SOCIETA
Journal of Society and Change
Volume 1, Issue 1, January–June 2026

Mapping digital song popularity characteristics as a foundation for popular song identification in society

Viona Anjani Greit Lumban Toruan[1]*, Darwis Robinson Manalu[2]

[1] [2] Universitas Methodist Indonesia, Medan, Indonesia

*Corresponding Author: vionaanjani406@gmail.com

Submission: 16 June 2026 Revision: 1 July 2026 Accepted: 11 July 2026

Abstract

Song popularity on digital platforms has increasingly influenced public music preferences. However, differences in popularity metrics across Spotify, YouTube, and TikTok have resulted in fragmented approaches to identifying popular songs, preventing a comprehensive representation of song popularity. This study aims to identify song popularity characteristics by integrating cross-platform popularity metrics using the K-Means clustering algorithm. A quantitative approach was employed using primary data collected from questionnaires administered to 71 respondents and secondary data consisting of 500 songs selected from the *Most Streamed Spotify Songs 2024* dataset. The data were analyzed using *Z-transformation* normalization and the K-Means clustering algorithm. The findings identified three principal popularity characteristics—*Low Popular*, *Viral Popular*, and *Stable Popular*—demonstrating that TikTok virality does not necessarily reflect overall song popularity across digital platforms. This study contributes by proposing a conceptual framework for *Popular Song Identification* based on the integration of cross-platform popularity metrics, offering a more comprehensive approach to identifying popular songs for society, the digital music industry, and the development of music recommendation systems.

Keywords: K-Means clustering; Song popularity; Spotify; YouTube; TikTok.

Abstrak

Popularitas lagu pada platform digital semakin memengaruhi preferensi musik masyarakat. Namun, perbedaan metrik popularitas pada Spotify, YouTube, dan TikTok menyebabkan identifikasi lagu populer masih dilakukan secara parsial sehingga belum mampu merepresentasikan popularitas lagu secara menyeluruh. Penelitian ini bertujuan mengidentifikasi karakteristik popularitas lagu melalui integrasi metrik lintas platform menggunakan algoritma *K-Means Clustering*. Penelitian menggunakan pendekatan kuantitatif dengan data primer berupa kuesioner yang melibatkan 71 responden dan data sekunder berupa 500 lagu yang dipilih dari dataset *Most Streamed Spotify Songs 2024*. Data dianalisis menggunakan normalisasi *Z-Transformation* dan algoritma *K-Means Clustering*. Hasil penelitian mengidentifikasi tiga karakteristik utama popularitas, yaitu *Low Popular*, *Viral Popular*, dan *Stable Popular*, serta membuktikan bahwa viralitas TikTok tidak selalu merepresentasikan popularitas lagu secara menyeluruh. Penelitian ini berkontribusi melalui pengembangan kerangka konseptual *Popular Song Identification* berbasis integrasi metrik popularitas lintas platform sehingga menawarkan pendekatan yang lebih komprehensif dalam mengidentifikasi lagu populer bagi masyarakat, industri musik digital, dan pengembangan sistem rekomendasi musik.

Kata Kunci: *K-Means Clustering*; popularitas lagu; Spotify; YouTube; TikTok

Introduction

Song popularity on digital platforms has traditionally been identified using a single dominant indicator, such as the number of Spotify streams, YouTube views, or the level of virality on TikTok (Wu, 2026). Although this approach appears straightforward, it fails to capture song popularity comprehensively because each platform measures different forms of user interaction. Spotify reflects the intensity of music consumption, YouTube indicates audience engagement with audiovisual content, whereas TikTok represents the dissemination of songs through user-generated content (Bello & Garcia, 2021; Cao, 2025; Munaro et al., 2021). These differences in measurement mechanisms imply that a song achieving high popularity on one platform does not necessarily attain the same level of popularity on others. Consequently, identifying genuinely popular songs has become increasingly complex, as no single indicator can adequately represent the overall characteristics of song popularity across the digital music ecosystem.

This complexity is not merely a technical issue in data processing but also has significant implications for stakeholders who rely on song popularity information. The music industry uses popularity indicators to formulate promotional strategies, streaming platforms incorporate them into recommendation systems (Aguiar & Waldfogel, 2021), while content creators and listeners depend on them to discover emerging songs (Hracs & Webster, 2021). When popularity identification relies solely on a single metric, the resulting decisions may overlook user interaction patterns on other platforms. For instance, a song that goes viral through short-form videos may be used millions of times within a short period without necessarily generating high streaming activity (Arrieta, 2025). Conversely, a song with a substantial number of streams may sustain long-term popularity despite experiencing limited virality on social media (Denisova, 2023). These circumstances suggest that song popularity is a multidimensional phenomenon requiring a more comprehensive analytical approach.

