Research Article

Uncovering Audio Features Shaping Popularity in Chart-Topping Songs: A Statistical Approach

Ruoqing Huang

Junior, Walnut High School, 400 N. Pierre Road, Walnut, CA 91789, USA

ABSTRACT

This study examines how audio features influence the popularity of songs on Spotify, focusing on tracks that have appeared on the Billboard Year-End Hot 100 charts. Using data from both Billboard and Spotify, the analysis explores the relationship between features such as Energy, Danceability, Loudness, and Instrumentalness and their impact on Spotify's Popularity scores. Correlation and regression models, ensemble learning, and clustering techniques were applied to uncover patterns and insights. Results show that high-energy, danceable, and loud songs tend to achieve higher popularity, while quieter and more experimental tracks are often less favored. Dimensionality reduction and clustering methods identified groups of songs with distinct audio profiles, highlighting the characteristics associated with varying levels of engagement. This research provides insights into the role of audio features in shaping song popularity, offering useful information for artists, producers, and industry professionals.

Keywords: Audio Features, Spotify Popularity, Billboard Hot 100, Correlation and Regression, Clustering Analysis, Dimensionality Reduction

INTRODUCTION

In the digital age, music has become a thriving and influential industry, reaching vast global audiences through streamlined distribution channels. Platforms like TikTok, Instagram, Spotify, and Apple Music have revolutionized how music is consumed and monetized, providing artists with new opportunities to connect with listeners and create significant economic value (1). Social media continues to amplify the music industry's growth, offering tools that enhance its impact and help artists expand their reach (2). In 2023, global recorded music revenues totaled $28.6 billion, with $19.3 billion generated from streaming—a 10% increase from the previous year (3). These trends underscore the growing demand for music and its ability to influence both culture and finances.

Within this thriving industry, popular songs occupy a central role, attracting attention not only for their commercial success but also for their unique musical qualities. Spotify's Popularity metric, which ranges from 0 to 100, measures listener engagement through streams,

Corresponding author: Ruoqing Huang, E-mail: judy.ruoqing.huang@gmail.com.

Copyright: © 2024 Ruoqing Huang. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received November 30, 2024**; Accepted** December 16, 2024
https://doi.org/10.70251/HYJR2348.24158174

activity, and interactions in real time. Meanwhile, Billboard's Year-End Hot 100 serves as a benchmark for top-performing songs, using a blend of radio airplay, sales, and streaming data. Although these metrics identify which songs succeed, the deeper question is: why are some songs more popular than others? Specifically, what role do their musical features play in their chart performance? This research investigates these questions by linking the audio features of Billboard chart-topping songs to their Spotify Popularity scores, offering insights into the elements driving their success.

To explore this, the study analyzes musical features such as Energy, Danceability, and Loudness to determine how these attributes impact song popularity. A variety of data analysis techniques—including correlation models, regression methods, ensemble learning, dimensionality reduction, and clustering—are applied to uncover both linear and non-linear relationships. Pearson and Spearman correlations highlight direct and rank-based trends, while Lasso and Elastic Net regression address multicollinearity and feature selection. Advanced methods, including Random Forest, Gradient Boosting, Principal Component Analysis (PCA), and ANOVA tests, provide deeper insights into feature importance, grouping patterns, and their combined effects on song performance.

By integrating Billboard and Spotify data, this research offers a comprehensive analysis of how audio features influence song popularity. The findings aim to deliver actionable insights for artists, producers, and marketers, helping them identify the attributes that resonate most with audiences in today's dynamic music industry. Through an examination of musical characteristics like Energy, Danceability, and Loudness, the study sheds light on how these features shape listener preferences and contribute to a song's success.

Data Description

This study uses two primary datasets: Billboard Year-End Hot 100 rankings sourced from Wikipedia and audio feature data collected from Spotify using the Spotify Web API (4).

The Billboard dataset includes the top 100 songs each year from 1946 to 2023, featuring details such as song titles, artists, and chart positions. This information provides a foundation for understanding trends in song rankings over time and serves as a basis for analyzing how musical features influence chart performance.

The Spotify dataset includes several audio features for each Billboard song, such as Energy (a measure of intensity and activity), Danceability (suitability for dancing based on tempo and rhythm), and Tempo (track speed in beats per minute). Other features include Loudness, Instrumentalness, Liveness, Speechiness, Valence, Key, Mode, and Time Signature (see Table 1). These attributes enable a detailed examination of the relationship between musical characteristics and a song's success.

Table 1. Descriptive Statistics of Audio Features Data

Feature	Data Type	Definition	Descriptive Statistics (Mean, Std, Min, Max)
Danceability	Numerical	How suitable a track is for dancing (0.0 - 1.0)	0.62, 0.16, 0, 1
Energy	Numerical	Intensity and activity in a track (0.0 - 1.0)	0.6, 0.2, 0, 1
Instrumentalness	Numerical	Likelihood of being instrumental (0.0 - 1.0)	0.03, 0.14, 0, 0.98
Liveness	Numerical	Detects live performance (0.0 - 1.0)	0.18, 0.14, 0.01, 0.99
Speechiness	Numerical	Presence of spoken words (0.0 - 1.0)	0.07, 0.08, 0, 0.88
Valence	Numerical	Positivity or negativity of a track's emotion (0.0 - 1.0)	0.60, 0.24, 0, 0.99
Loudness	Numerical	Overall volume level in decibels (dB)	-8.59, 3.70, 37.44, -0.81
Tempo	Numerical	Speed of the track (BPM)	119.41, 28.30, 0, 232
Key	Categorical	Musical key (0 = C, 1 = C♯/D♭, etc.)	5.39, 3.69, 0, 11
Mode	Categorical	Modality (1 = major, 0 = minor)	0.63, 0.48, 0, 1
Time Signature	Numerical	Beats per measure (e.g., 3/4, 4/4, etc.)	4, 0.52, 0, 5
Popularity	Numerical	Popularity metric on Spotify (0 - 100)	53.96, 18.59, 0, 92

METHODS AND MATERIALS

This section outlines the data collection process, data processing steps, and the analytical techniques used to evaluate the relationship between Audio Features and Spotify Popularity. The analysis follows a structured, multi-method approach designed to assess correlations, synthesize findings, validate results through hypothesis testing, and explore grouping patterns.

Data Collection

The data for this study was collected from Wikipedia, providing information about the Billboard Year-End Charts Hot 100 Songs, including Year, Rank, Title, and Artist. The Title and Artist were used to search the Spotify Web API for corresponding Popularity and Audio Feature data. Initially, 7,700 entries were expected, but due to data capture issues, the final dataset consists of 6,689 songs with combined Billboard and Spotify data.

Data Preprocessing

Before conducting the analysis, the data underwent preprocessing to ensure that the audio features were suitable for modeling. Since these features are measured on different scales, normalization was performed to bring all features to a comparable scale. This step is particularly important for methods such as Lasso regression, Elastic Net regression, Linear regression, and Principal Component Analysis (PCA), which are sensitive to the scale of input variables.

The normalization process involved using StandardScaler, which standardizes each feature to have a mean of 0 and a standard deviation of 1. This approach ensures that no single feature dominates the models due to its scale, allowing for more accurate interpretations of feature importance and interactions. Additionally, the categorical variables Key and Mode were converted to categorical data types and subsequently one-hot encoded. This transformation created binary columns for each category, which supports compatibility with the analytical models, especially in regression contexts where categorical data cannot be directly incorporated.

Normalized data, with one-hot encoding applied to Key and Mode, was used in all subsequent analyses, ensuring consistency across different methods.

Statistical Analysis. The analysis focused on three types of methods: Linear Relationships Analysis, which studies straightforward trends and connections between audio features and Spotify popularity; Non-Linear Modeling with Ensembles, which explores more complex patterns using tools like Random Forest and Gradient Boosting; and Pattern Exploration and Clustering, which groups songs with similar audio features to uncover how these groupings relate to popularity differences. These methods work together to provide a well-rounded understanding of what drives song success.

Linear Relationships. To assess linear relationships, Pearson's Product-Moment Correlation Coefficient was applied (5). This measures the linear association between two continuous variables, calculated as

$$r = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\Sigma(X_i - \overline{X})^2}\sqrt{\Sigma(Y_i - \overline{Y})^2}}$$

where X_i and Y_i are individual observations, and $\overline{X}$ and $\overline{Y}$ are their respective means. This coefficient quantifies how increases in a feature correspond to increases or decreases in popularity.

Additionally, Spearman's Rank Correlation Coefficient was used to capture rank-based relationships (6), reflecting non-linear trends between features and popularity. It is calculated as

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)}$$

where d_i is the difference between the ranks of two variables and n is the number of observations. This method is particularly useful for detecting monotonic relationships, even when they are not linear.

