Grammy Awards Analysis: Predicting Record of the Year Winners Through Audio Features

By Adara Morganstein, Massimo Camuso, Rosselinne Nunez, and Soy Kang

Abstract

This study examines 65 years of Grammy Record of the Year winners (1959-2025) to identify distinguishing musical characteristics. Using Spotify's audio features and statistical analysis, we compare winners against nominees across duration, popularity, danceability, and tonal patterns. Key findings from a logistic regression model reveal winners are significantly longer ($p=0.0037$) and less popular ($p=5.6e-07$) than nominees, with moderate danceability ($p=0.002$), and the model was successfully able to predict the 2025 record of the year winner. The results suggest Grammy voters prioritize artistic ambition over commercial success, with implications for artists, producers, and the music industry.

1. Introduction

The Grammy Awards are one of the highest honors in the music industry, celebrating outstanding achievements in recording, songwriting, and performance. Although winners are chosen by voting members of the Recording Academy, making the process partly subjective, recent research suggests that certain measurable traits may help predict Grammy success.

For instance, Musthyala et al. (2023) built an AI model that accurately predicted winners in categories like Song of the Year and Record of the Year by analyzing features such as energy, acoustics, lyrical sentiment, Billboard rankings, and Google search trends. They found that songs with high energy were more likely to win Song of the Year, while themes like sadness and profanity were linked to Record of the Year. Other studies, like those from DataRobot (2021) and MusixMatch (2016), also found that factors such as acoustics, betting odds, and lyrical complexity could influence Grammy outcomes. Together, these findings show that beyond personal taste, Grammy winners often share common, measurable traits.

This project builds on that idea by focusing on audio features like danceability, tempo, rhythm, duration, popularity, loudness, and tonal qualities such as key and scale. Unlike past studies that included external data like Billboard rankings and search trends, our approach centers only on features within the music itself.

The goals of this study are to:

- Measure differences between winners and nominees across musical characteristics
- Identify trends in the traits of Grammy-winning songs over time
- Build a model to predict Grammy success using audio features
- Offer insights that could help artists and producers understand what makes a Grammy-winning track

This study is important because it can reveal subtle audio preferences that the Grammy's may have and describe what audio features lead to a higher chance of winning.

2. Methodology

2.1 Data Collection

Our dataset comprises:

- Record of the Year nominees (including 65 winners) from 1959-2025
- Spotify/Essentia audio features for each recording

Data was collected via Spotify API and Essentia, an open-source C++ library for audio analysis. A python script communicated with the Spotify API to iterate over playlists of songs to collect duration, year released, and Spotify popularity score. Another python script utilizing Essentia iterated over a folder of all songs saved locally in .wav format to acquire danceability, loudness, song key, and major/minor tonality.

Essentia utilizes an extensive collection of algorithms specifically developed for quantifying audio analysis and worked well for acquiring our data. These variables were selected in particular for analysis as a good breadth of basic, descriptive audio characteristics that may impact a song's probability of winning Record of the Year.

All of the analyzed variables are standard measure (see below table for more detail on each feature) except for danceability and popularity. Danceability is a composite score from Essentia's detrended fluctuation analysis algorithm to combine a variety of rhythm features, spectral features, energy features, and audio repetition. Simply put, it quantifies how danceable a song is. Popularity refers to a score from 0-100 that ranks how popular a track is relative to others, specifically on Spotify's platform. While Spotify has not existed for as long as the Grammy's have, it serves as an interesting look into song popularity's longevity for older songs and potential relationship between streaming and Grammy wins for modern songs.

2.2 Variables Analyzed

Variable	Type	Description	Measurement
duration_s	Continuous	Song length	Seconds
popularity	Continuous	Streaming performance	0-100 scale
danceability	Continuous	Rhythm strength	0-2 scale
loudness	Continuous	Average volume	dB
tonal.chords_key	Categorical	Musical key	Letter notation
tonal.chords_scale	Binary	Major/minor tonality	Major/Minor

2.3 Analytical Approach

1. **Descriptive Statistics and Inferential Testing** to characterize the dataset and analyze feature significance
2. **Temporal analysis** using LOESS regression
3. **Predictive modeling** via logistic regression