One promising approach to addressing these differences is clustering analysis based on the integration of popularity metrics (Kanavos et al., 2023). Unlike conventional ranking approaches, which compare the value of each indicator independently, clustering groups objects according to similarities across multiple variables simultaneously (Li et al., 2018). Through this approach, songs are no longer classified according to the dominance of a single platform but rather according to interaction patterns formed by the combination of multiple popularity metrics. This approach is particularly relevant because it minimizes bias arising from differences in data scales across platforms while producing groups of songs with similar characteristics. Therefore, clustering serves not only as a data grouping technique but also as an analytical approach for understanding the structure of song popularity more comprehensively than relying on individual indicators in isolation.

Studies on popular song identification have been conducted using a variety of approaches and data sources. A study by Maasø and Hagen (2020) demonstrated that the number of streams can be used to predict song popularity, although it reflects user behavior only on streaming platforms. Rahardjo et al. (2024) found that social media virality contributes to increasing song exposure; however, it does not adequately explain the consistency of popularity across different digital platforms. Kowald et al. (2021) employed clustering techniques to group songs based on audio characteristics, metadata, and user preferences, yet they did not integrate cross-platform popularity metrics as the basis for identifying popular songs. Martin-Gutierrez et al. (2020) also applied machine learning algorithms to predict song popularity, but their work emphasized predictive performance rather than the underlying popularity characteristics. Consequently, previous studies have generally treated popularity as a set of isolated indicators, providing only a limited understanding of how interactions across digital platforms collectively shape song popularity.

Based on these research gaps, this study aims to identify popular songs by integrating popularity metrics from Spotify, YouTube, and TikTok using the K-Means clustering algorithm. Unlike previous studies, this research not only classifies songs according to their popularity characteristics but also synthesizes the clustering results into a conceptual framework of *Popular Song Identification*, which explains the relationships among integrated popularity metrics, the clustering process, and popular song identification. The primary contribution of this study lies in developing an identification approach that conceptualizes song popularity as a multidimensional phenomenon rather than as the representation of a single platform or a single indicator. By adopting this perspective, the study is expected to provide a more comprehensive foundation for the development of music recommendation systems, digital music industry analytics, and future research on popular song identification within the digital platform ecosystem.

Method

This study employed a quantitative approach using the K-Means clustering method to identify song popularity characteristics across digital platforms (Steinley, 2006). The research utilized both primary and secondary data sources. Primary data were collected through an online questionnaire administered to 71 respondents to examine users' perceptions of the process of identifying popular songs across different digital platforms. Secondary data were obtained from the *Most Streamed Spotify Songs 2024* dataset available on Kaggle, which contains 4,600 songs. From this dataset, 500 songs were selected based on their *All Time Rank* to serve as the research sample, as they represent songs with the highest levels of exposure on Spotify, YouTube, and TikTok. The combination of these two data sources provided an empirical basis for examining both song popularity characteristics and user behavior in identifying popular songs.

The study employed six popularity metrics representing three major digital platforms: *Spotify Streams*, *Spotify Popularity*, *YouTube Views*, *YouTube Likes*, *TikTok Views*, and *TikTok Likes*. The primary data were first subjected to validity and reliability testing to ensure that the research instrument consistently measured respondents' perceptions. The secondary data then underwent a preprocessing stage, including attribute selection, data cleaning, and normalization using the *Z-transformation* method (Graf, 2004). The normalization process was performed to eliminate differences in measurement scales among variables, ensuring that each popularity metric contributed equally to the clustering process and preventing variables with larger numerical ranges from dominating the analysis.

Data analysis was conducted using the K-Means clustering algorithm implemented in RapidMiner with the number of clusters (k) set to three. The number of clusters was determined based on the objective of grouping songs according to popularity characteristics derived from the integration of cross-platform popularity metrics. The clustering results were subsequently interpreted using centroid values, cluster membership distribution, and the characteristics of each cluster to identify patterns of song popularity across digital platforms. Beyond generating data classifications, the clustering results were synthesized into a conceptual framework for *Popular Song Identification*. Consequently, the analysis not only produced computational groupings but also proposed a conceptual model that explains the relationships among integrated popularity metrics, the clustering process, and the identification of popular songs in the digital platform era.

Results and Discussion

Popularity metrics as a challenge in identifying popular songs

The differences in popularity indicators used by each streaming platform mean that song popularity can no longer be measured using a single standardized parameter (Lee & Lee, 2018). Spotify relies on the number of streams and its popularity score, YouTube emphasizes views and likes, whereas TikTok evaluates song popularity based on the number of views and user interactions with content featuring the song (Colley et al., 2022). Consequently, users must compare multiple metrics simultaneously before determining whether a song is genuinely popular. To address this challenge, this study utilized primary data collected through a user survey and secondary data obtained from a cross-platform song dataset. The characteristics of the research data are presented in Table 1, which summarizes the data sources, sample size, and analytical procedures underlying the clustering process.