Following correlation analysis, three regression models were applied. Lasso regression was used for feature selection, applying L1 regularization to shrink some coefficients to zero, effectively selecting key variables. The objective function for Lasso regression is

$$min_\beta(\Sigma_{i=1}^{n}(y_i - X_i\beta)^2 + \lambda\Sigma_{j=1}^{p}|\beta_j|)$$

where β_j are the regression coefficients, λ is the regularization parameter, y_i is the observed value, and X_i are the feature values (7).

Elastic Net regression combines L1 and L2 regularization to handle multicollinearity while allowing for variable selection. The objective function for Elastic Net is

$$min_\beta(\Sigma_{i=1}^{n}(y_i - X_i\beta)^2 + \lambda_1\Sigma_{j=1}^{p}|\beta_j| + \lambda_2\Sigma_{j=1}^{p}\beta_j^2)$$

where λ_1 controls the L1 regularization and λ_2 controls the L2 regularization (8).

Finally, a linear regression model was used as a baseline, capturing how each feature linearly influences Spotify popularity. The model is expressed as $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$, where y is the dependent variable (popularity), $X_1, X_1, \dots, X_p$ are the independent variables (audio features), β_0 is the intercept, $\beta_0, \dots, \beta_p$ are the coefficients, and ϵ is the error term (9).

Together, these correlation and regression methods provided foundational insights into both linear and non-linear relationships between features and popularity, guiding the further steps of our analysis.

Non-Linear Modeling. Random Forest is a tree-based method that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting. The final prediction is given by

$$\hat{y} = \frac{1}{B}\Sigma_{b=1}^{B} T_b(X)$$

where B is the number of trees, $T_b(X)$ is the prediction from tree b, and $\hat{y}$ is the averaged prediction across all trees (10). This method detected the most influential audio features while capturing complex relationships that may not be linearly interpretable.

Gradient Boosting is an ensemble technique that builds models sequentially, where each subsequent model aims to correct the errors made by the previous one. The objective function is minimized as

$$f(x) = \Sigma_{m=1}^{M} \alpha_m h_m(x)$$

where α_m is the weight of the -th model and $h_m(x)$ is a weak learner. These models are combined iteratively to form a strong prediction (11). This technique is particularly effective for complex datasets, as it iteratively refines predictions to reduce errors. Gradient Boosting proved valuable for highlighting nuanced feature contributions and identifying interactions that Random Forest alone might overlook.

Combining the results from these ensemble methods provided a comprehensive view of non-linear relationships among audio features and their contributions to Spotify popularity. This analysis offered a deeper understanding of feature importance and interactions, guiding further exploration in Pattern Exploration and Clustering.

Pattern Exploration. The analysis incorporated visualization, dimensionality reduction, clustering, and hypothesis testing to examine how combinations of features influence Spotify popularity, capturing both individual and collective effects.

A heatmap visually represents data as a color-coded matrix, with each cell's color intensity indicating the value at a specific intersection in the data. Each value is normalized to a range (e.g., 0 to 1) using

$$M'_{ij} = \frac{M'_{ij} - \min(M)}{\max(M) - \min(M)}$$

where min(M) and max(M) are the dataset's minimum and maximum values. The normalized values are then mapped to a color gradient, such as blue to red, to visually highlight patterns based on relative intensities (12).

Principal Component Analysis (PCA) was employed for dimensionality reduction, transforming features into uncorrelated principal components that maximize data variance and simplify complex data while retaining key feature combinations (13). PCA accomplishes this by solving the eigenvalue problem, $Sv_i = \lambda_i v_i$, where S is the covariance matrix of the data, λ_i represents the eigenvalue, and v_i is the corresponding eigenvector (principal component). By focusing on the components with the largest eigenvalues, PCA highlights dominant patterns that drive song popularity.

To identify natural groupings of songs based on similar audio attributes, K-means clustering was applied. This algorithm minimizes the sum of squared distances between data points and their assigned cluster centroids (14), calculated as

$$\min \Sigma_{i=1}^{n} \Sigma_{k=1}^{K} r_{ik} \|x_i - \mu_k\|^2$$

where x_i is a data point, μ_k is the cluster centroid, and r_{ik} is a binary indicator denoting whether point i is assigned to cluster k. K-means clustering grouped songs based on their musical attributes, providing insights into how songs with similar features perform in terms of Spotify popularity.

Hierarchical clustering organizes data into nested clusters through either merging or splitting groups, using distance measures like Euclidean distance, where the distance between two points and is defined as

$$d(x, y) = \sqrt{\Sigma_{i=1}^{n} (x_i - y_i)^2}$$

Linkage criteria determine inter-cluster distances: single linkage uses the minimum, complete linkage the maximum, and average linkage the mean distance. Ward's linkage, which minimizes within-cluster variance, calculates inter-cluster distance as

$$d(C_i, C_j) = \frac{|C_i||C_j|}{|C_i| + |C_j|} \left\| \bar{x}_{C_i} - \bar{x}_{C_j} \right\|^2 \qquad (15; 16).$$

Hierarchical clustering can follow agglomerative (bottom-up) or divisive (top-down) approaches, either merging individual points into larger clusters or progressively splitting one large cluster into smaller groups.

To evaluate whether significant differences existed between predefined groups of Spotify popularity scores, ANOVA (Analysis of Variance) was applied (17). ANOVA compares group means to test if they differ significantly, calculating the F -statistic as $F = \frac{\text{Between} - \text{group variance}}{\text{Within} - \text{group variance}}$.

A high F statistic suggests that group means are not all drawn from the same population. If significant differences were detected, Tukey's Honest Significant Difference (HSD) test was used post-hoc to identify specific group differences. The Tukey HSD test statistic is calculated as

$$q = \frac{\overline{x_\iota} - \overline{x_\jmath}}{SE}$$

where $\overline{x_\iota}$ and $\overline{x_\jmath}$ are the group means and *SE* is the standard error (18).

These combined methods of dimensionality reduction, clustering, and hypothesis testing integrated findings from previous steps, highlighting both the individual impact of each audio feature and their collective influence on Spotify popularity.

RESULTS

Correlation and Regression

The correlation and regression analysis focused on examining the relationships between Audio Features and Spotify Popularity. Both Pearson's correlation and Spearman's correlation were used to assess linear and rank-order relationships (see Table 2). Among the features analyzed, Energy and Loudness showed the strongest positive correlations with Spotify Popularity, suggesting that tracks with higher values in these features tend to be

Table 2. Coefficient of Correlation Models for Spotify Popularity

Feature	Pearson Correlation (r)	Pearson p-value (p)	Spearman Correlation (ρ)	Spearman p-value (p)
Danceability	0.1969	1.94E-59	0.1758	1.51E-47
Energy	**0.277**	4.16E-118	**0.2407**	9.47E-89
Instrumentalness	-0.2247	2.65E-77	-0.1683	1.11E-43
Liveness	-0.0681	2.48E-08	-0.0386	0.002
Speechiness	0.1024	4.57E-17	0.1552	2.52E-37
Valence	-0.0297	0.015	-0.1249	1.15E-24
Loudness	**0.285**	3.37E-125	**0.4121**	1.52E-272
Tempo	0.0356	0.004	0.0364	0.003
Time Signature	-0.2247	2.65E-77	-0.1683	1.11E-43
Key (0 = C)	-0.0261	0.033	-0.0237	0.053
Key (1 = C♯/D♭)	0.0492	5.68E-05	0.0474	1.05E-04
Key (2 = D)	0.0026	0.831	0.0039	0.749
Key (3 = D♯/E♭)	-0.0512	2.79E-05	-0.0444	2.84E-04
Key (4 = E)	0.0052	0.67	0.002	0.871
Key (5 = F)	-0.0063	0.609	-0.0028	0.822
Key (6 = F♯/G♭)	0.0317	0.01	0.0347	0.004
Key (7 = G)	-0.0293	0.016	-0.0315	0.01
Key (8 = G♯/A♭)	0.0336	0.006	0.0329	0.007
Key (9 = A)	-0.0328	0.007	-0.035	0.004
Key (10 = A♯/B♭)	-0.0263	0.032	-0.0302	0.014
Key (11 = B)	0.0467	1.31E-04	0.0458	1.78E-04
Mode (0 = Minor)	0.1238	3.07E-24	0.1225	9.09E-24
Mode (1 = Major)	-0.1238	3.07E-24	-0.1225	9.09E-24

more popular. In contrast, Instrumentalness exhibited a negative correlation, implying that songs with fewer vocal components are less popular.

Loudness has the strongest positive correlation with Spotify Popularity among all features, with Pearson correlation $r = 0.285$ and Spearman correlation $\rho = 0.4121$, both highly significant ($p < 0.0001$). This suggests that louder songs are generally more popular.

Energy is also strongly positively correlated with popularity ($r = 0.277$ for Pearson and $\rho = 0.2407$ for Spearman), indicating that more energetic songs tend to rank higher in popularity.