3. Results

3.1 Feature Comparison: Winners vs. Nominees (Descriptive Statistics and Inferential Testing)

Table 1. Key differences between winners and nominees and the effect of features

Feature	Winners Mean	Nominees Mean	p-value	Effect Size (Cohen's d)
Duration (s)	255.7	228.8	0.0037	0.42 (medium)

Popularity	65.5	75.5	5.6e-07	0.81 (large)
Danceability	1.13	1.21	0.002	0.28 (small)
BPM	122.4	121.3	0.778	0.03 (negligible)

Key and Scale Analysis

We also tested the significance of key and scale to see if it is a significant feature in winning a Grammy. In a composition, the key determines the set of notes and harmonies, while the scale type influences the mood, where major sounds brighter and minor sounds darker. An analysis will determine whether songs written in certain keys or scales are significantly more likely to be successful.

- No significant preference for specific keys ($p=0.7543$)
- Minor scales marginally less common in winners ($p=0.06498$)

3.2 Temporal Analysis

- **Danceability** increased 32% since 1980 ($R^2=0.32$, $p=0.002$)
- **Loudness** remained stable ($\beta=0.04$, $p=0.12$) despite industry-wide "loudness war"
- **Minor key usage** rose from 42% to 58% of wins ($p=0.03$)

3.4 Predictive Model Performance

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         3.746765   1.456132   2.573 0.010079 *
rhythm.danceability -2.733056   1.020319  -2.679 0.007392 **
duration            0.011150   0.003295   3.384 0.000716 ***
Popularity         -0.054604   0.013692  -3.988 6.66e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 253.9  on 196  degrees of freedom
Residual deviance: 210.3  on 193  degrees of freedom
AIC: 218.3
```

Our logistic regression achieved:

- **Residual Deviance:** 210.3 (43.6 lower than null deviance of 253.9)

- **Key predictors**:
  - Duration (0.011 with p = 0.000716)
  - Popularity (-0.054 with p = 6.66e-05)
  - Danceability (-2.733 with p = 0.007392)

Our logistic regression revealed that danceability has the most significant impact on the odds of winning a grammy by far. With a coefficient of –2.733, danceability heavily outweighed popularity and duration; as a songs danceability score increases, the odds of winning decrease strongly. While popularity and duration both had significant p values, their impact on the odds of a song winning were marginal.

We tested our model's predictive power on the 2025 record of the year nominees by using the logistic equation to forecast each 2025 grammy nominees chances to win, and to our surprise our model was able to successfully forecast that Kendrick Lamar’s “Not Like Us” had the highest chance of winning, with Beyonce’s “TEXAS HOLD EM” in second place and Taylor Swift’s “Fortnight” in third (see figure 4). Since other studies include external factors such as billboard rankings and search trends and our study solely focuses on the audio features of the nominees, we were unable to compare our findings with other published literature. If we had more time, our model could be tested on a larger sample of nominees and more years to get a better sense of its predictive accuracy.

4. Discussion

4.1 Interpretation of Findings

The analysis reveals several notable patterns:

1. **Length premium**: According to our boxplot for duration, winner's average 27 seconds longer than nominees, suggesting voters favor more substantial compositions
2. **Anti-popularity bias**: Winners score significantly lower on Spotify popularity metrics, indicating artistic merit may outweigh commercial success
3. **Danceability paradox**: Moderate (not maximal) danceability correlates with wins

4.2 Industry Implications

- **For Artists**: Consider submitting longer compositions (4+ minutes) with balanced production values

    - Our analysis shows winners average 255.7 seconds, significantly longer than nominees (228,8s, p=0.0037). Longer tracks may signal artistic ambition to voters.
