Unveiling the Dynamics of Music Popularity on Spotify: Temporal Trends and Popularity Drivers

s2772634 **s2880814** **s2816523** **s2808241**

Abstract

The rapid expansion of music streaming platforms and increasing competition among various platforms has created a vast opportunity to understand the track popularity, artist success, and popular songs. We operate as skunkworks team within Spotify investigating the Top 200 global chart dataset (2017 – 2023) consisting of audio features, artist metadata and popularity metrics. Through exploratory data analysis, we examine temporal streaming patterns across years, months, and weeks, while examining whether it's the artist or song audio features that drives the success of the track. We build Random forest, Logistic Regression, SVM, and XGBoost models with accuracies ranging from 50-70%. Our findings contribute to music information retrieval (MIR) research by explaining how multi level feature aggregation improve forecasting in streaming environments, with future work exploring Graph Neural Networks for enhanced prediction in successful collaboration.

1 Introduction

Digital streaming platforms like Spotify [2], YouTube Music [3], and Apple Music [1] have transformed music consumption, as they generate vast behavioral and acoustic data that fuels Hit Song Science (HSS) research [19, 20]. While audio features such as energy, valence, and danceability correlate with popularity [12, 17], success is shaped by more than sound alone—cultural impact [18], collaboration networks [11, 16], and artist identity play critical roles [15]. Despite recognition of social media influence [18] and collaboration patterns [16], very few studies integrate long-term global chart analysis with artist-level features and collaboration networks [15].

We fix this by analyzing Spotify's Global Top 200 charts (2017–2023) through machine learning-driven exploratory analysis. Using audio features, time-aware artist metadata, and a points-based popularity metric, we investigate: **(1) Are there any temporal patterns in what day(s) of the week and/or months experience the most streams?** and **(2) Which factors, artist reputation or audio characteristics, drive song popularity?** Our approach uses evidence-based insights into the evolving dynamics of global music popularity.

2 Data Preprocessing

2.1 Dataset Description

This dataset contains Spotify Top 200 tracks from 2017-01-01 to 2023-05-29, and consists of 651,936 rows and 20 columns. Each track contains features like ranking position, points (popularity proxy) artist name, Nationality, and various Spotify audio features (Danceability, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, and Valence) which we used for our analysis.

2.2 Data Aggregation and Feature Scaling

- **Audio Feature Standardization**:* The audio feature - loudness ranged from -34,475 to 1,509 dB which is an extreme range that could dominate over the model and hence it was normalized by 1000. Post correction it ranged between -34.48 to 1.51 dB ensuring a balanced contribution.
- **Unique Song Aggregation**: * The original dataset contained entries (651,936 rows) comprising multi-row-per-song due to artist collaboration. To transform this into a unique song dataset to preserve meaningful information while eliminating redundancy, we aggregated all the features and grouped them by the song id.

2.3 Target Variable Engineering

In order to investigate what drives popularity of the song, we need a robust measure of success and so we created a 3-class target variable using percentile-based threshold

Success Class	Threshold Points	Percentage of Songs	Distribution
Low	$\leq$ 2,901	70.0%	6,413 songs
Moderate	2,901 - 19,884	20.0%	1,832 songs
Hit	$>$ 19,884	10.0%	916 songs

Table 1: Target Variable Classification Based on Accumulated Chart Points

The thresholds stated in Table 1 are based on the 70th percentile at 2,901 points and 90th percentile at 19,884 points of the points distribution; this would result in balanced classes, reflecting real-world music success patterns but still having enough samples to drive robust modeling. The final classification achieved near exact proportional splits (70.0% / 20.0% / 10.0%) matching our percentile-based design goals.

