Analyzing Seasonal and Life Cycle Patterns in Music Popularity: A Clustering and Predictive Modeling Approach to Song Trajectories

s2665822 **s2693044** **s2716530** **s2751141**

Abstract

Understanding the life cycles of song popularity is crucial for the music industry, influencing decisions on production, release timing, and marketing. This study uses Spotify's "Top 200 Songs" dataset (2017–2023) to explore how features like energy, danceability and seasonal trends drive a song's trajectory. KMeans clustering identified three distinct groups: Cluster 0, mainstream tracks with high energy and danceability; Cluster 1, acoustic tracks with longer peak durations; and Cluster 2, high-energy tracks that peak quickly and sustain popularity. Seasonal trends reveal that energetic songs thrive in summer, while reflective, acoustic tracks are more successful in winter and fall. These findings highlight the importance of aligning song releases with seasonal preferences to optimize audience engagement. A range of regression models are explored for prediction of the time from a song's release to its peak, providing the foundation for potential success in future approaches with supplementary data or reclassification into binary targets.

1 Introduction

The rise and fall of songs in the music industry reflect the ever-shifting dynamics of audience behaviour. This report analyses patterns in song trajectories after release, through examination of key measures like time to peak popularity, duration at peak, and rate of decline. We also explore seasonal shifts in song popularity and intrinsic audio features of songs as drivers for higher rankings. This approach aligns with existing research on digital engagement patterns, which have shown that certain attributes, including energy level and danceability, can significantly influence the duration and intensity of a song's popularity [21, 33], while seasonality can influence listener behavior [20, 27]. This report applies cluster analysis and predictive modelling techniques to group songs with similar features and then predict the trajectory of a song's popularity. The goal is to build a framework for categorizing songs based on life cycle and seasonal profiles, developing predictive insights to guide strategic decisions around release timing, promotional strategies, and content targeting. The results of this research could benefit music producers, marketers, and streaming platforms by identifying patterns that reflect both inherent song qualities and the cyclical nature of audience preferences. In essence, this project aims to bridge the gap between music content characteristics and the broader seasonal and life cycle dynamics of music consumption, contributing to understanding what drives a song's trajectory to peak popularity, sustainment, and eventual decline over time.

2 Data Preparation

We used the "Top 200 Spotify Songs" dataset provided to us, which includes daily top 200 songs from 01/01/2017 to 29/05/2023. Each entry represents a song's ranking for that day, along with features such as energy and artist nationality. We refer to these as "hits" since a song may appear multiple times if it stays in the top rankings. We also gathered release dates using the Spotify for Developers

API [31], though we couldn't get a response from Spotify about data usage guidelines. Since we're not using the data commercially, we moved forward. To match release dates, we used song IDs, as matching titles led to inconsistencies. We saved dates in batches of 20 to handle API limits, and for missing dates, we defaulted to January 1st or manually filled in the gaps. This approach follows best practices for missing data handling [16]. There were no other missing values. We standardized column names and removed redundant columns, then converted the date column to the correct format. Since features like danceability and energy had different scales, we normalized them to a 0-1 range, so that each feature is equally weighted in our models. For feature engineering, we focused on *time_to_peak*, *duration_at_peak*, and *rate_of_decline* to track a song's popularity journey. We also added *release_month* and *day_of_week* to capture seasonal trends [2]. Regional factors like nationality and continent were included to explore cultural influences on song popularity. To handle outliers, we used the interquartile range method for features like danceability and energy, and quantile clipping for more skewed features like speechiness [1]. We then performed KMeans clustering, staging this before data splitting so that we could still test against meaningful validation labels during prediction, an approach taken in [8] and [5]. Finally, we split the data into training, validation, and test sets, with 80% for training and 20% for testing, further splitting the test set for validation. This approach ensures our model will generalize well to new data [10]. We also ensured songs didn't appear in both training and testing sets by working at the song level before splitting and then merging back into the full data. Each step in data preparation was aimed at creating a clean, well-organized dataset to reveal patterns in song popularity, seasonality, and regional trends.

3 Exploratory Data Analysis (EDA)

In this section of the report, we conducted an in-depth Exploratory Data Analysis (EDA) to examine how song characteristics and life cycle metrics evolve over different time frames and seasons, and whether they influence a song's popularity. We began by processing the *train_data* set, which included musical features as well as life cycle metrics - release date, peak rank, time to peak, duration at peak and rate of decline. To determine differences between more and less successful songs, we split the data into songs attaining a rank of at least 20 (1023 songs) and songs which did not meet this benchmark (4839 songs).