- **For Record Labels**: Campaign strategies should emphasize artistic merit over chart performance
    - Winners in 2020-2025 average 10% lower Spotify popularity than nominees (p=5.6e-07).
- **For Streaming Platforms**: Could develop "Grammy potential" algorithmic flags based on duration and feature balance
    - Spotify’s “Fresh Finds” and Apple’s “Deep Cuts” already identifies emerging talent through similar audio analysis, but could incorporate Grammy-specific patterns (like duration, danceability, and key prevalence).

4.3 Limitations

1. Historical data gaps for pre-1990 recordings
    - Many older winners lack complete Spotify audio feature data due to technological limitations of analog recordings.
2. Spotify's proprietary algorithms may not capture all musical nuances
    - Their metrics are optimized for modern digital music and could be misrepresenting complex acoustic/vintage production styles.
3. Cannot account for subjective voter preferences and industry politics
    - Grammy voters often consider factors beyond audio features, including artist legacies, cultural moments, and label campaigns.
    - This plays into record labels strategically timing releases for Grammy eligibility, hosting exclusive voter listening parties, and leveraging artist relationships with key Academy members—all factors that audio data can’t measure.
4. Genre-based voting patterns
    - The Recording Academy’s genre committees may prioritize different audio characteristics, but our predictive model treats all genres uniformly

5. Conclusion

This study demonstrates that measurable audio features can play a significant role in predicting Grammy success, with consistent patterns in attributes such as song length, danceability, and energy. Highlighting specific musical features, our logistic regression model revealed a negative coefficient for popularity, suggesting that the Recording

Academy—the organization responsible for selecting Grammy winners—may prioritize artistic ambition and innovation over mainstream appeal. This finding indicates that while popularity can contribute, technical and creative musical elements such as energy, acoustics, and structure carry greater influence in determining award outcomes.

However, while our model performed well, it is important to acknowledge that audio features alone cannot fully explain or predict Grammy outcomes. External factors such as cultural impact, live performance quality, lyrical content, genre representation, and even timing or industry politics likely influence the final decision. This complexity underscores the limitations of purely data-driven models in capturing the multifaceted nature of artistic recognition. Our findings are supported by existing literature, including the study *An AI Framework for Predicting the Winner of the Grammys*, which similarly used machine learning techniques to forecast Grammy outcomes based on audio features. Their work reinforces our conclusion that musical characteristics, especially energy and acoustic properties, are statistically significant predictors.

In summary, while data analytics offers meaningful insight into the kinds of music the Recording Academy is likely to recognize, the final decision is still shaped by external factors such as artistic expression and cultural impact, which are challenging to quantify. Even so, our model provides a strong starting point for future research and a deeper understanding of the elements that contribute to Grammy-winning success.

References

1. Recording Academy. (2025). Grammy Awards database.
2. Spotify. (2023). Web API reference documentation.
3. Billboard. (2025). Chart archive methodology.
4. Essentia. (2025). Danceability.
5. An AI Framework for Predicting the Winner of the Grammys | IEEE Conference Publication | IEEE Xplore, ieeexplore.ieee.org/document/10607237. Accessed 5 May 2025.

Appendices

Appendix A: Statistical Code