2.4 Time-Aware Artist Feature Engineering

The features mentioned in 2 were calculated to interpret the artist historical performance before the song release. For each song released at time t, we can calculated:

Feature	Calculation	Interpretation
Artist_Past_Avg_Points	mean(points before t)	Historical performance baseline
Artist_Past_Max_Points	max(points before t)	Peak capability indicator
Artist_Experience	count(songs before t)	Career longevity measure
Artist_Days_Since_Last	t - max(dates before t)	Release frequency pattern
Is_Debut	1 if no previous songs, else 0	First-release effect

Table 2: Time-Aware Artist Feature Engineering

2.5 Collaboration Dynamics

The features listed in 3 were created as they directly address whether popularity is determined by combined artists and audience groups, testing whether collaborative efforts produce a multiplicative rather than an additive success effect

Feature	Description	Purpose
Is_Collaboration	Binary flag (1 if multiple artists, 0 if solo)	Tests synergistic effects: whether artist combinations create popularity beyond individual capabilities
Num_Artists	Count of contributing artists per song	Measures network size impact: whether more artists correlate with increased audience reach

Table 3: Collaboration Feature Engineering

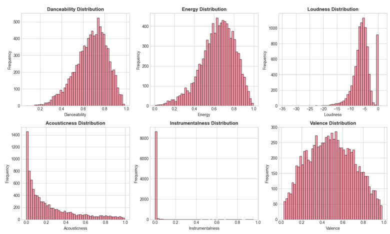
3 Exploratory Data Analysis

3.1 Univariate Analysis

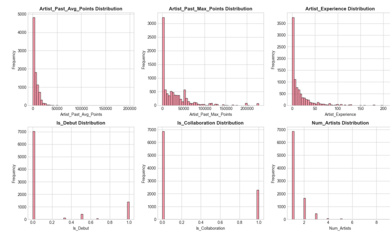
We examined the distribution of all audio and artist features to understand their statistical properties and identify patterns relevant to modeling.

Figure 1a shows the distribution of audio features: Danceability (μ=0.67), Energy (μ=0.64), Loudness (μ=–5.82 dB), Valence (μ=0.49) have a **Normal Distribution**, good for parametric models while Speechiness (Mdn=0.08) & Acousticness (Mdn=0.14) have a **Right-skewed distribution**, showing dominance of vocal electronic tracks. Instrumentalness is heavily concentrated at 0 (75th percentile = 0), almost all charting songs are vocal.

Figure 1b show the distribution of all the artist features. the Features Artist_Past_Avg_Points (Mdn=3,459 vs Mean=5,784) and Artist_Past_Max_Points (Mdn=18,127 vs Mean=29,859) have a very strong **Right Skew Distribution**. Artist_Experience has a Mdn=7, Artist_Days_Since_Last Release has a Mdn=28 indicating active but not highly prolific artists. There is a 80.8% non-debut artists and 25.1% collaborations.



(a) Audio Features

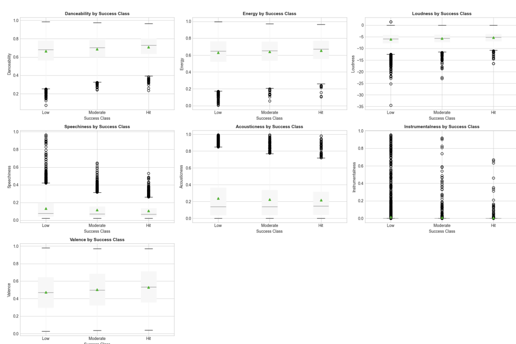


(b) Artist Features

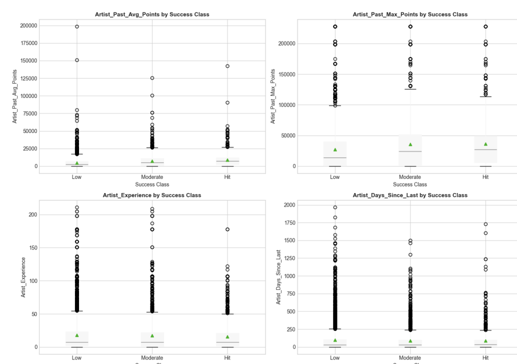
Figure 1: Univariate analysis of Audio features and Artist Features

3.2 Bivariate Distribution

The Bivariate analysis tests our hypothesis of which either the audio properties or artist reputation have stronger associations with success by examining how each of them correlate with the target feature - success class, thereby identifying which of them discriminate between the class low, moderate and High.



