Kenneth Nguyen, Jackson Le, Joanne Lee

Professor Jason Williamson

DS 2002

March 20, 2025

Data Project 1: Reflection Report

Throughout this project, we encountered several challenges. Still, the most significant one was dealing with how Spotify deprecated several of its API features in December 2024. This included the endpoint we had planned to use to obtain audio features for each song, a large part of our plan. This forced us to pivot and explore other ways of gathering information using the Spotify API, such as using other endpoints. However, once we determined how to adapt our use of the Spotify API, actually calling the API and incorporating the data was simple. Additionally, the data analysis and visualization part of the project was very straightforward largely due to the cleaning we did on the dataset beforehand which made it well-structured and easy to manipulate.

That said, the data cleaning stage itself was far more intensive and time-consuming than we expected. The multiple Kaggle datasets required extensive wrangling. We had to drop unnecessary columns, manage duplicate entries, and filter the data to only include songs within the Grammy eligibility range. Additionally, we had to do meticulous steps when merging the datasets such as matching and normalizing shared columns across datasets to not cause conflicts. The entire process was tedious but luckily set us up for an easier time afterwards.

Looking towards the future, a utility like this could be highly valuable for future data projects, specifically those in the music industry. Even with Spotify's recent API limitations, a similar approach to ours could be applied to data sourced from other streaming platforms or publicly available music databases. Our project could easily serve as a template for anyone looking to explore the factors that contribute to both commercial success and critical acclaim in the music world and help build strategies around artistry, songwriting, and marketing.