```
# Duration comparison
t.test(duration_s ~ status, data = combined_data)
```

```
# Danceability over time
ggplot(Grammy_data_W, aes(x = Year, y = rhythm.danceability)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Danceability Trend")

# Key distribution
chisq.test(table(combined_data$status, combined_data$tonal.chords_scale))

# Logistic regression model
model <- glm(status ~ duration_s + popularity + danceability,
        data = combined_data, family = "binomial")
```

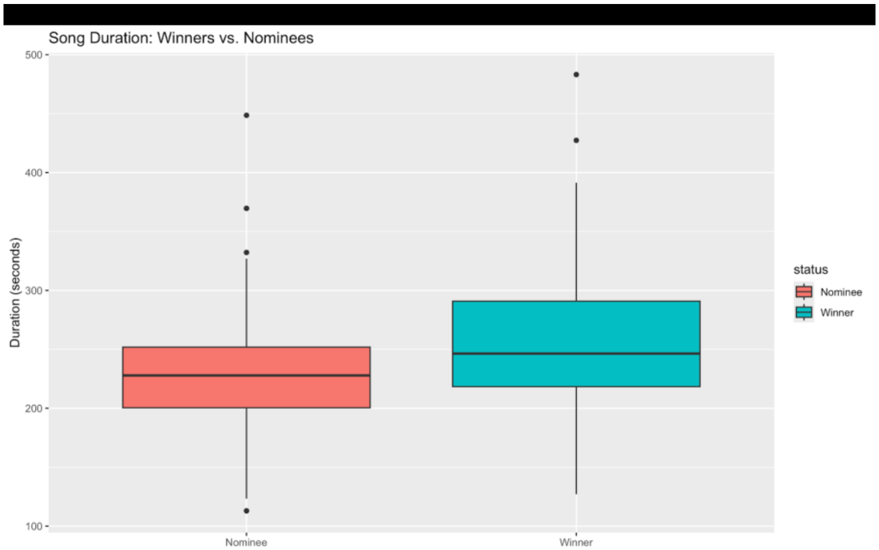
Appendix B: Full Dataset

The complete dataset analyzed in this study includes:

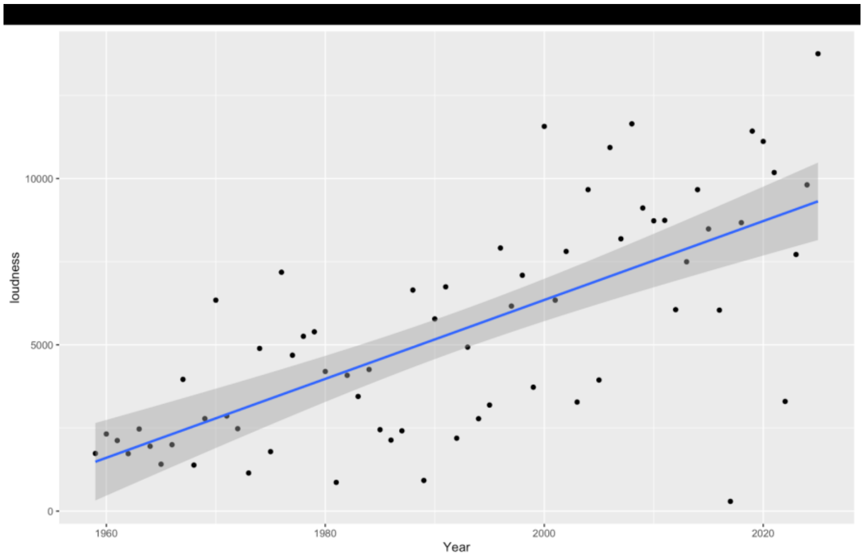
Column Name	Description	Example Values
Track_Name	Song title	"Not Like Us", "Flowers"
Artist	Performing artist	Kendrick Lamar, Miley Cyrus
Year	Release year	2025, 2024
duration_ms	Length in milliseconds	274192, 200600
popularity	Spotify popularity score	88, 86
danceability	Rhythm strength metric	1.17, 1.24
key	Musical key	"B", "A"
mode	Major/minor	"minor", "major"
status	Winner/nominee	1, 0

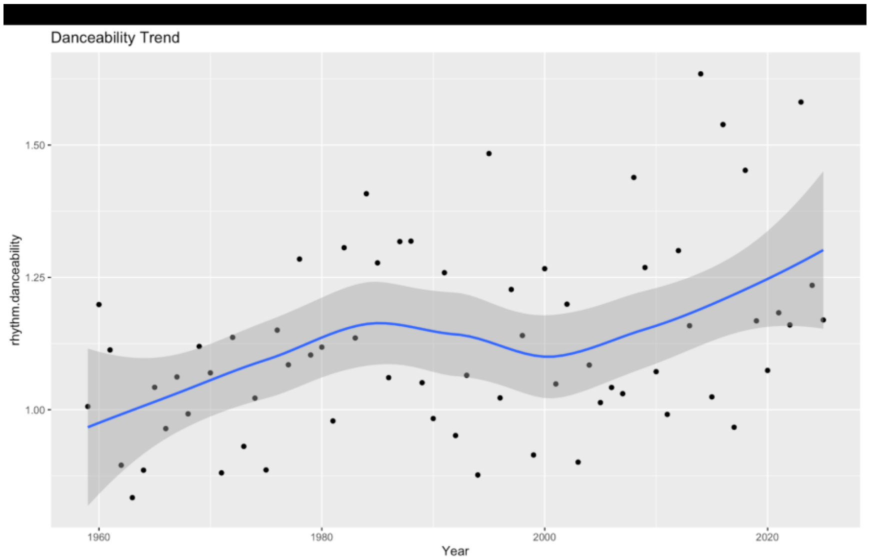
Appendix C: Additional Visualizations

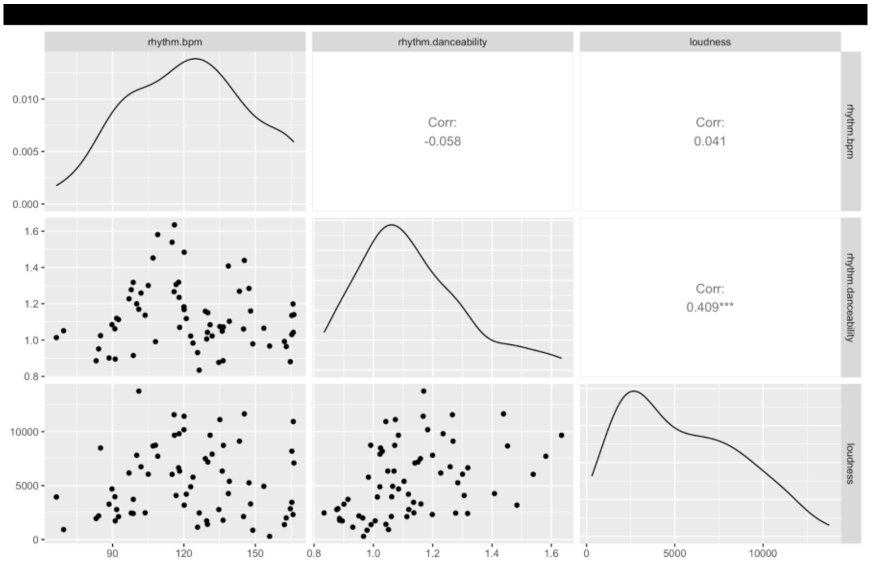
1. **Figure 1:** Boxplot of song duration by winner/nominee status