Table 1. Summary of research data

Component	Description
Primary Data	Google Forms questionnaire

Number of Respondents	71 respondents
Secondary Data	*Most Streamed Spotify Songs 2024*
Dataset Source	Kaggle
Total Dataset	4,600 songs
Research Sample	500 songs
Sampling Technique	Top 500 based on *All Time Rank*
Algorithm	K-Means clustering
Software	RapidMiner
Number of Clusters	3
Normalization Method	*Z-transformation*

Source: Authors' calculations (2026).

As shown in Table 1, the study involved 71 respondents as the primary data source to capture users' perceptions of the process of identifying popular songs. This sample size was considered sufficient to provide an initial understanding of user behavior when utilizing streaming platforms. In addition, the study analyzed 500 songs selected from a total of 4,600 songs in the dataset, representing approximately 10.9% of the entire dataset. The sample was intentionally limited to the highest-ranked songs based on *All Time Rank*, allowing the analysis to focus on songs with the greatest exposure across Spotify, YouTube, and TikTok. Furthermore, the inclusion of six variables derived from these three platforms demonstrates that the study did not rely on a single popularity indicator but instead integrated multiple metrics with distinct characteristics through a normalization process before clustering.

The characteristics of the research data were further examined through descriptive statistics for each variable included in the clustering analysis. These statistics, presented in Table 2, summarize the minimum, maximum, mean, and standard deviation of each popularity metric. Descriptive statistics are essential because each platform generates data on different numerical scales, making direct comparisons inappropriate. The greater the variation in value ranges among variables, the higher the likelihood that one attribute will dominate the others if the data are not normalized beforehand. Therefore, descriptive statistics provide an important foundation for understanding the initial characteristics of the dataset prior to applying the K-Means clustering algorithm.

Table 2. Research variables and descriptive statistics

Platform	Variable	Minimum	Maximum	Mean	Standard Deviation
Spotify	Spotify Streams	2,354,781	3,912,467,523	624,385,112	712,641,355

Spotify	Spotify Popularity	18	100	73.42	16.81
YouTube	YouTube Views	5,621,870	5,486,317,925	851,247,930	963,214,582
YouTube	YouTube Likes	52,314	41,285,463	6,428,115	7,394,511
TikTok	TikTok Views	1,284,512	18,562,437,115	2,634,428,721	3,247,186,544
TikTok	TikTok Likes	13,512	2,418,563,874	312,854,731	421,513,614

Source: Authors' calculations (2026).

The results presented in Table 2 indicate that *TikTok Views* reached a maximum value of 18,562,437,115, substantially exceeding both *YouTube Views* (5,486,317,925) and the average *Spotify Streams* (624,385,112). This substantial difference suggests that song dissemination on TikTok has a much broader reach than on other platforms because a single song can be reused across millions of user-generated videos. Meanwhile, *Spotify Popularity* ranged from 18 to 100, with an average score of 73.42, indicating that most songs included in the sample were relatively popular. On the other hand, the exceptionally high standard deviations for *TikTok Views* (3,247,186,544) and *TikTok Likes* (421,513,614) reveal considerable variation in the data distribution. This pattern suggests that only a small number of songs achieve extremely high levels of virality, whereas the majority receive substantially lower user engagement. Such an uneven distribution reinforces the argument that clustering provides a more appropriate analytical approach than evaluating each popularity indicator separately.

Data quality is determined not only by the number of observations but also by the ability of the research instrument to measure the phenomenon consistently. Therefore, before analyzing the survey responses, the questionnaire was evaluated using validity and reliability tests. These tests were conducted to ensure that each questionnaire item accurately represented respondents' perceptions of the challenges associated with identifying popular songs across digital platforms. The results of the instrument evaluation are presented in Table 3. Meeting both validity and reliability criteria confirms that respondents' perceptions can be used to support the interpretation of the clustering results, ensuring that the findings are grounded not only in computational analysis but also in users' real-world experiences.

Table 3. Instrument validity and reliability test results

Test	Item/Method	r-value / Value	Threshold	Status
Validity	P1	0.572	> 0.235	Valid
Validity	P2	0.723	> 0.235	Valid
Validity	P3	0.678	> 0.235	Valid
Validity	P4	0.638	> 0.235	Valid

Validity	P5	0.578	> 0.235	Valid
Validity	P6	0.553	> 0.235	Valid
Validity	P7	0.647	> 0.235	Valid
Validity	P8	0.332	> 0.235	Valid
Validity	P9	0.238	> 0.235	Valid
Validity	P10	0.471	> 0.235	Valid
Reliability	Cronbach's Alpha	0.746	> 0.700	Reliable

Source: Authors' calculations using SPSS (2026).