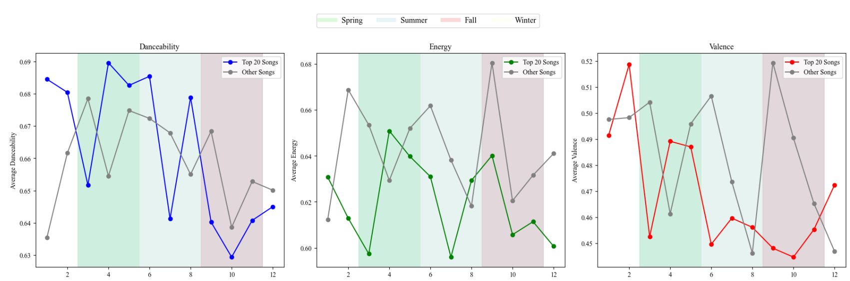

Figure 1: *Average Song Features by Release Month (Top 20 vs Other Songs), with seasonal boundaries*

We then examined the seasonal impact on song performance by identifying release months and grouping these into four seasons: Winter (December, January, February), Spring (March, April, May), Summer (June, July, August), and Fall (September, October, November). This helped us identify patterns in the success of songs based on which season they were released. For example, we observed that energetic songs performed better in the summer, while slower, more reflective songs thrived in the winter [25]. We also observed that Winter had the highest valence for Top 20 Songs, suggesting that songs with a more positive, upbeat tone resonated more with audiences during this season. Similarly, Winter had longer durations at peak (Mean of 3.35 days for Top 20 songs) and slower rates of decline (54.67 for Top 20 songs), possibly due to less competition in the season [22] as seen in Figure 2. We calculated monthly averages for each of the musical features, comparing Top 20 Songs and Other Songs. The Top 20 Songs had higher danceability in Spring, suggesting that rhythmic songs are often released during this season to build momentum and be ready for the energetic atmosphere of Summer. Summer saw higher energy levels for these songs, highlighting that energetic tracks thrive during festivals and outdoor events [17]. Additionally, we found that the time to peak for Top 20 Songs was shorter in Spring (Mean of 23.69 days), suggesting that songs released during this time tend to

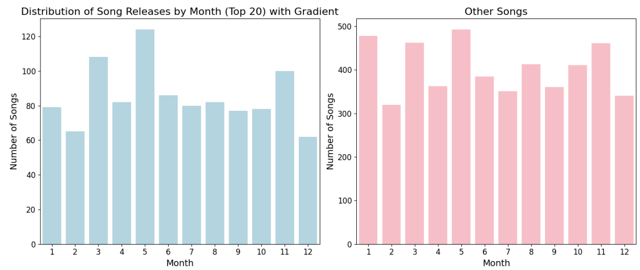

Figure 2: *Distribution of Songs Released by Month for Top 20 and Other Songs*

gain popularity more quickly, possibly due to increased media and social activity. In contrast, Winter releases are slow to grow in popularity and reach their peak (54.22 days) [26].

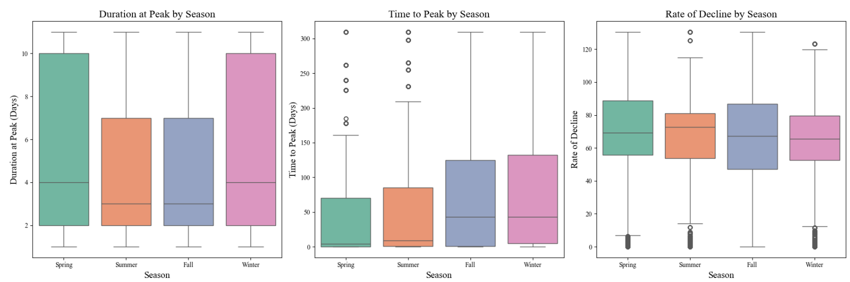

Figure 3: *Distribution of Life Metric by Season on the entire data set*

By using scatterplots and regression lines, we analyzed the relationship between song characteristics and life cycle metrics. We found a negative correlation between danceability and time to peak, suggesting that more danceable songs took longer to reach peak popularity, perhaps because they are more dependent on social events or activities to gain traction. The distribution of life cycle metrics for both Top 20 Songs and Other Songs revealed that Top 20 Songs exhibited more consistent durations at peak, whereas Other Songs displayed greater variability in their performance, highlighting the difference between high-ranking and lower-ranking songs. These insights from the EDA provide strong evidence for why cluster analysis would be a beneficial next step in the analysis process. The patterns observed in seasonal trends and life cycle metrics suggest that groups of songs with similar characteristics and trajectories could be identified through clustering [3]. For instance, songs released in Spring showed quicker rises to peak rank, while Winter releases stayed at the top for longer [7]. Clustering could help us distinguish between songs with short-term success and those with long-lasting popularity [7].

