight key patterns and genre shifts over time.

**How will I know I have Succeeded?** You will meet expectations on the case study when you follow the criteria in the rubric below.

Formatting	<ul style="list-style-type: none"> <li>• One GitHub repository link submitted to UVA Canvas</li> <li>• Create a new GitHub repository titled 'Music_Genre_CS' containing the following: <ul style="list-style-type: none"> <li>◦ README.md file</li> <li>◦ LICENSE.md file</li> <li>◦ SCRIPTS folder</li> <li>◦ DATA folder</li> <li>◦ OUTPUT folder</li> </ul> </li> </ul>
README.md	<ul style="list-style-type: none"> <li>• <u>Goal</u>: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings.</li> <li>• Use <b>markdown headers</b> to divide content. <ul style="list-style-type: none"> <li>◦ Make an H2 (##) section explaining the contents of the repository</li> </ul> </li> <li>• Section 1: Software and platform section <ul style="list-style-type: none"> <li>◦ The type(s) of software you used for the project.</li> <li>◦ The names of any add-on packages that need to be installed with the software.</li> <li>◦ The platform (e.g., Windows, Mac, or Linux) you used.</li> </ul> </li> <li>• 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. </li> <li>• Section 3: Instructions for reproducing your results.   In this section, you should give <b>explicit step-by-step instructions</b> 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. </li> </ul>
SCRIPTS folder	<ul style="list-style-type: none"> <li>• <u>Goal</u>: This folder contains all the source code for your project.</li> <li>• Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results.</li> <li>• Your code must include EDA visualizations, a graph of music genre popularity over time, a graph of the Prophet model's prediction, and calculations on the model's accuracy (i.e MAE, RMSE, MAPE, etc.)</li> <li>• Throughout all your scripts, you should include <b>copious comments</b> explaining what each command or sequence of commands accomplishes and what the purpose is.</li> </ul>

DATA folder	<ul style="list-style-type: none"> <li>● <u>Goal</u>: This folder contains all of the data for this project.</li> <li>● You should AT LEAST the data include the initial data, and the final data analyzed.</li> <li>● If your data does not fit in GitHub use a single file explaining the process to obtain the dataset.</li> <li>● A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics.</li> </ul> <p>This file should be organized in sections, with a section for <b>each</b> dataset analyzed. Each section should begin with a statement of what the unit of observation is--that is, it should explain what kind of object each row of the data file represents.</p> <p>After that, you should include a subsection for each variable in the analyzed dataset.</p> <p>More information:  <a href="https://www.projecttier.org/tier-protocol/protocol-4-0/root/data/analysisdata/data-appendixfile">https://www.projecttier.org/tier-protocol/protocol-4-0/root/data/analysisdata/data-appendixfile</a> </p>
OUTPUT folder	<ul style="list-style-type: none"> <li>● <u>Goal</u>: This folder contains all of the output generated by your project, e.g. figures, tables, etc.</li> <li>● Importantly, any information like tables, figures shown in your presentation should be here.</li> <li>● Use <b>informative names</b> for your files.</li> </ul>
References	<ul style="list-style-type: none"> <li>● All references should be listed at the end of the document</li> <li>● Use IEEE Documentation style (<a href="#">link</a>)</li> </ul>

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from [Streifer & Palmer \(2020\)](#).

### References

- [1] Defferrard, M., Benzi, K., Vandergheynst, P., & Bresson, X. (2017). FMA: A Dataset For Music Analysis [Dataset]. UCI Machine Learning Repository.  
<https://archive.ics.uci.edu/dataset/386/fma+a+dataset+for+music+analysis>
- [2] GeeksforGeeks. (2024, November 21). *Time Series Analysis using Facebook Prophet*.  
<https://www.geeksforgeeks.org/time-series-analysis-using-facebook-prophet/>
- [3] Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, 2(5), 150081. <https://doi.org/10.1098/rsos.150081>