As presented in Table 3, all questionnaire items produced *r*-values exceeding the *r*-table threshold (0.235), indicating that every item satisfied the validity criterion. The highest correlation coefficient was observed for P2 (0.723), suggesting that the statement concerning the need to search for popular songs across multiple platforms had the strongest relationship with the underlying construct. In contrast, P9 yielded the lowest correlation coefficient (0.238). Although it was the smallest value among all items, it still exceeded the *r*-table value, thereby fulfilling the validity requirement. Furthermore, the Cronbach's Alpha coefficient of 0.746 indicates good internal consistency because it surpasses the minimum acceptable threshold of 0.70. These findings confirm that respondents' perceptions provide a reliable basis for explaining the challenges users encounter when identifying popular songs across digital platforms.

After the research instrument had been confirmed as both valid and reliable, the next stage involved analyzing respondents' perceptions of the popularity metrics used across different digital platforms. A summary of the survey results is presented in Table 4. The mean score for each item was used to describe respondents' overall tendencies regarding the process of identifying popular songs. Higher mean scores indicate stronger agreement with the corresponding statement. Consequently, the survey results not only illustrate how users perceive popularity indicators on individual platforms but also reveal the challenges they face when comparing popularity information derived from different digital sources.

Table 4. User survey results

Item	Statement	Mean	Interpretation
P1	It is difficult to identify currently popular songs	2.82	Neutral
P2	I have to search for songs on multiple platforms	2.97	Neutral
P3	I find it confusing to use trending song search features	2.93	Neutral
P4	I use more than one music platform	3.35	Agree
P5	I switch platforms because information is incomplete	3.34	Agree

P6	It is difficult to compare song popularity across platforms	3.20	Neutral–Agree
P7	Spotify popularity is measured by the number of streams	3.66	Agree
P8	YouTube popularity is measured by the number of views	3.73	Agree
P9	TikTok popularity is measured by the frequency of soundtrack usage	3.76	Agree
P10	Differences in popularity indicators are confusing	3.28	Neutral–Agree

Source: Authors' calculations from questionnaire data using SPSS (2026).

The results in Table 4 show that P9 achieved the highest mean score (3.76), indicating that most respondents agreed that song popularity on TikTok is primarily determined by how frequently a song is used as a soundtrack in user-generated content. Similarly, high mean scores for P8 (3.73) and P7 (3.66) demonstrate that respondents clearly recognized the differences in popularity indicators between YouTube and Spotify. Meanwhile, P4 (3.35) and P5 (3.34) suggest that users tend to rely on multiple music platforms because a single platform does not provide sufficiently comprehensive popularity information. Most importantly, the responses to P10 (3.28) and P6 (3.20) reveal that respondents continue to experience difficulties when comparing song popularity across platforms due to differences in the metrics employed. These findings confirm that the diversity of popularity indicators not only reflects the unique characteristics of each platform but also constitutes a practical challenge for users attempting to identify genuinely popular songs. This evidence further justifies the application of the K-Means clustering algorithm as an approach for integrating multiple popularity metrics into more interpretable popularity groups, as discussed in the following subsection.

Identifying song popularity characteristics on digital platforms

Song popularity in the digital era has long been perceived as a universal measure, implying that a song popular on one platform should exhibit a similar level of popularity across other platforms (Ta et al., 2026). However, the findings of this study reveal a different reality. Song popularity is shaped by heterogeneous patterns of user interaction, leading each platform to generate a distinct representation of popularity. Consequently, a single indicator is no longer sufficient to explain a song's popularity comprehensively. These findings reinforce the results presented in the previous subsection, demonstrating that differences in popularity metrics constitute the primary challenge in identifying popular songs. To examine this phenomenon, the K-Means clustering algorithm was applied to 500 songs, resulting in three clusters with distinct centroid characteristics, as presented in Table 5.

Table 5. Centroid values of the K-Means clustering results

Variable	Cluster 0	Cluster 1	Cluster 2
Spotify Streams	-0.345	-1.012	1.344

Spotify Popularity	-0.103	-0.590	0.405
YouTube Views	-0.343	-0.554	1.324
YouTube Likes	-0.403	-0.621	1.558
TikTok Views	-0.112	9.736	0.241
TikTok Likes	-0.126	8.970	0.310

Source: Authors' calculations using RapidMiner (2026).