Instrumentalness has a strong negative correlation ($r = -0.2247$ for Pearson and $\rho = -0.1683$ for Spearman), showing that songs with higher instrumental content tend to be less popular.

Mode has a significant impact on popularity: For Minor (Mode = 0), there is a positive correlation with Spotify Popularity, meaning that songs in a minor key tend to be more popular. Conversely, for Major (Mode = 1), there is a negative correlation with popularity, indicating that songs in a major key tend to be less popular.

Key shows smaller, but still significant, correlations with Spotify Popularity. Some keys, such as C♯/D♭ (Key = 1) and B (Key = 11), are positively correlated with popularity, suggesting that songs in these keys tend to be slightly more popular. On the other hand, keys like D♯/E♭ (Key = 3) and A (Key = 9) are negatively correlated with popularity, meaning that songs in these keys tend to be less popular. However, these correlations are generally weaker than those for other audio features such as Loudness and Energy, which have a much stronger influence on popularity.

The regression models further reinforced these findings. Lasso regression and Elastic Net regression identified Energy, Loudness, and Danceability as key predictors of popularity. These features had significant coefficients across multiple models, confirming their predictive power. Linear regression also pointed to the influence of these features, while others, such as Valence and Tempo, had a smaller impact. Overall, the correlation and regression analyses provided strong evidence that features like Energy are crucial in determining a song's popularity on Spotify.

In the regression analysis (see Table 3a and Figure 1), Energy emerged as the strongest predictor of Spotify Popularity across all three models. The coefficients for Energy were 3.8423 in Lasso, 3.5092 in Elastic Net, and 3.9718 in linear regression, indicating that more energetic songs tend to be significantly more popular.

Alongside Energy, Danceability emerged as the second most influential feature, with coefficients of 2.4334 in Lasso, 2.3362 in Elastic Net, and 2.4983 in linear regression. These results suggest that songs with higher danceability tend to perform well, though their influence is slightly weaker than Energy.

Instrumentalness consistently showed negative coefficients, indicating that songs with more instrumental content tend to be less popular. Similarly, Mode (1 = Major) had negative coefficients across all models, suggesting that songs in a major key are generally less popular than those in a minor key. The impact of Key was

Table 3a. Coefficient of Regression Models for Spotify Popularity

Feature	Lasso Coefficient (β)	Elastic Net Coefficient (β)	Linear Regression Coefficient (β)
Danceability	2.4334	2.3362	2.4983
Energy	**3.8423**	**3.5092**	**3.9718**
Instrumentalness	-3.6291	-3.4821	-3.7245
Liveness	-1.0847	-1.0762	-1.1452
Speechiness	-0.1606	-0.1805	-0.4467
Valence	-1.7667	-1.5255	-2.0298
Loudness	1.28	1.552	1.1749
Tempo	0.7327	0.7497	0.8597
Time Signature	1.0986	1.1323	1.1954
Key (1 = C♯/D♭)	0.0086	0.3884	1.1749
Key (2 = D)	0.00E+00	0.2579	1.1867
Key (3 = D♯/E♭)	-0.3242	-0.7324	-2.8251
Key (4 = E)	0.00E+00	0.00E+00	-0.0619
Key (5 = F)	0.00E+00	0.00E+00	0.0611
Key (6 = F♯/G♭)	0.00E+00	0.1599	1.2629
Key (7 = G)	0.00E+00	-0.2153	-0.5469
Key (8 = G♯/A♭)	0.6343	0.8081	2.3501
Key (9 = A)	-0.7577	-0.804	-1.6011
Key (10 = A♯/B♭)	-0.241	-0.5214	-1.4988
Key (11 = B)	0.00E+00	0.4085	1.278
Mode (1 = Major)	-2.3885	-2.1466	-2.868

Note: β represents the regression coefficient for each feature in the model.

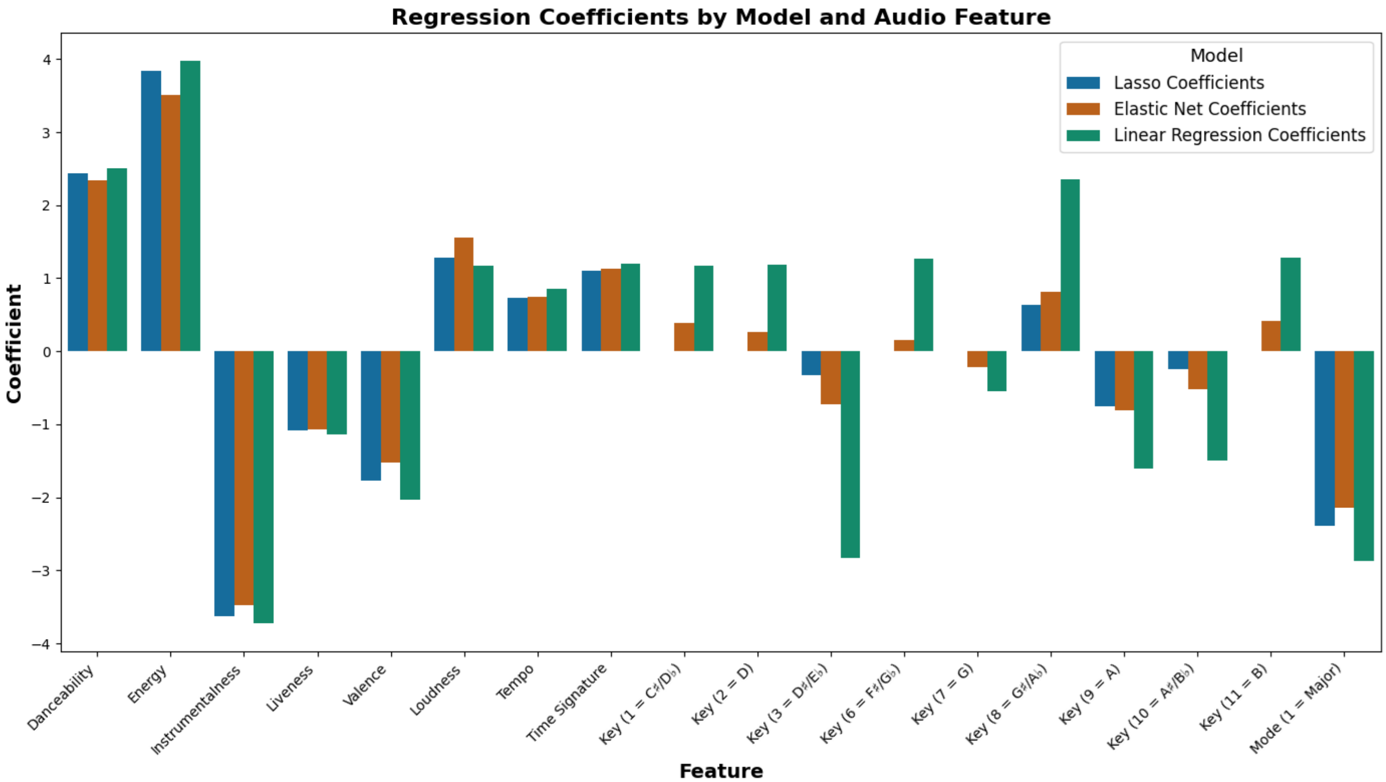

Figure 1. Regression Coefficients by Model and Audio Feature.

less significant, with some keys showing small positive or negative associations with popularity, such as Key (1 = C♯/D♭) and Key (11 = B) having positive associations, while Key (3 = D♯/E♭) had a negative association.

To avoid multicollinearity, Key (0 = C) and Mode (0 = Minor) were dropped, allowing the remaining categories to be interpreted relative to these baselines. Randomness was controlled by setting a random_state in the train-test split and in Lasso and Elastic Net models to ensure consistency in the results.

The model performance was similar across all three approaches (see Table 3b), with linear regression achieving the best performance with an R^2 score of 0.2006 and an MSE of 282.7022. These results show that approximately 20% of the variance in Spotify Popularity is explained by the selected audio features, with Energy, and Danceability being the most influential factors.

Table 3b. Performance of Regression Models for Spotify Popularity

Performance Metrics	Lasso	Elastic Net	Linear Regression
***R*² Score**	0.2003	0.1996	**0.2006**
MSE	282.8025	283.0441	**282.7022**

Both the correlation and regression analyses identified Energy, Danceability, and Loudness as key predictors of Spotify Popularity, though their relative importance varied between the two approaches. In the correlation analysis, Loudness had the strongest linear relationship with popularity, followed by Energy and Danceability. In contrast, the regression analysis revealed Energy as the strongest predictor, followed by Danceability and Loudness. Despite these linear relationships, the correlation coefficients were moderate, and the regression models explained only about 20% of the variance in popularity. This suggests that non-linear interactions between audio features could play a more significant role in determining a song's success. To explore these complexities, ensemble methods will be employed, as they can model both linear and non-linear dynamics, providing a more comprehensive understanding of the factors driving popularity.