(a) Audio Features by Target Class



(b) Artist Features by Target Class

Figure 2: Bivariate analysis: Feature distributions across success classes (Low/Moderate/Hit)

The Audio features in Figure 2a (Appendix) show Hits only show modest lifts: Danceability (+6.5%); Energy (+3.8%); Loudness (+13%, +0.76 dB); and Valence (+12%). Speechiness, Acousticness, and Instrumentalness are slightly lower. But heavy IQR overlap meaning the audio traits barely separate Low vs. Hit songs reflecting mastering norms more than creative differences.

Figure 2b (Appendix) shows the Artist Features emerging as the most potent predictors of success. Hit songs are from artists with much stronger track records, as demonstrated by the 86.8% increase in Artist_Past_Avg_Points (μ=4,855 $\rightarrow$ μ=9,068) and a 35% increase in Artist_Past_Max_Points. This increase suggests that selective, high-quality output is more important than volume, as indicated by the decrease in Artist_Experience by 13.1%. Artists releasing slightly more frequently fare better, as demonstrated by an 8.4% drop in Days_Since_Last.

3.3 Correlation Matrix

The results of the correlation analysis 4 (Appendix) are particularly critical to our approach to modeling. Strong inter-feature correlations, especially those between Energy and Loudness at 0.56, and among artist historical metrics, such as Artist_Past_Avg_Points and Artist_Past_Max_Points at 0.59, point to significant feature redundancy that calls for dimensionality reduction. More importantly, the weak correlations with our target variable-the fact that no audio feature exceeds 0.33 for Valence and the artist metrics show only modest relationships, with Artist_Past_Avg_Points at 0.14-no single feature is dominating predictive power. This directly supports our ensemble modeling strategy and explains the consistent advantages seen in our experiments, where artist-based features tend to always outperform those that are audio-only.

4 Temporal Analysis

4.1 Temporal Streaming Trends

Figure 3 and Table 4 tells us how global events and platform evolution reshaped music consumption patterns during 2017 to 2023. But we excluded 2023 data from temporal analysis since it only contains five months. We use points as a measure of streaming activity.

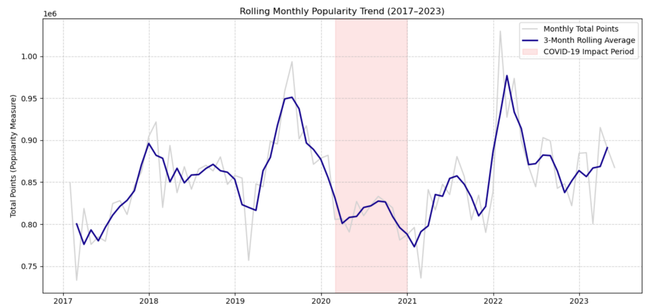


Figure 3: Year-over-Year Average Streaming Points Trend Affected by COVID-19

Year	Total Points	Unique Songs	Collab %	Longevity (Days)
2017	9,818,064	1,551	24.89%	72.9
2018	10,368,011	1,924	29.11%	53.6
2019	10,619,250	1,765	30.55%	53.1
2020	9,792,540	2,004	25.37%	42.3
2021	9,879,331	1,822	24.34%	42.2
2022	10,743,685	1,589	31.08%	32.0

Table 4: Comprehensive temporal statistics showing pandemic disruption across metrics

- **Pre-Pandemic Growth (2017–2019)**: Steady growth to 10.62M points in 2019, driven by platform expansion and rising global adoption.[9]
- **COVID-19 Disruption (2020)**: Sharp 7.8% drop in 2020 (total points),the dataset's sharpest decline, as lockdowns cut commute listening[4]; collaborations fell from 30.55% → 25.37% (Table 4). Podcasts diverted the attention from music to podcast during this phase indicating due to pandemic restrictions in person studio collaborations were limited collaborations.[7]
- **Sluggish Recovery (2021)**: Minimal rebound (+0.9%) was observed despite vaccines coming out for covid, and adoption of remote work and podcast persisted.[21]
- **Post-Pandemic Surge (2022)**: Streaming jumped 8.8% to 10.74M (Table 4), fueled by return-to-office and increased the demand for live music.[9]
- **Accelerating Consumption Cycles**: Song chart longevity collapsed 56%—from 73 days (2017) to 32 days (2022)—driven by TikTok's viral mechanics. Songs now spike rapidly through social media trends but fade quickly, fundamentally shortening the "hit lifespan." This reflects algorithmic discovery replacing traditional radio-driven slow-burn hits.[21]
- **Weekday Patterns**: Weekend engagement peaks at +2.2% (Saturday, Table 8) above weekly average, consistent with leisure-time listening replacing work/commute consumption.[13]