4 Cluster Analysis

The EDA has provided valuable insights into how song characteristics, seasonal factors, and life cycle metrics influence song success. These observations naturally lead to cluster analysis, which will allow us to group songs with similar life cycle patterns and feature profiles. We evaluated four clustering algorithms—KMeans, DBSCAN, Agglomerative Clustering, and Gaussian Mixture—using a range of clustering quality metrics. We focused on the Silhouette Score, Davies-Bouldin Score, Calinski-Harabasz Score, and Inertia to analyse the performance of each model. According to [28], a higher silhouette value (a measure of degree of separation between clusters) indicates better cluster separation, which is critical for interpretability. KMeans yielded the highest silhouette score (0.214), suggesting that its clusters were the most distinct compared to the others, while DBSCAN resulted in a negative score (-0.283), indicating poor clustering quality. The Davies-Bouldin Score (a measure of compactness and separation of clusters) also favoured KMeans with the lowest score (1.531), indicating better separation. This metric aligns with [11], who emphasize that lower Davies-Bouldin scores leads to more desirable clustering results. Calinski-Harabasz (a measure of the ratio of between-cluster dispersion to within-cluster dispersion) proved KMeans as the top performer with a score of 113.623, compared to much lower scores for DBSCAN (2.045) and Gaussian Mixture (53.118). Based on these metrics, KMeans demonstrated an optimal clustering performance and was selected as the model for further analysis. We analysed silhouette scores and the elbow method graph to determine the ideal number of clusters. Although the silhouette score for 2 clusters (0.2366) is slightly higher than that for 3 clusters (0.2247), there is only a marginal difference between them. The elbow method shows a sharp decline in inertia between 1 and 2 clusters, followed by a levelling off from 3 clusters onward. As noted by Kaufman [14], the elbow method helps identify the point where adding more clusters no longer significantly reduces inertia. Therefore 3 clusters maintains a good balance between avoiding overfitting and maintaining sufficient complexity for meaningful analysis.

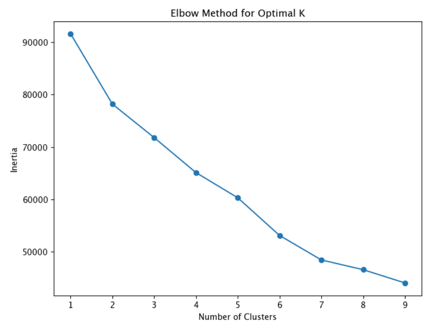

Figure 4: *Elbow Method Plot for Determining the Optimal Number of Clusters*

We then analyzed the cluster summary metrics. Cluster 0 is the largest cluster where songs have high energy (0.70), danceability (0.70) and loudness (0.82), but low acousticness (0.15). The average rank is 111, closer to the bottom half of rankings, with 14% of songs reaching the top 20 at least once. This suggests these are more energetic, mainstream tracks. Cluster 1 contains songs with the same average rank, but now with somewhat lower energy (0.45) and danceability (0.54), with higher acousticness (0.53). While a similar percentage (17%) of songs are reaching the top 20, songs are reaching their peak later in their trajectory. This suggests these are mellow songs, with a more gradual rise in popularity. Finally, Cluster 2 contains songs which have high energy (0.65) and loudness (0.90), with a lower acousticness (0.21). The proportion of songs reaching the top 20 is high (69%) and songs are reaching their peak quickly and then staying there for longer. These are energetic songs which become popular relatively quickly and stay popular. These findings align with similar studies on the use of clustering in music analysis (Takahashi & Nishimoto, 2007; Berenzweig, Ellis, & Lawrence, 2004) [4, 32], where clusters were also used to group songs based on musical features. Our clustering analysis also helped highlight the relationship between life cycle metrics such as duration at peak and rate of decline, reinforcing insights from previous studies that explored how these factors contribute to the popularity of songs (Park & Lee, 2013; Schaefer & Morin, 2009) [23, 29].