Ensemble Modeling

To capture non-linear relationships between audio features and Spotify popularity, Random Forest and Gradient Boosting models were applied. These ensemble methods provided insights into feature importance and revealed complex interactions, enhancing our understanding of factors driving popularity.

As shown in Table 4a, Loudness emerges as the most influential feature in both models, with particularly high importance in Gradient Boosting (α = 0.4165). This suggests that louder songs may be more popular on Spotify, likely due to their ability to capture attention and fit high-energy listening preferences. Other features, such as Valence (a measure of musical positivity), Instrumentalness, and Energy, also hold significant importance, indicating that upbeat and dynamic qualities are associated with higher popularity.

In contrast, features like Time Signature, Key, and Mode hold relatively low importance scores, implying minimal impact on popularity compared to more expressive attributes like Loudness and Danceability. This highlights that listeners may prioritize nuanced musical qualities over structural aspects when it comes to popularity.

Performance metrics in Table 4b reveal that the R^2 score for Random Forest is 0.3409, explaining approximately 34% of the variability in Spotify popularity, while Gradient Boosting has a slightly lower R^2 of 0.3035. Mean Squared Error (MSE) values, 233.0692 for Random Forest and 246.2963 for Gradient Boosting, indicate that Random Forest offers slightly better predictive accuracy.

While the previous analyses highlighted the influence of individual audio features on Spotify popularity, they do not capture the collective impact of multiple features acting in concert. To address this, clustering techniques emerges. This approach identifies natural groupings of songs with similar profiles, providing a more comprehensive view of how combined attributes influence popularity.

Clustering Patterns

To understand how combinations of audio features influence Spotify popularity, a heatmap is used to visualize correlations among features, highlighting patterns that may impact popularity. Principal Component Analysis (PCA) is then applied to simplify the dataset while preserving key variations. Clustering algorithms, such as K-means and hierarchical clustering, group songs with similar feature profiles, revealing how sets of audio characteristics collectively relate to popularity. The stability and interpretability of these clusters are validated to ensure that the identified patterns are meaningful and reliable. This approach provides deeper insights into the interactions between audio features and listener preferences.

Table 4a. Feature Importance of Ensemble Models for Spotify Popularity

Feature	Random Forest Importance (α)	Gradient Boosting Importance (α)
Danceability	0.1048	0.1028
Energy	0.1123	0.1003
Instrumentalness	0.0929	0.1201
Liveness	0.0908	0.0445
Speechiness	0.0847	0.0252
Valence	0.1126	0.1188
Loudness	**0.2354**	**0.4165**
Tempo	0.0851	0.0394
Time Signature	0.0065	0.0048
Key (1 = C♯/D♭)	0.0067	0.002
Key (2 = D)	0.0056	6.06E-04
Key (3 = D♯/E♭)	0.0054	7.14E-04
Key (4 = E)	0.0037	0.0022
Key (5 = F)	0.0048	2.79E-04
Key (6 = F♯/G♭)	0.0056	0.00E+00
Key (7 = G)	0.0046	1.93E-04
Key (8 = G♯/A♭)	0.0069	1.67E-04
Key (9 = A)	0.0041	0.00E+00
Key (10 = A♯/B♭)	0.0056	0.0015
Key (11 = B)	0.0055	0.0025
Mode (1 = Major)	0.0058	0.0015

Note: α represents the weight of importance.

Table 4b. Performance of Ensemble Models for Spotify Popularity

Performance Metrics	Random Forest	Gradient Boosting
***R*² Score**	**0.3409**	0.3035
MSE	**233.0692**	246.2963

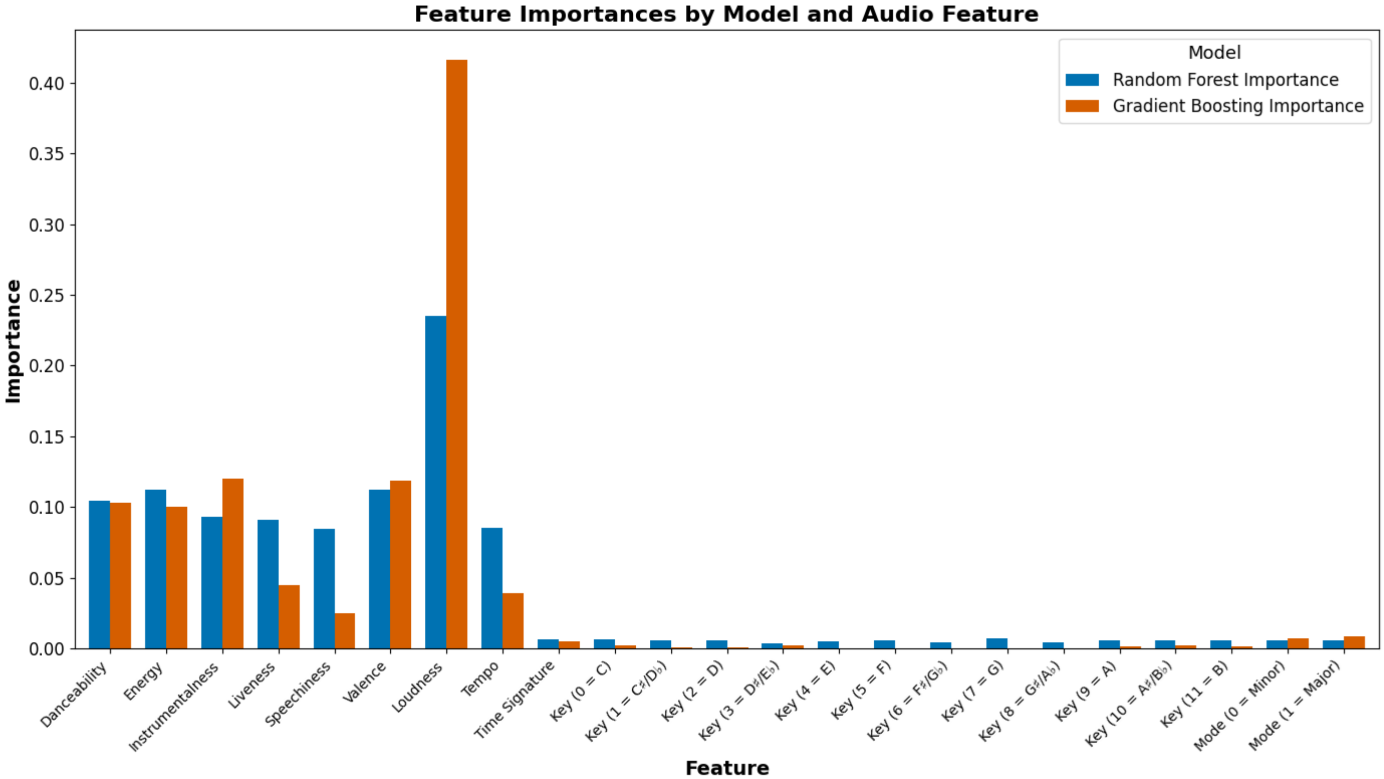

Figure 2. Feature Importances by Model and Audio Feature.

The heatmap shown in Figure 3 visualizes normalized results across various analytical methods, including Pearson and Spearman correlations, Lasso, Elastic Net, and linear regression coefficients, as well as feature importance scores from Random Forest and Gradient Boosting models. Normalizing these results ensures consistent comparison across methods, making it possible to identify which audio features exert the strongest influence on popularity.

Loudness and Energy emerge as the most influential features across methods, displaying high positive values that highlight their strong correlation with popularity. Notably, Loudness reaches values of 3.99 in Gradient Boosting and 2.96 in Spearman, indicating its significant role in driving listener engagement. Similarly, Energy shows consistently high positive values, including 2.54 in Lasso and 2.12 in Pearson correlation, underscoring its importance as a key factor in popular music.

Conversely, Instrumentalness and Mode (1 = Major) display consistent negative associations across methods. Instrumentalness has values such as -2.4 in Elastic Net and -2.16 in Pearson, suggesting that songs with less instrumental content may resonate more with listeners. Mode (1 = Major) also shows a prominent negative effect, seen in Spearman (-1.22) and Lasso (-1.56), implying that songs in minor keys might be more popular. Valence similarly shows negative values across methods (e.g., -1.23 in Spearman), suggesting that emotional positivity does not directly correlate with popularity.

This heatmap provides a foundational understanding of these feature relationships, setting the stage for dimensional reduction and further analysis with PCA.

The PCA analysis of audio features reveals key dimensions underlying the dataset's structure. The scree plot (Figure 4) shows that the first five principal components (PC1–PC5) capture approximately 66% of the variance, with PC1 and PC2 alone accounting for nearly 35%. This indicates that much of the variation in audio features can be summarized by a few principal components, making them central to understanding the dataset's structure.