The most significant finding presented in Table 5 demonstrates that TikTok virality is not synonymous with cross-platform popularity. This is evident in Cluster 1, which recorded exceptionally high centroid values for *TikTok Views* (9.736) and *TikTok Likes* (8.970), while all Spotify and YouTube indicators remained below the average, including *Spotify Streams* (-1.012), *Spotify Popularity* (-0.590), *YouTube Views* (-0.554), and *YouTube Likes* (-0.621). This pattern indicates that extensive use of a song in TikTok content does not automatically translate into increased streaming activity or music video consumption. In other words, social media virality primarily reflects a short-lived surge in public attention, whereas cross-platform popularity requires consistent user engagement across multiple digital ecosystems. These findings suggest that relying on a single metric to identify popular songs may produce misleading classifications because it overlooks the distinct characteristics of each platform.

Beyond revealing differences in popularity characteristics, the clustering results also demonstrate that digital song popularity is distributed unevenly and concentrated among only a small number of songs. This finding is particularly important because it highlights the highly competitive nature of the digital music ecosystem. Only a limited number of songs consistently capture public attention, while the majority receive relatively low levels of exposure. The distribution of cluster membership presented in Table 6 provides a clearer illustration of this imbalance and indicates the extent to which specific popularity characteristics dominate the overall dataset.

Table 6. Distribution of clustering results

Cluster	**Number of Songs**	**Percentage**	**Category**
Cluster 0	395	79.0%	Low Popular
Cluster 1	2	0.4%	Highly Popular
Cluster 2	103	20.6%	Moderately Popular
Total	500	100%	—

Source: Authors' calculations using RapidMiner (2026).

As shown in Table 6, 395 songs (79.0%) were classified into Cluster 0, whereas only two songs (0.4%) exhibited the characteristics of Cluster 1. In contrast, 103 songs (20.6%) belonged to Cluster 2, representing a more stable pattern of popularity. This distribution confirms that digital song popularity follows a highly unequal pattern, in which user attention is concentrated on only a small number of exceptionally successful songs, while the majority receive relatively limited exposure. The dominance of Cluster 0 further indicates that merely being available on digital platforms does not guarantee widespread popularity. Therefore, these findings demonstrate that the principal challenge in identifying popular songs lies not only in the differences among platform-specific metrics but also in the inherently uneven distribution of popularity itself. This observation provides a strong foundation for interpreting the characteristics of each cluster in the following discussion.

The clustering results reveal not only differences in popularity levels but also distinct interaction characteristics associated with each group of songs across digital platforms. These findings directly address the challenge identified in the previous subsection, where users experienced difficulty identifying popular songs because of inconsistencies among platform-specific popularity metrics. Through the clustering process, these diverse indicators were successfully simplified into three clearly defined groups, allowing the identification process to move beyond dependence on a single metric. The characteristics of each cluster are presented in Table 7, illustrating how the dominance of Spotify, YouTube, and TikTok shapes different patterns of song popularity. Consequently, the clustering results not only produce meaningful data groupings but also establish a more interpretable classification of digital song popularity.

Table 7. Characteristics of each cluster

Aspect	**Cluster 0**	**Cluster 1**	**Cluster 2**
Number of Songs	395	2	103
Spotify Streams	Low	Very Low	High
Spotify Popularity	Low	Low	High
YouTube Views	Low	Low	High
YouTube Likes	Low	Low	High
TikTok Views	Low	Very High	Moderate
TikTok Likes	Low	Very High	Moderate
General Characteristics	Songs with consistently low performance across all platforms	Viral songs dominated by TikTok engagement	Songs with strong performance on Spotify and YouTube and relatively high engagement on TikTok
Example Songs	*Fortnight, Saturn, Training Season*	*Beautiful Things, Die With A Smile*	*Espresso, BIRDS OF A FEATHER, Too Sweet*

Source: Authors' calculations (2026).

The findings presented in Table 7 indicate that Cluster 0 represents more than simply a group of unpopular songs; rather, it characterizes the majority of digital songs that have not yet succeeded in attracting widespread public attention. The presence of 395 songs within this cluster demonstrates that most songs perform poorly across Spotify, YouTube, and TikTok simultaneously. In contrast, Cluster 1 represents a unique phenomenon, consisting of only two songs that achieved exceptionally high levels of interaction on TikTok. This suggests that virality is a highly exclusive characteristic possessed by only a small number of songs. Meanwhile, Cluster 2 exhibits a more stable popularity profile, characterized by strong performance on Spotify and YouTube while maintaining relatively high engagement on TikTok. These findings demonstrate that the success of a song is determined more by consistent performance across multiple platforms than by dominance on a single platform alone.