4.2 Consumption Timing: Weekday and Monthly Patterns

The table 8 shows us the streaming patterns by day of weeks shows consistent weekend effects. Saturday has the highest streaming points(8,888,057 total points) 2.2% higher than weekdays average. Sunday show's the second most highest streaming points with +0.6% higher relative average then normal days, while Monday to Thursdays' the streams remain relatively stable showed a relatively low average than weekdays varying between 0% and -0.5% . Streaming increases from Fridays on Spotify we can easily see the relative difference between the weekdays and weekends pattern. Monthly patterns 7 shows that December being the highest ranking streaming months due to holiday and festival season. In mid year summer months also shows elevated engagement between(July – August). February typically indicates lowest monthly total points.January 2022 shows great performance (10,381 average points), possibly due to post holidays consumption. These temporal insights provide the strategic insights for release timing, with weekend releases and specific months offering potential advantages.

5 Learning Methodologies

5.1 Methodology Overview

In order to assess whether the artist reputation or the audio characteristics drive the popularity of a song, we employed a comparative modeling framework using four different machine learning algorithms across three models - Audio Only (Danceability, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Valence); Artist Only (Artist_Past_Avg_Points, Artist_Past_Max_Points, Artist_Experience, Artist_Days_Since_Last', Is_Debut, Is_Collaboration, Num_Artists); and Combined (Audio Features + Artist Features). This multi-algorithm approach was used to ensure that our findings are robust across the different learning paradigms and not artifacts of one single model.

5.2 Algorithm Selection & Rationale

The following four algorithms 5 were selected representing distinct paradigms to ensure an intricate coverage.

Algorithm	Rationale & Learning Paradigm
Random Forest [5]	Ensemble tree-based learner [6] powerful to treat the outliers and non-linear relationships. It handles all the imbalanced classes well and provides interpretable feature importance.
Logistic Regression [14]	Linear baseline model with comprehensible coefficients. It tests whether the relationships are fundamentally linear or require non-linear modeling.
XGBoost [8]	State-of-the-art gradient boosting algorithm. It is benchmark for structured data, excels with complex feature interactions and class imbalance.
Support Vector Machine [10]	Margin-based learner using kernel methods (RBF). It tests the class separability in high-dimensional transformed space.

Table 5: Algorithm Selection Rationale and Learning Paradigms

5.3 Evaluation Strategy and Metrics

The data was split as 80/20 test-train with propotional sampling maintaining the class distribution (70%-Low, 20%-Moderate, 10%-Hit). A 5-fold stratified Cross-Validation was chosen over a single split due to: (1) reduces variation in measurement of performance, (2) provides a strong evaluation across the multiple data partition, (3) stratification preserves the class balance. This was performed only on the training set in order to prevent any data leakage. The metrics used were: (1) Accuracy - overall correctness; (2) Weighted F1-Score - harmonic mean of precision/recall weighted by class support, addresssing all the imbalance. Feature Scaling was also performed: StandardScalar was applied to Linear Regression(distance-based) and Support Vector Machine(scale-sensitive); the tree-methods use raw features(scale-variant).

5.4 Hyperparameter Tuning Strategy

Hyperparameters 9 (Appendix) were selected to: (1) prevent overfitting (dataset: 9,161 songs), (2) handle class imbalance (70-20-10). All algorithms used class_weight='balanced'. Tree models: max_depth=6-10, n_estimators=100. LR: multinomial solver. SVM: RBF kernel, C=1.0. XGBoost: learning_rate=0.1. Values from preliminary grid search. Artist superiority across all configurations confirms finding is not hyperparameter-dependent.