The PCA loading table (see Table 5) and biplot (Figure 5) provide insights into how specific features contribute to each component. PC1 has high loadings for Energy (0.5872), Danceability (0.4385), and Loudness (0.4151), suggesting it represents an intensity dimension. This alignment indicates that songs scoring high on PC1 likely emphasize energy and loudness. In contrast, PC2 is dominated by Valence (0.6558) with a strong negative loading for Loudness (-0.5602), indicating a possible mood-related dimension that contrasts positivity with loudness.

Further components capture more specific aspects; PC3 emphasizes Tempo (0.6181) and Liveness (0.5106), suggesting a dimension related to performance style. Components PC4 and PC5 are characterized by high loadings for Instrumentalness, potentially representing instrumental versus vocal qualities.

Overall, this PCA analysis suggests that the primary dimensions—intensity, mood, and performance—capture the core structure within the audio features. By clustering songs on the principal components or key audio features, K-means Clustering and Hierarchical Clustering will help segment songs into natural groupings, revealing distinct musical categories or trends and providing insight into relationships among clusters.

The optimal number of clusters is determined using both the Elbow Method and the Silhouette Score (19, 20).

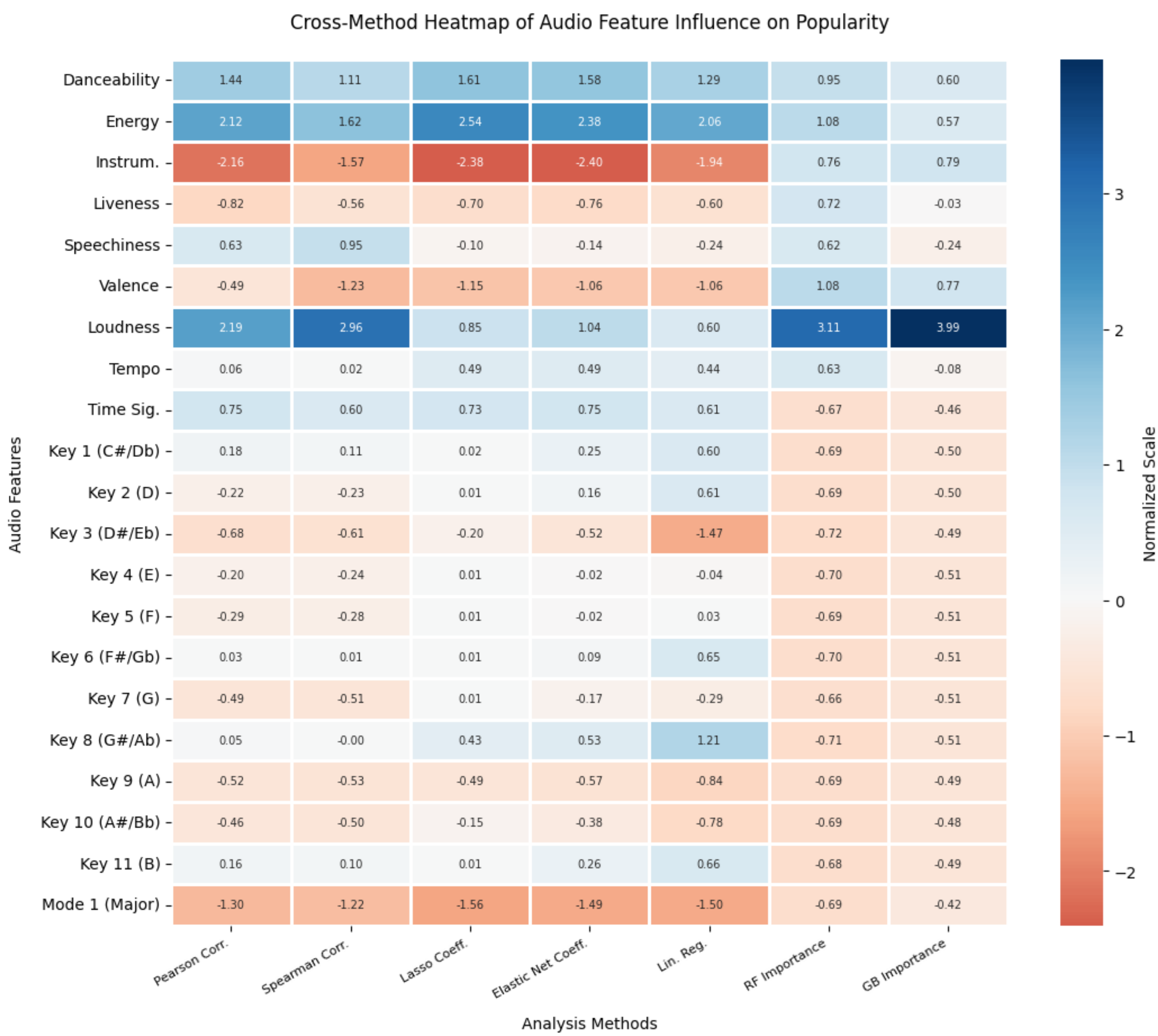

Figure 3. Cross-Method Heatmap of Audio Feature Influence on Popularity.

Table 5. PCA Loading Result for Audio Features

Feature	PC1	PC2	PC3	PC4	PC5
Danceability	**0.4385**	0.1952	-0.4705	0.0676	0.0346
Energy	**0.5872**	-0.0284	0.275	-0.0308	0.1306
Instrumentalness	-0.1266	0.1289	0.1139	0.5656	0.7932
Liveness	0.0474	-0.0641	0.5106	0.6203	-0.5092
Speechiness	0.3079	-0.2878	-0.1536	0.2028	-0.0475
Valence	0.2469	0.6558	0.1243	-0.0538	-0.0894
Loudness	0.4151	-0.5602	0.0363	-0.0442	0.1531
Tempo	0.0968	0.0515	0.6181	-0.4729	0.2358
Time Signature	0.3189	0.3326	-0.0614	0.1396	-0.0574
Key (1 = C♯/D♭)	0.0209	-0.0062	-0.021	-6.12E-04	-3.72E-04
Key (2 = D)	-0.0056	0.0028	0.006	-0.0033	0.0022
Key (3 = D♯/E♭)	-0.0096	0.0012	0.0034	-0.0024	0.002
Key (4 = E)	-0.0032	0.0017	0.0064	8.92E-04	-0.0061
Key (5 = F)	-0.0041	0.0045	-7.43E-04	-0.0013	0.0031
Key (6 = F♯/G♭)	0.0084	-0.0035	-0.0018	0.0062	0.0033
Key (7 = G)	-0.0038	2.93E-04	0.0129	0.0136	-0.0128
Key (8 = G♯/A♭)	0.0042	-0.0011	-0.0087	-0.0011	0.0028
Key (9 = A)	-0.0035	0.0079	0.0075	-0.0118	0.0098
Key (10 = A♯/B♭)	-0.006	2.75E-04	-0.0093	0.0018	-0.0046
Key (11 = B)	0.0138	-0.0029	-0.0013	0.0018	-0.0085
Mode (1 = Major)	-0.0675	0.0155	0.046	-0.0154	-0.0199
Explained Variance (%)	*19.2*	*15.56*	*11.81*	*9.9*	*9.65*

Note: PC – Principal Component.

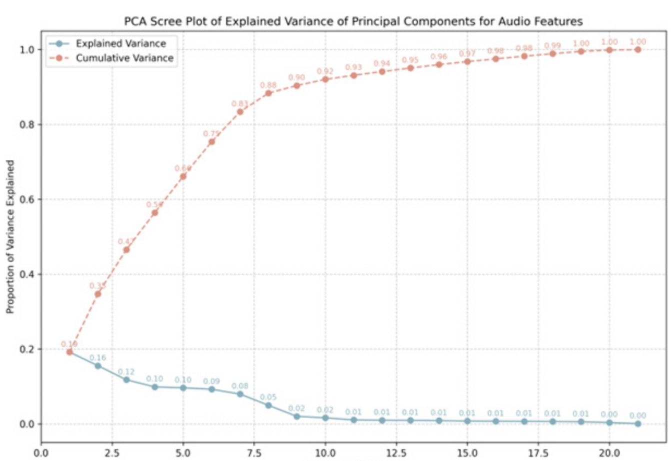

Figure 4. Explained Variance by Principal Component.

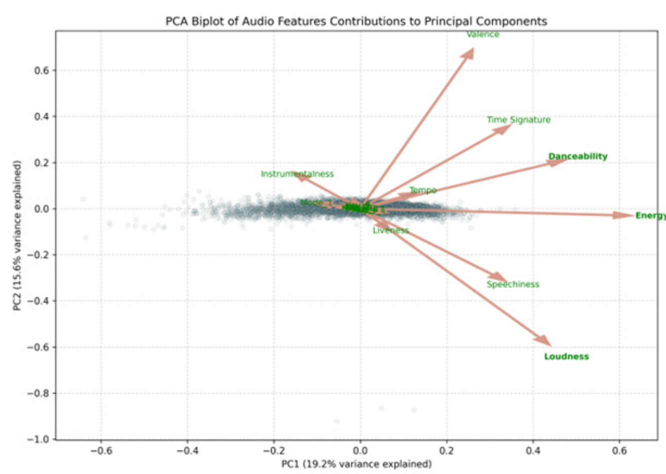

Figure 5. PCA Biplot of Audio Features Contributions to Principal Components.

