

CS3 Rubric – Case Study Create

DS 4002 – Instructor: Loreto Alonzi

Due: Monday after Presentation week of project cycle 3, via Canvas

Submission format:

- **Upload link to GitHub repo to Canvas AND submit on canvas**

Individual Assignment

Why am I doing this? You're joining Spotify's Data Science Innovation Team to investigate whether lyrics alone can be used to predict a song's genre. As part of this case study, you'll use NLP and machine learning to uncover genre-specific patterns in word choice, emotion, rhyme, and structure. This project is your opportunity to learn how machines interpret language and apply that to the real-world challenge of music classification.

- Course Learning Objective: prepare findings for presentation to your peers.

What am I going to do? This case study asks you to build and evaluate a genre classification model using only the lyrical content of songs. You will clean and explore a large dataset, extract NLP features, apply machine learning models, and analyze the linguistic elements that define genres.

- Preprocessing notebook or script
- Feature engineering notebook
- Model training + evaluation script
- 1–2 page reflection on genre-specific findings
- Organized GitHub repo with README, code, data, and references

All of this will be submitted electronically via a link to a GitHub repository built for the case study AND selected documents will be submitted by hard copy as well.

Tips for success:

- Use visuals to support your insights (genre word clouds, confusion matrices, etc.)
- Keep your code and outputs well-commented and organized
- Prioritize clarity over complexity
- Use the README to orient your user quickly and completely

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	<ul style="list-style-type: none">• Topic – The topic of your assignment is to determine lyrical genre classification using NPL and ML.

	<ul style="list-style-type: none"> • Repository – A GitHub repo (and cloud storage folder if necessary) containing all materials <ul style="list-style-type: none"> ○ Submit a link to the repo ○ Contents <ul style="list-style-type: none"> ▪ Various Materials (data, code, articles, etc.) ▪ README ○ Use pdf format when possible ○ For code and data products use the appropriate format for whatever it is
Repository	<ul style="list-style-type: none"> • <u>Goal:</u> This is a GitHub repository containing all digital materials for the case study. • The README should explain all parts of the repo • Use folders to organize the major categories of materials, code, etc.
Materials	<ul style="list-style-type: none"> • <u>Goal:</u> Include materials necessary for a student to engage with the case study. • Data, code, reproduction steps • Place ALL materials in the repo that is easy to find