Case Study Rubric: Song Lyrics

Why am I doing this? This is your opportunity to put into practice all the skills you have learned throughout your data science classes. It will capture the full lifecycle from data capturing, exploratory data analysis, model testing, and conclusion / results. Additionally, it allows you to reflect on and reproduce the work of one of your previous classmates.

- Course Learning Objective: recreate results from peers and apply learned skills.

What am I going to do? You will use the skills you have learned in this class to first download the Spotify song lyrics dataset off Kaggle. From here, you will first explore the data by performing EDA to find actionable insights. After this, you will continue cleaning the data in a digestible format for the model. Explore and experiment on a variety of models before settling on a preferred one. From here, hyperparameters can be altered and adjusted to optimize the results. Lastly, conclude on what you have found.

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) containing all project files. ○ **Repository Contents:** ○ README.md ○ LICENSE.md ○ DATA folder with all your data ○ SCRIPTS folder with all your source code
README.md	● The motivation and inspiration behind the project and the main deliverable. ● Restating packages required, description of the data, and instructions how to reproduce results.
Source Code File	● Include a well-documented Jupyter Notebook for data cleaning and preprocessing. ● Outline and describe how you obtained, cleaned and utilized the data during preprocessing.

	● Clearly show your exploratory analysis before any models begin and comment on the insights they reveal. ● Include all experimental models used before arriving at the preferred model and well documented reasoning as to why.
Analysis and Results	● Summarize your model accuracy in a separate report. ● Include recall, precision, and F1 score so that it can be clearly understood how the model evaluated individual classes. ● Reflect on the findings you delivered and how they can be transformed into actionable recommendations for the topic motivation.

Acknowledgements: This structure is pulled from Streifer & Palmer (2020).