$K = 3$ is chosen as it provides a balanced solution: the Elbow plot shows diminishing WCSS reduction beyond $K = 3$, while the Silhouette Score remains relatively high, indicating well-separated clusters (see Figure 6).

With K set to 3, K-means clustering was applied for the dataset's principal components (PC1 and PC2), derived from our previous PCA. The resulting clusters are well-separated in PCA space, validating the choice of three clusters (see Figure 7).

Following the insights gained from K-means clustering, which segmented the songs into three distinct groups, Hierarchical Clustering was applied with the same number of clusters, based on the PCA results. This approach explored the nested structure of the dataset, revealing how songs group at various levels of similarity (see Figure 8).

A heatmap (see Figure 9) was created to visualize and compare normalized mean values for Popularity and top 3 audio features in PC1—Loudness, Danceability, and Energy—across the identified clusters from both K-means and hierarchical clustering methods. The heatmap clearly illustrates the distinct characteristics of each cluster. For example, clusters with high normalized values for Loudness and Energy tend to correlate with higher Popularity, while clusters with lower values for these features generally correspond to lower popularity levels. This comparative heatmap analysis reinforces the observations from the individual clustering methods and provides a consolidated view of how audio features vary by cluster. These differences underscore the role

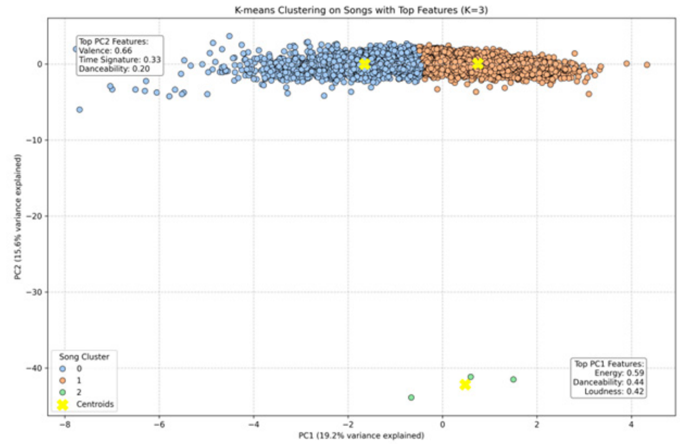

Figure 7. K-means Clustering on Principal Components.

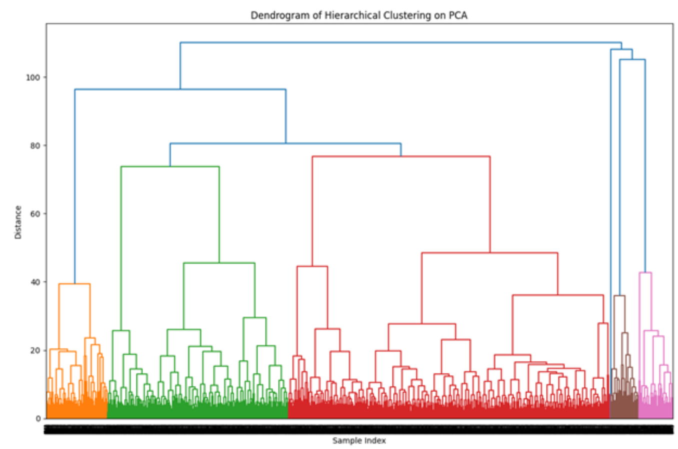

Figure 8. Dendrogram of Hierarchical Clustering on PCA.

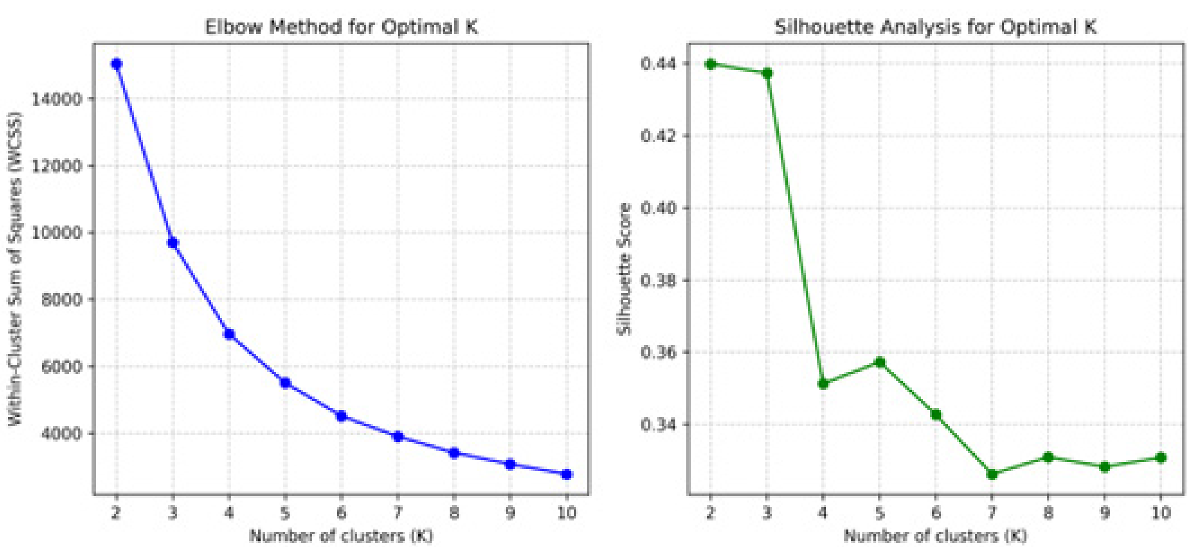

Figure 6. Optimal Number of Clusters.

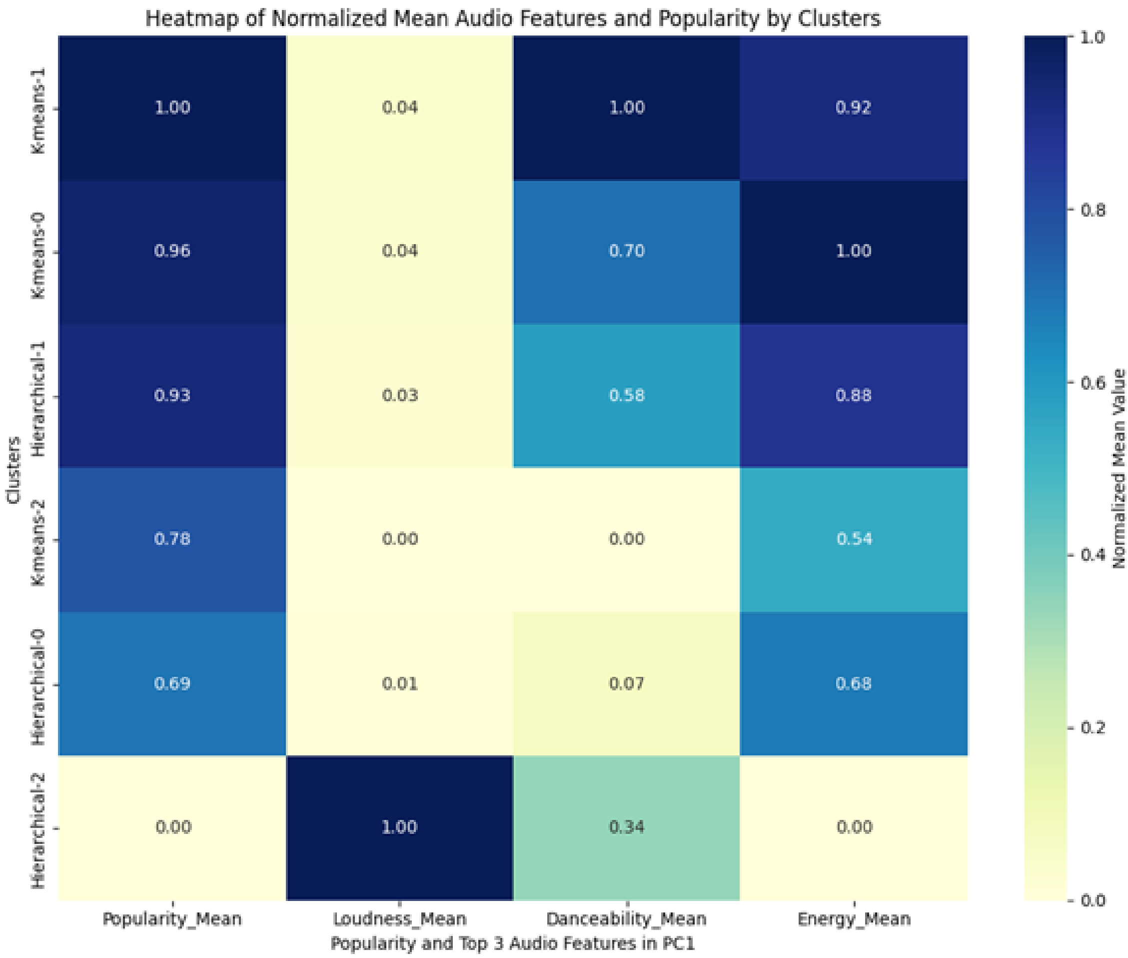

Figure 9. Dendrogram of Hierarchical Clustering on PCA.

of specific musical elements in influencing popularity, offering insights into how different styles or moods appeal to listeners.