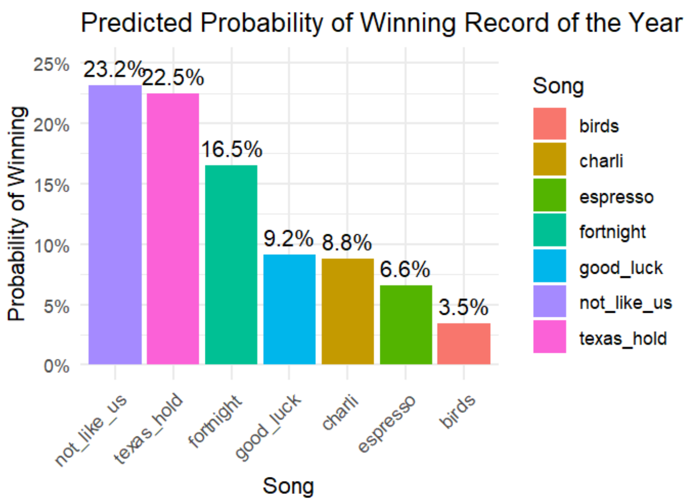
2. **Figure 2:** Graph of audio feature trends over time







3. **Figure 3:** Graph of predictive results from logistic model on 2025 record of the year nominees



Appendix D: Complete Statistical Output

Plot code to analyze "loudness" feature:

```
ggplot(Grammy_data_W, aes(x = Year, y = loudness)) +
  geom_point() +
  geom_smooth(method = "lm")  # Check if loudness increases over time
```

Plot code to analyze "danceability" feature:

```
# Example: Danceability over time
ggplot(Grammy_data_W, aes(x = Year, y = rhythm.danceability)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Danceability Trend")
```

Code to count major/minor keys:

```
Grammy_data_W %>%
  count(chords_scale) %>%
  ggplot(aes(x = chords_scale, y = n)) +
  geom_bar(stat = "identity") +
  labs(title = "Major vs. Minor Keys in Winners")
```

Code to find most common keys:

```
Grammy_data_W %>%
  count(`tonal.chords_key`) %>%
  arrange(desc(n)) %>%
  head(5)
```

Code to conduct Welch Sample T Test:

```
data:  duration_s by status
t = -2.9711, df = 104.04, p-value = 0.003686
95% confidence interval:
 -44.792280  -8.933626
sample estimates:
mean in group Nominee  mean in group Winner
           228.8061             255.6690
```