**Executive Summary**

This project analyzes the intersection of music popularity, gender, and critical acclaim by examining three datasets: the Top 10 Spotify Songs dataset (2010–2019), the Make Music Equal Pronoun & Gender Dataset, and the Grammy Awards Winners dataset. Using Python for data processing and Power BI for data visualization, the goal is to uncover trends in music attributes that influence commercial success versus award recognition and to determine if there is a notable difference between awards given to various genders.

**Motivation**

In the era of music streaming, there's a growing conversation regarding the disconnect between what the public listens to and what the industry recognizes. Many Grammy-winning songs are not among the most-streamed tracks, and vice versa. This project aims to investigate that gap and understand what qualities, including gender, may distinguish popular hits from award-winning songs. Insights can potentially support music professionals, marketers, and analysts in identifying patterns in music success, and provide fans with a deeper understanding of how recognition is earned.

**Data Question**

What are the musical characteristics and trends that differentiate commercially successful songs from Grammy-winning songs, and is there a correlation with gender?   
  
Articles:  
[What’s Wrong with the Grammys?](https://nique.net/opinions/2025/02/14/whats-wrong-with-the-grammys/)  
[Grammys 2023: They Still Can’t Get Rap Right](https://rollingstoneindia.com/grammys-2023-they-still-cant-get-rap-right/)  
[Are the Grammys Sexist?](https://www.cantgetmuchhigher.com/p/many-women-won-grammys-this-year)

**Minimum Viable Product (MVP)**

The final capstone will include:

1. A Power BI dashboard showing:

* Trends in top songs' musical features (e.g., BPM, energy, valence)
* Grammy winners by genre, year, and artist overlap
* Comparison visualizations of popularity vs. awards

1. A Python-based Jupyter Notebook for data cleaning and merging.
2. A PowerPoint presentation summarizing findings, visuals, and interpretations.

Intended audience: educators, music analysts, data enthusiasts, and the general public.

**Schedule (through <date of demo day>)**

1. Get the Data (May 4, 2025)

2. Clean & Explore the Data (May 24, 2025)

3. Create Presentation of your Analysis (June 17, 2025)

- Should be a presentation, but could include a Jupyter Notebook or

dashboard in Excel, Tableau, or PowerBI

4. Internal demos (June 21, 2025)

5. Demo Day!! (July 10, 2025)

**Data Sources**

[DataCamp Spotify Music Set](https://www.datacamp.com/datalab/datasets/dataset-python-spotify-music?utm_source=chatgpt.com)  
[Kaggle Grammy Winners Set](https://www.kaggle.com/datasets/johnpendenque/grammy-winners-and-nominees-from-1965-to-2024)  
[Make Music Equal Pronoun & Gender Dataset](https://makemusicequal.chartmetric.com/pronoun-gender-database)

**Known Issues and Challenges**

1. Data Matching:

The biggest known challenge, currently, is that artist and song titles are inconsistently formatted across datasets (e.g., variations like "Beyoncé" vs "Beyonce"). I will need to standardize artist/song names through fuzzy string matching and manual validation to address this.

1. Data Cleaning Needs:

* Normalize casing and punctuation in song/artist fields
* Resolve artist/group distinctions (Grammy entries may reference artists groups rather than individuals)
* Handle duplicate or ambiguous entries in Grammy data
* Resolve song/album distinctions (Grammy entries may refer to albums rather than individual songs)

Evaluate the billboard top 100