6 Results

Table 6 shows a clear pattern that artist based features are more predictive of song's success than audio features. Artist models outperform Audio models across all algorithms by significant margins. The

	Audio		Artist		Combined	
Algorithm	**Test Acc.**	**F1**	**Test Acc.**	**F1**	**Test Acc.**	**F1**
RF	51.9%	0.550	**53.0%**	**0.567**	59.1%	0.616
LR	43.6%	0.476	**51.3%**	**0.549**	53.8%	0.570
XGB	69.6%	0.579	**70.0%**	**0.628**	69.9%	0.626
SVM	40.7%	0.457	**49.9%**	**0.539**	52.0%	0.560

Table 6: Algorithm Performance Comparison Across Feature Sets

fundamental reasons behind this is that artist related features such as such as historical performance, release patterns, and collaboration patterns are stable indicators and career level trends that strongly shape how well their new tracks do on the charts. In contrast, audio features tend to cluster within a narrow stylistic range (similar loudness, energy levels, and overall structure) for charting songs, thereby making it limiting to meaningfully separate Low, Moderate, and Hit categories.

XGBoost consistently achieves the highest accuracies ($\approx 70\%$) across feature sets, which reflects its strength in handling non-linear interactions and heterogeneous feature spaces. This aligns with the expectation that popularity is influenced by multiple interacting factors rather than any single dominant feature. On the other hand Random Forest shows moderate improvements when moving from Audio to Combined features, but its gains thin out for Artist only features, which outputs declining performance even when the strongest predictors are already present.

Furthermore, the Combined features does not always surpass Artist only performance. Particularly for XGBoost audio features contribute relatively little additional signal once artist reputation is accounted already. This supports the broader interpretation that success on streaming platforms is heavily path dependent: songs perform well largely because they are released by artists with established audience bases, prior chart momentum, or collaborative reach. Whereas their acoustic profiles alone are insufficient to predict high popularity independently.

7 Conclusion and Future work

In 2020, we have seen collaborations suddenly disappear from recording studios, and streaming also dropped by 7.8% 4 of Spotify's streaming points. This wasn't a coincidence, it was the popularity of artist ecosystems breaking. Our analysis reveals that while audio features provide the notes, artists write the music that truly resonates.

Across every algorithm we tested, from Random Forest to XGBoost, the same pattern emerged: songs carrying the weight of artist reputation consistently outperformed those relying solely on audio characteristics. The correlation matrix whispered this truth audio features barely whispered to success (max r=0.33) 4, while artist histories spoke volumes.

But the most telling rhythm emerged from the timeline itself. As song longevity collapsed from 73 days to just 32 between 2017-2022 4, a new reality crystallized: in today's algorithm-driven landscape, artist branding provides the lasting echo that outlives any viral moment. The weekend streaming peaks and holiday surges? Merely surface rhythms to the deeper current of artist influence. The industry's lesson is clear: you can master the perfect beat, but without the artist's story behind it, the music fades quickly. True success lies not in optimizing danceability, but in cultivating the artists who make people want to dance.

While we've shown that the artist reputation drives success, the COVID-19 pandemic revealed a missing piece: as studio collaborations declined, TikTok and Reels became music's new discovery engines. This suggests artist influence now extends into viral social ecosystems. We propose pioneering **Social Aware Success Prediction** using **Cross-Platform Influence Graphs** that map: TikTok virality patterns and challenge propagation Multi-platform artist social connectivity and Temporal relationships between social trends and streaming spikes Using Graph Neural Networks, we'll model how social media provides the initial spark while artist reputation sustains the fire, finally explaining why some tracks break through their 32-day lifespan to become enduring cultural moments.

References

[1] Apple music - web player. *https://music.apple.com/us/new*. [Accessed: 2025-11-19].

[2] Spotify - web player: Music for everyone. *https://open.spotify.com/*. (Accessed: 2025-11-19).

[3] Youtube music. *https://music.youtube.com/*. (Accessed: 2025-11-19).

[4] Luis Aguiar and Joel Waldfogel. Covid-19 and music consumption. *Information Economics and Policy*, 57:100935, 2021.