Hierarchical Cluster 2 stands out with a Loudness mean of 1.0, yet it has the lowest Popularity. This suggests that while Loudness generally correlates with higher Popularity, extreme Loudness alone may not attract listeners, highlighting a nuanced relationship between loudness and appeal.

To determine whether song popularity significantly differs across clusters identified by K-means and Hierarchical Clustering, a one-way Analysis of Variance (ANOVA) test was performed. This analysis evaluates whether the mean popularity scores vary across clusters, which helps confirm if the songs were successfully grouped based on distinct popularity profiles. For each clustering method, popularity scores were grouped by cluster, and the F-statistic and p-value were calculated to check if the differences in means were statistically significant (see Table 6).

The results of the ANOVA test for the K-means clusters yielded an F-statistic of $F = 265.515$ and a p -value of $p < 0.0001$, indicating a significant difference in popularity across clusters. Similarly, for the Hierarchical Clustering, the ANOVA test produced an F-statistic of $F = 192.1302$

Table 6. ANOVA Test Results for Popularity Across K-means and Hierarchical Clusters

Clustering Method	F-statistic	P-value
K-means	265.515	1.09E-111
Hierarchical	192.1302	7.39E-82

with a p -value of p < 0.0001. Both tests show statistically significant differences, suggesting that both clustering techniques successfully identified clusters with distinct Popularity profiles (Figure 10a and 10b).

Following the ANOVA test, Tukey's Honest Significant Difference (HSD) test was used to determine which specific pairs of clusters differed significantly in song popularity. This post-hoc test identifies cluster pairs with statistically significant differences in popularity, offering a more detailed understanding of the clustering results.

Tukey's HSD test was applied to the popularity scores within each pair of clusters. This test compares the mean popularity difference between each pair and calculates a confidence interval for these differences, this approach determines whether they are statistically significant at a 95% confidence level. (i.e., $p < 0.05$).

The Tukey's HSD results for the K-means clusters reveal significant differences between all cluster pairs. The mean Popularity difference between Cluster 0 and Cluster 1 is 2.4798 ($p < 0.05$), while the difference between Cluster 0 and Cluster 2 is -10.6656 ($p < 0.001$). The largest difference was observed between Cluster 1 and Cluster 2, with a mean difference of -13.1453 ($p < 0.001$), indicating a pronounced distinction in Popularity (see Figure 11a) (Table 7).

For the Hierarchical Clustering, Tukey's HSD results indicate significant differences between Cluster 0 and Cluster 1 (Δ Mean = 14.0084, $p < 0.001$) and between Cluster 1 and Cluster 2 (Δ Mean = -54.5591, $p < 0.001$). These results suggest that, especially in the hierarchical clusters, distinct levels of popularity are present, further validating the effectiveness of this clustering method (see Figure 11b).

The significant results from both ANOVA and Tukey's HSD tests provide strong evidence that the clusters generated by both K-means and Hierarchical Clustering contain distinct Popularity groups. These findings imply that audio feature-driven clustering is effective in segmenting songs with unique popularity profiles, potentially highlighting different musical preferences or trends in listener behavior across these groups.

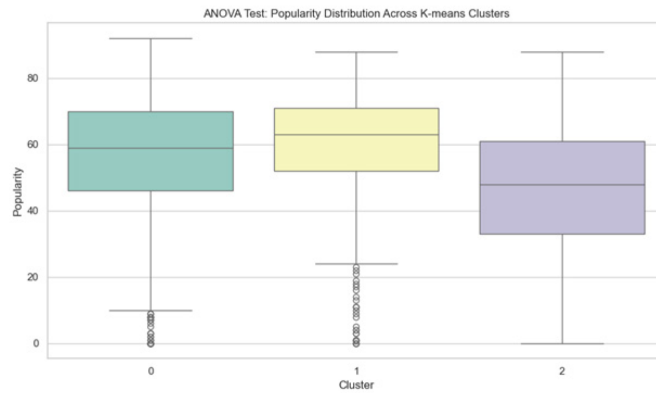

Figure 10a. ANOVA Test: Popularity Distribution Across K-Means Clusters.

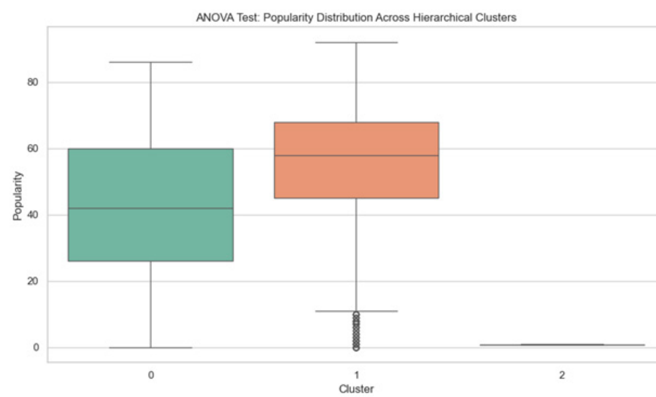

Figure 10b. ANOVA Test: Popularity Distribution Across Hierarchical Clusters.

Table 7. Tukey's HSD Test Results for Pairwise Popularity Differences Across K-means and Hierarchical Clusters

Clustering Method	Group1	Group2	Mean Difference	P-Value	Lower CI	Upper CI	Reject
K-means	0	1	2.4798	0.0028	0.7174	4.2421	TRUE
K-means	0	2	-10.6656	0	-11.8232	-9.508	TRUE
K-means	1	2	-13.1453	0	-15.039	-11.2517	TRUE
Hierarchical	0	1	14.0084	0	12.2737	15.7432	TRUE
Hierarchical	0	2	-40.5507	0.0003	-65.0829	-16.0185	TRUE
Hierarchical	1	2	-54.5591	0	-79.0421	-30.0762	TRUE

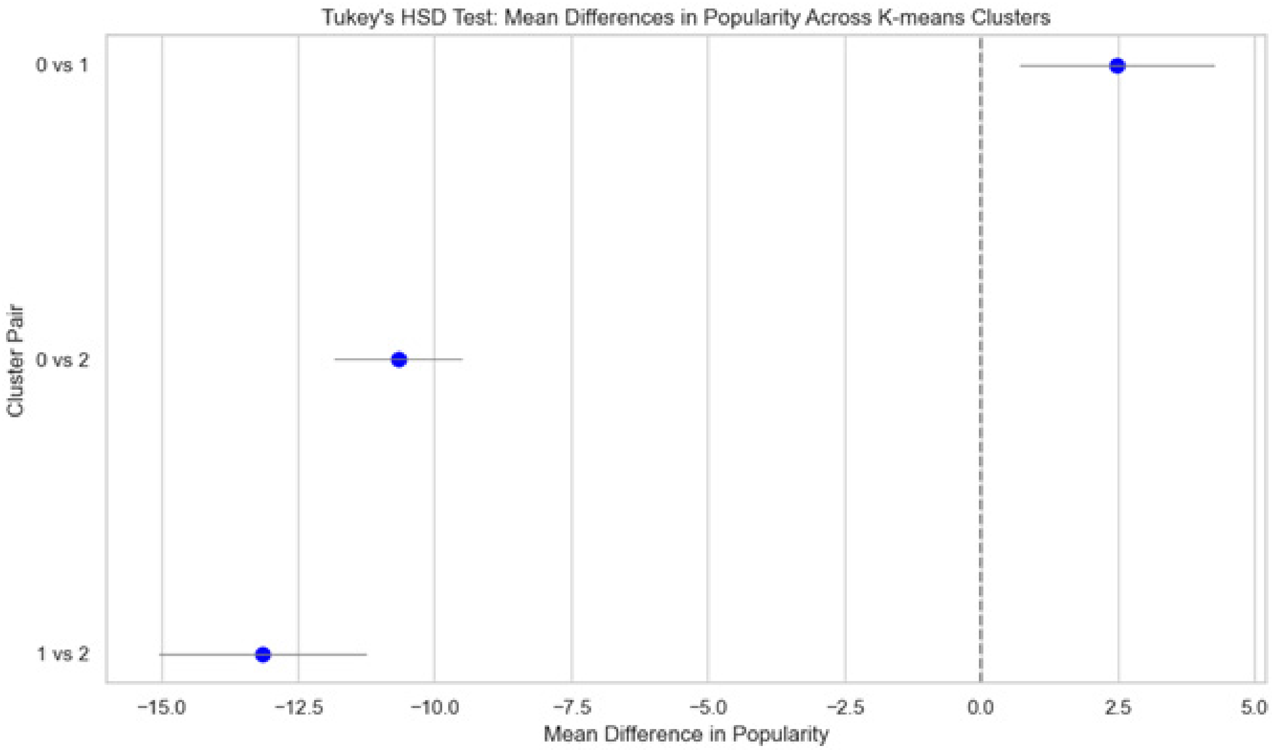

Figure 11a. Tukey's HSD Test: Mean Differences in Popularity Across K-means Clusters.