These findings have important implications for the way popular songs are identified on digital platforms. Traditionally, popularity has been assessed using a single indicator, such as the number of Spotify streams or TikTok views. However, the results of this study demonstrate that such an approach fails to capture the multidimensional nature of song popularity because each platform reflects different patterns of user interaction. By applying the K-Means clustering algorithm, six popularity indicators were successfully integrated into three principal characteristics: low popularity, social media-driven viral popularity, and stable cross-platform popularity. Consequently, this study contributes not only a data clustering model but also a more comprehensive framework for understanding digital song popularity. These findings directly address the research objective by demonstrating that K-Means clustering effectively simplifies the complexity of multiple popularity metrics into a more informative classification system, thereby facilitating the identification of popular songs more effectively than relying on a single indicator in isolation.

The role of popularity clustering in popular song identification

Song popularity on digital platforms has traditionally been identified using a single dominant indicator, such as the number of Spotify streams, YouTube views, or the level of virality on TikTok (Oliveira et al., 2025). Although this approach appears straightforward, it provides only a partial representation of song popularity because each platform captures user behavior through different mechanisms. The findings presented in the previous subsection demonstrate that songs that go viral on TikTok do not necessarily perform well on Spotify or YouTube. This discrepancy indicates that relying on a single metric may lead to biased conclusions when identifying popular songs. Therefore, an approach capable of integrating multiple indicators into a more comprehensive identification framework is required. Based on the clustering results, this study proposes a conceptual framework for *Popular Song Identification*, as illustrated in Figure 1.

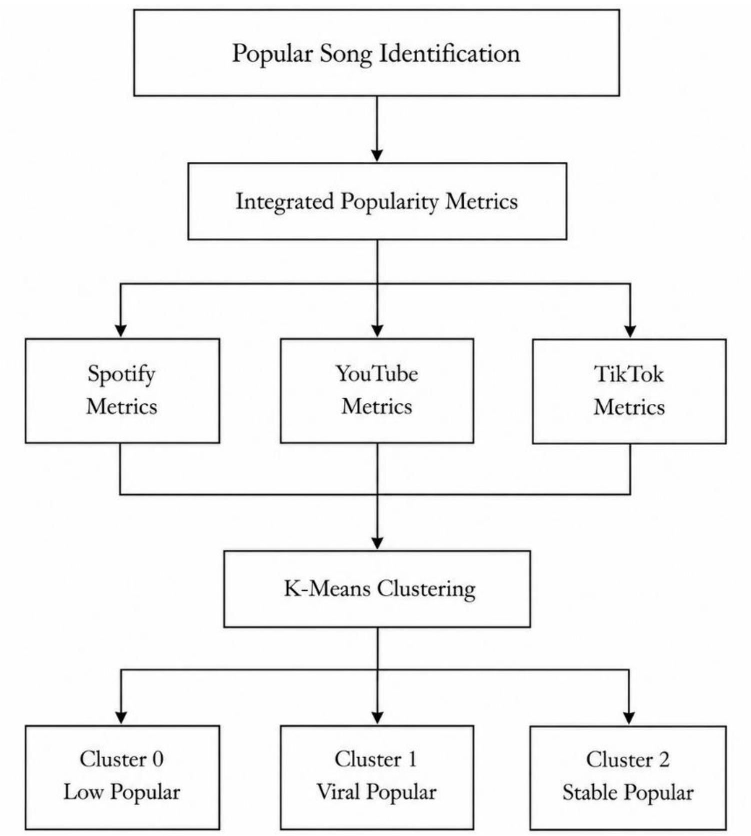

Figure 1. Conceptual framework of popular song identification based on integrated popularity clustering

Source: Developed by the authors (2026).

Figure 1 was developed as a synthesis of the study's overall findings rather than as an illustration of the clustering algorithm itself. Unlike a conventional flowchart that merely describes analytical procedures, this conceptual framework demonstrates the logical relationships among popularity metrics, data integration, clustering results, and popular song identification. The framework emphasizes that identifying popular songs should not rely on a single indicator but instead should incorporate multiple metrics representing different patterns of user interaction across Spotify, YouTube, and TikTok. Consequently, Figure 1 serves as a conceptual model that translates computational findings into an interpretative framework for understanding song popularity in a more comprehensive manner.

The first component of Figure 1 is *Integrated Popularity Metrics*, which combines Spotify, YouTube, and TikTok metrics as representations of three distinct forms of digital interaction. Spotify reflects music consumption intensity through the number of streams and popularity scores, YouTube captures audience engagement through views and likes, whereas TikTok represents song dissemination through the use of audio in user-generated content. The integration of these three metrics demonstrates that song popularity is a multidimensional phenomenon that cannot be reduced to a single indicator. This finding is consistent with the study by Trunfio and Rossi (2021), which argued that evaluating digital popularity requires the integration of multiple user interaction indicators. Likewise, Han and Anderson (2026) showed that relying on a single metric tends to introduce platform-specific bias. Therefore, the present study reinforces the argument that integrating multiple popularity metrics is a prerequisite for achieving a more representative identification of song popularity.