[5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[7] Sara Fischer. Spotify now hosts over 2.2 million podcasts as platform pivots during pandemic. https://www.axios.com, 2020. Axios Report.

[8] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[9] Christian Hansen. Longitudinal shifts in music spending and consumption during covid-19. *Psychology of Music*, 2022.

[10] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[11] Tristan Herd, M. Graff, and L. Scharff. Modeling collaborative patterns in popular music networks. *Empirical Studies of the Arts*, 2022.

[12] Youngmoo Kim et al. Predicting music popularity on streaming platforms using acoustic features. *Journal of New Music Research*, 49(4):381–395, 2020.

[13] Amanda Krause and et al. Music listening predicted improved life satisfaction in university students during early stages of the covid-19 pandemic. *Frontiers in Psychology*, 13:842239, 2022.

[14] Michael P LaValley. Logistic regression. *Circulation*, 117(18):2395–2399, 2008.

[15] G. Macia-Fernández and colleagues. Artist collaboration networks and global music popularity. *Information Processing & Management*, 2023.

[16] Gabriel P Oliveira, Mariana Santos, Danilo B Seufitelli, Anisio Lacerda, and Mirella M Moro. Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, pages 726–732, 2020.

[17] François Pachet and Pierre Roy. Hit song science. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2008.

[18] Matthew J. Salganik, Peter Dodds, and Duncan Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311:854–856, 2006.

[19] Markus Schedl. The lfm-1b dataset for music retrieval and recommendation. *Proceedings of the ACM International Conference on Multimedia Retrieval*, 2016.

[20] Aaron van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

[21] Bin Ye and Kenneth Law. Cultural consumption and novelty during covid-19 mobility restrictions. *Multimedia Tools and Applications*, 2022.