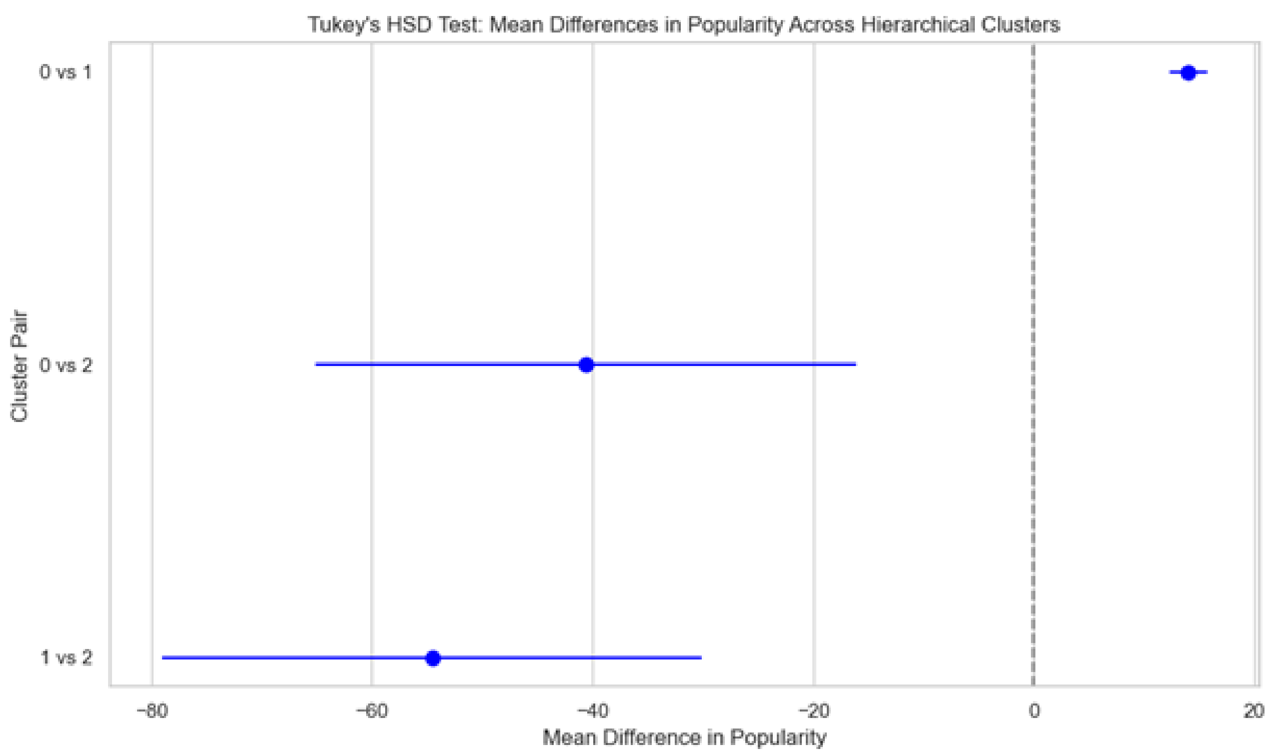

Figure 11b. Tukey's HSD Test: Mean Differences in Popularity Across Hierarchical Clusters.

CONCLUSION

This study provides insights into the relationship between audio features and song popularity on Spotify, highlighting key factors that drive listener engagement. High-energy, danceable, and loud songs consistently achieve higher popularity, while quieter and more experimental tracks tend to be less favored. Through a combination of linear and non-linear models, Loudness, Energy, and Danceability emerge as the most influential features. These findings are further validated by ensemble and clustering analyses, which provide a deeper understanding of how these attributes contribute to a song's success.

Results from PCA, K-means clustering, and Hierarchical Clustering offer a comprehensive view of the influence of audio features. Loudness, Danceability, and Energy, identified as the primary components in PCA, account for the most substantial variance in the dataset. Clustering reinforces these findings by grouping songs with similar feature profiles, revealing that clusters with high Loudness and Energy consistently correspond to greater popularity. However, outliers, such as Hierarchical Cluster 2, suggest that extreme values in Loudness may not always guarantee higher popularity, pointing to a nuanced relationship where balance is key.

Statistical tests like ANOVA and Tukey's HSD confirm significant differences in popularity between clusters, validating the effectiveness of clustering in segmenting songs based on their audio features. These results demonstrate how specific musical qualities resonate with listeners and influence popularity, offering valuable insights for artists and producers aiming to connect with audiences.

While these findings provide a strong foundation for understanding song popularity, the study is limited to audio features and does not account for lyrical or contextual elements that may also play a role. Future research could include these variables for a more holistic model. Nonetheless, this study underscores the importance of data-driven approaches in shaping music industry strategies and offers a basis for further exploration of musical characteristics across genres and platforms.

ACKNOWLEDGEMENTS

The author expresses heartfelt gratitude to the anonymous reviewers for their insightful comments and suggestions, which significantly enhanced this paper.

REFERENCES

1. Ivo. How TikTok is changing the music industry. *Stereofox*. 2021, February, Retrieved from https://www.stereofox.com/resources/how-tiktok-is-changing-the-music-industry/ (accessed on 2024-7-18).
2. Grome, J. Utilizing social media in the music industry. *Forbes Business Council*. 2023, September 22, Retrieved from https://www.forbes.com/sites/forbesbusinesscouncil/2023/09/22/utilizing-social-media-in-the-music-industry/ (accessed on 2024-7-18).
3. Statista Research Department. Global revenue of the music industry. *Statista*. 2024, May 29. Retrieved from https://www.statista.com/statistics/272305/global-revenue-of-the-music-industry/ (accessed on 2024-7-18).
4. Spotify for Developers. Get audio features. *Spotify Web API documentation*. Retrieved from https://developer.spotify.com/documentation/web-api/reference/get-audio-features (accessed on 2024-7-20).
5. Pearson K. Note on regression and inheritance in the case of two parents. *Proc R Soc Lond*. 1895; 58: 240–242. https://doi.org/10.1098/rspl.1895.0041.
6. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904; 15 (1): 72–101. https://doi.org/10.2307/1412159.
7. Tibshirani R. Regression shrinkage and selection via the lasso. J *R Stat Soc Series B Stat Methodol*. 1996; 58 (1): 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
8. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005; 67 (2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.
9. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 6th ed. Wiley; 2021. ISBN: 978-1119578727.
10. Breiman L. Random forests. *Mach Learn*. 2001; 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.
11. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001; 29 (5): 1189–1232. https://doi.org/10.1214/aos/1013203451.
12. Wilkinson L, Friendly M. The history of the cluster heat map. *Am Stat*. 2009; 63 (2): 179–184. https://doi.org/10.1198/tas.2009.0033.
13. Jolliffe IT. Principal component analysis. In: Lovric M, editor. International encyclopedia of statistical science. *Springer*; 2011; p. 1094–1096. https://doi.org/10.1007/978-3-642-04898-2_455.
14. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. *University of California Press*; 1967; p. 281–297.

15. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 1963; 58 (301): 236–244. https://doi.org/10.1080/01621459.1963.10500845.
16. Kaufman L, Rousseeuw PJ. Finding groups in data: An introduction to cluster analysis. *Wiley*; 1990. ISBN: 978-0471878766. https://doi.org/10.1002/9780470316801.
17. Fisher RA. Statistical methods for research workers. *Oliver and Boyd*; 1925.
18. Tukey JW. Comparing individual means in the analysis of variance. *Biometrics*. 1949; 5 (2): 99–114. https://doi.org/10.2307/3001913.
19. Thorndike RL. Who belongs in the family? *Psychometrika*. 1953; 18 (4): 267–276. https://doi.org/10.1007/BF02289263.
20. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987; 20: 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.