The second component is *K-Means Clustering*, which integrates multiple popularity metrics into groups with similar characteristics. The findings indicate that the clustering process successfully identified three primary popularity characteristics: *Low Popular*, *Viral Popular*, and *Stable Popular*. These characteristics demonstrate that song popularity evolves through different interaction patterns depending on the dominant platform. This finding is consistent with the study by Chaudhry et al. (2023), which demonstrated that clustering techniques are effective for uncovering hidden patterns within high-dimensional datasets. Similarly, Ruggeri et al. (2020) found that grouping data based on similarity produces more informative interpretations than ranking objects using a single indicator. However, the present study differs from Gu et al. (2018) by not only generating data clusters but also interpreting each cluster as a representation of cross-platform song popularity characteristics.

The final component of Figure 1 is *Popular Song Identification*, which represents the primary outcome of the proposed conceptual framework. Popular song identification is no longer determined by the dominance of a single platform but by cluster characteristics derived from the integration of multiple popularity metrics. This approach enables songs to be classified according to more comprehensive interaction patterns, distinguishing between songs that are merely viral on social media, songs that maintain stable popularity across multiple platforms, and songs that have yet to achieve substantial public exposure. This finding supports the study by Jatain et al. (2022), which emphasized the importance of multidimensional approaches in digital popularity analysis. However, the present study extends previous research by introducing a conceptual framework that integrates popularity metrics, the clustering process, and popular song identification into a single comprehensive model.

Overall, Figure 1 demonstrates that the primary contribution of this study lies not merely in applying the K-Means algorithm but in reconstructing the process of identifying popular songs through the integration of multiple popularity metrics. The findings confirm that song popularity is a multidimensional phenomenon that cannot be adequately represented by a single indicator or a single platform. The proposed conceptual framework successfully simplifies the complexity of six popularity metrics into three principal characteristics that are easier to interpret during the popular song identification process. Consequently, this study not only develops a data-driven clustering model but also proposes a new perspective in which popular song identification should adopt a cross-platform approach that considers the full spectrum of digital user interactions rather than evaluating virality or streaming counts independently.

Conclusion

This study demonstrates that song popularity across digital platforms is a multidimensional phenomenon that cannot be accurately identified using a single popularity indicator. Clustering analysis based on six popularity metrics derived from Spotify, YouTube, and TikTok successfully revealed three principal popularity characteristics: *Low Popular*, *Viral Popular*, and *Stable Popular*. These findings demonstrate that social media virality does not necessarily represent overall song popularity because each platform reflects different forms of user interaction. Consequently, the clustering approach effectively simplifies the complexity of multiple popularity metrics into a more representative classification, allowing popular song identification to be based on integrated interaction patterns rather than on a single platform or indicator.

This study contributes conceptually by developing a *Popular Song Identification* framework based on the integration of cross-platform popularity metrics while demonstrating that the K-Means algorithm can be effectively employed to identify song popularity characteristics in a more comprehensive manner. Nevertheless, the study has several limitations, including the use of a single dataset, six popularity indicators, and one clustering algorithm, meaning that the resulting characteristics may not fully capture the dynamics of the broader digital music ecosystem. Future research is therefore encouraged to expand the dataset by incorporating additional platforms such as Apple Music and Shazam, include supplementary indicators such as *growth rate*, *engagement rate*, and user sentiment, and compare the performance of K-Means with other clustering algorithms, including Hierarchical Clustering, DBSCAN, and Gaussian Mixture Models, to develop a more adaptive and accurate framework for popular song identification.

References

Aguiar, L., & Waldfogel, J. (2021). Platforms, power, and promotion: Evidence from Spotify playlists. *The Journal of Industrial Economics, 69*(3), 653–691. https://doi.org/10.1111/joie.12263

Arrieta, A. (2025). The limits of virality: Music creators and platform negotiation in the era of short-form video. *Social Media + Society, 11*(4). https://doi.org/10.1177/20563051251388000

Bello, P., & Garcia, D. (2021). Cultural divergence in popular music: The increasing diversity of music consumption on Spotify across countries. *Humanities and Social Sciences Communications, 8*(1), 182. https://doi.org/10.1057/s41599-021-00855-1

Cao, H. (2025). Exploring the promotion of musical intangible cultural heritage under TikTok short videos. *Scientific Reports, 15*(1), 21772. https://doi.org/10.1038/s41598-025-09723-3

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. *Symmetry, 15*(9), 1679. https://doi.org/10.3390/sym15091679