A Appendix

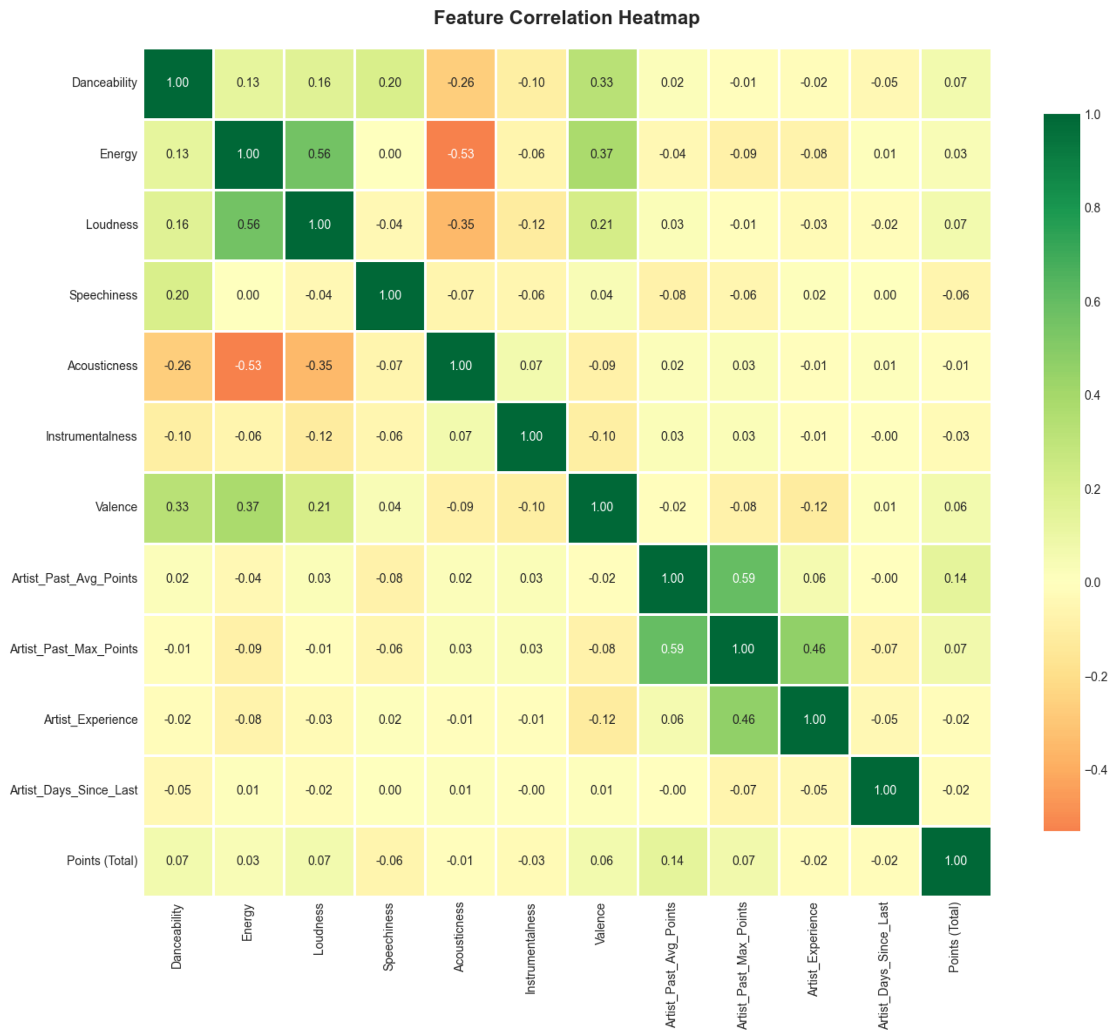


Figure 4: Feature Correlation Heatmap

Month	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Points	5.33M	4.78M	5.18M	4.97M	5.10M	5.01M	5.25M	5.28M	5.05M	5.15M	4.97M	5.15M

Table 7: Monthly patterns

Day	**Total Points**	**Relative to Avg**
Monday	8,694,074	-0.5%
Tuesday	8,691,338	-0.5%
Wednesday	8,706,934	-0.3%
Thursday	8,729,148	+0.0%
Friday	8,751,691	+0.3%
Saturday	8,888,057	+2.2%
Sunday	8,759,639	+0.6%

Table 8: Weekend effect with Saturday peak streaming

Model	Hyperparameters (Search Space, with Optimal in Bold)
Logistic Regression	`multi_class`: '**multinomial**' `solver`: ['liblinear', 'saga', **lbfgs**] `max_iter`: [**1000**] `class_weight`: [**'balanced'**, None]
Random Forest	`n_estimators`: [**100**, 200, 300] `max_depth`: [5, **10**, 20, None] `class_weight`: [**'balanced'**, 'balanced_subsample', None]
XGBoost	`n_estimators`: [**100**, 200, 300] `learning_rate`: [0.01, **0.1**, 0.2] `max_depth`: [3, 4, 5, **6**, 7]
SVM	`C`: **1.0** `kernel`: ['linear', '**rbf**', 'poly', 'sigmoid'] `gamma`: ['**scale**', 'auto', 0.001, 0.01, 0.1] `class_weight`: ['**balanced**', None]

Table 9: Hyperparameter search space and optimal hyperparameters for each model.

B Declaration and Contribution

Declaration:
By submitting this assignment, we confirm that all work presented in this report is our own and has been completed in accordance with the University's Student Conduct and Assessment Regulations. All members contributed equally to the study, and there is no conflict of interest to declare.

Author Contributions:
All members helped with writing, checking, and finalising the report. s2772634 and s2880814 worked together on data preprocessing, including cleaning the dataset, aggregating unique songs, scaling audio features, and creating the artist and collaboration features. s2816523 and s2808241 carried out the exploratory data analysis and produced the univariate, bivariate, correlation, and temporal analysis.For modelling, s2772634 and s2816523 built and tuned the machine learning models across the audio-only, artist-only, and combined feature sets. s2880814 and s2808241 analysed the model outputs, prepared the tables and figures, and explained the key performance results. All members contributed equally to interpreting the findings, ensuring consistency across sections, and completing the report as a group.

C Acknowledgment

This report was completed in accordance with the University of Edinburgh's guidance on the use of Generative AI tools in academic work. Notably, we used ChatGPT to check for any grammatical inconsistencies and certain wordings of phrases, as Latex speller check demonstrated a certain inability to check for grammatical mistakes.