Colley, L., Dybka, A., Gauthier, A., Laboissonniere, J., Mougeot, A., Mowla, N., Dick, K., Khalil, H., & Wainer, G. (2022). Elucidation of the relationship between a song's Spotify descriptive metrics and its popularity on various platforms. *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 241–249. https://doi.org/10.1109/COMPSAC54236.2022.00042

Denisova, A. (2023). Viral journalism. Strategy, tactics and limitations of the fast spread of content on social media: Case study of the United Kingdom quality publications. *Journalism, 24*(9), 1919–1937. https://doi.org/10.1177/14648849221077749

Graf, U. (2004). z-Transformation. In *Applied Laplace transforms and z-transforms for scientists and engineers* (pp. 77–113). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-7846-3_2

Gu, H., Wang, J., Wang, Z., Zhuang, B., Bian, W., & Su, F. (2018). Cross-platform modeling of users' behavior on social media. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 183–190. https://doi.org/10.1109/ICDMW.2018.00035

Han, S., & Anderson, C. K. (2026). The platform matters: Selection and measurement bias in online reviews. *Cornell Hospitality Quarterly, 67*(1), 42–53. https://doi.org/10.1177/19389655251327536

Hracs, B. J., & Webster, J. (2021). From selling songs to engineering experiences: Exploring the competitive strategies of music streaming platforms. *Journal of Cultural Economy, 14*(2), 240–257. https://doi.org/10.1080/17530350.2020.1819374

Jatain, D., Singh, V., & Dahiya, N. (2022). A multi-perspective micro-analysis of popularity trend dynamics for user-generated content. *Social Network Analysis and Mining, 12*(1), 147. https://doi.org/10.1007/s13278-022-00969-7

Kanavos, A., Karamitsos, I., & Mohasseb, A. (2023). Exploring clustering techniques for analyzing user engagement patterns in Twitter data. *Computers, 12*(6), 124. https://doi.org/10.3390/computers12060124

Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M., & Lex, E. (2021). Support the underground: Characteristics of beyond-mainstream music listeners. *EPJ Data Science, 10*(1), 14. https://doi.org/10.1140/epjds/s13688-021-00268-9

Lee, J., & Lee, J.-S. (2018). Music popularity: Metrics, characteristics, and audio-based prediction. *IEEE Transactions on Multimedia, 20*(11), 3173–3182. https://doi.org/10.1109/TMM.2018.2820903

Li, Z., Nie, F., Chang, X., Yang, Y., Zhang, C., & Sebe, N. (2018). Dynamic affinity graph construction for spectral clustering using multiple features. *IEEE Transactions on Neural Networks and Learning Systems, 29*(12), 6323–6332. https://doi.org/10.1109/TNNLS.2018.2829867

Maasø, A., & Hagen, A. N. (2020). Metrics and decision-making in music streaming. *Popular Communication, 18*(1), 18–31. https://doi.org/10.1080/15405702.2019.1701675

Martin-Gutierrez, D., Hernandez Penaloza, G., Belmonte-Hernandez, A., & Alvarez Garcia, F. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access, 8*, 39361–39374. https://doi.org/10.1109/ACCESS.2020.2976033

Munaro, A. C., Hübner Barcelos, R., Francisco Maffezzolli, E. C., Santos Rodrigues, J. P., & Cabrera Paraiso, E. (2021). To engage or not engage? The features of video content on YouTube affecting digital consumer engagement. *Journal of Consumer Behaviour, 20*(5), 1336–1352. https://doi.org/10.1002/cb.1939

Oliveira, G. P., Da Silva, A. P. C., & Moro, M. M. (2025). On the causal relationship between music virality and success. *IEEE Access, 13*, 122782–122791. https://doi.org/10.1109/ACCESS.2025.3589173

Rahardjo, E. Z., Alifa, J. M., Setiawan, S. Z., Gunawan, A. A. S., & Setiawan, K. E. (2024). Viral melodies: Exploring the factors influencing music virality in TikTok engagement. *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 159–164. https://doi.org/10.1109/ICSINTESA62455.2024.10748012

Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. (2020). Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes, 18*(1), 192. https://doi.org/10.1186/s12955-020-01423-y

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology, 59*(1), 1–34. https://doi.org/10.1348/000711005X48266

Ta, N., Jiao, F., Lin, C., & Shen, C. (2026). A computational analysis of the platformization of music: Comparing hit songs on TikTok and Spotify. *Information, Communication & Society, 29*(3), 1041–1059. https://doi.org/10.1080/1369118X.2025.2539297

Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing, 2021*(3), 267–292. https://doi.org/10.1007/s43039-021-00035-8

Wu, Y. (2026). Utilizing artificial intelligence to predict and interpret trends in digital music consumption culture. *Humanities and Social Sciences Communications*. https://doi.org/10.1057/s41599-026-07452-0