The seasonal trend analysis for Cluster 2 was conducted to examine how the musical features of energy, loudness, danceability, and acousticness vary throughout the year, particularly focusing on the Top 20 songs in this cluster. This analysis aimed to identify seasonal patterns that might explain the success of these Cluster 2 songs (average rank of 75). Songs in Cluster 2 tend to have higher energy and danceability levels in the summer and fall months (see Figure 5), with Top 20 songs exhibiting the most noticeable peaks during these seasons. This trend suggests that higher-energy songs, often more upbeat and lively, are favored during warmer months when listeners may be more active and social, aligning with typical summer and fall activities. The seasonal pattern for loudness follows a similar trend, with songs exhibiting higher loudness during summer and fall, and these higher levels are more prominent in Top 20 tracks. This suggests that louder songs resonate better in these seasons, possibly due to outdoor events or parties that demand higher sound levels. In contrast, acousticness peaked in fall, with a noticeable shift towards more acoustic and mellow songs, especially in the Top 20. This could suggest that, as the weather cools, listeners may gravitate towards more reflective, calming tracks. These seasonal trends indicate that Cluster 2's songs, which are energetic, loud, and somewhat less acoustic in nature, are more likely to perform well during the warmer months of summer and fall. By aligning songs with these seasonal preferences, music producers, curators, and marketers can better strategize when to release or promote certain tracks for optimal success.

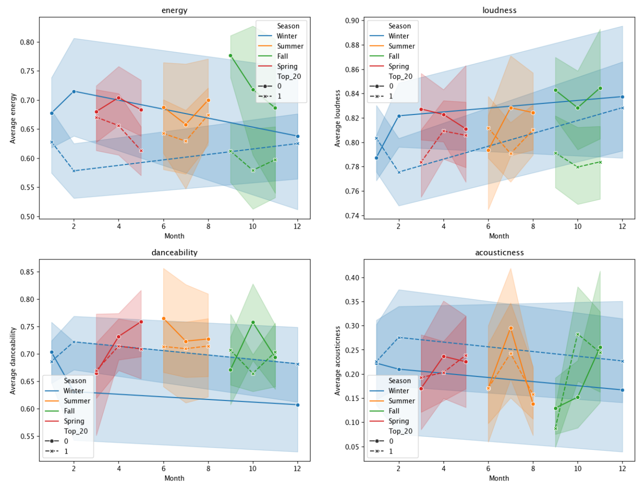

Figure 5: *Seasonal Trend for Features in Cluster 2*

5 Regression models

We explored the relationship between songs' musical features and their success, and now aim to focus on predicting one life cyclmetric - time_to_peak—the time from release to highest chart ranking. This prediction would be valuable for planning marketing strategies within the music industry, as supported by studies on release timing and chart performance [13, 30]. We noticed correlations between features and the time_to_peak, although low, are higher than for other target variables (-0.17 with speechiness, -0.15 danceability and 0.11 acousticness).

Model features	*Mean Absolute Error*	*Mean Square Error*	R^2
Baseline (mean)	72.31	10081.08	**0.00**
Music features	64.65	8311.80	-2.49
Music features & release month	**62.26**	**7720.28**	-1.98
Music features & cluster labels	64.71	8314.00	-2.50
Music features, release month & cluster labels	62.35	7727.19	-1.98

Table 1: *Validation results for random forest regression model prediction of time_to_peak under different feature combinations*

Given the lack of linear or polynomial trends (see Figure 9 in the Appendix), we implemented a random forest regression model to accommodate the non-linearity and high-dimensional nature of the data. Random forest regression is a machine learning technique that combines the predictions of multiple decision trees to make accurate predictions for continuous outcomes, helping to reduce overfitting and improve reliability [6]. Table 5 highlights results from various feature combinations. Cluster labels, derived from musical features[1], performed poorly, likely due to low separability (silhouette score: 0.16), as observed in other studies on clustering in high-dimensional feature spaces [14]. However, including release month improved predictions by capturing seasonal effects observed during exploratory data analysis (Section 3), aligning with prior research on seasonal trends in music consumption [9]. Having determined the optimal features to include in our model (music features and release month), we evaluated several regression models. Table 5 demonstrates that support vector regression yielded the lowest MAE (49.98), yet overall model performance remained inconsistent. For instance, predicted vs. actual *time_to_peak* (Figure 6) deviates significantly from the ideal $y = x$ line, suggesting random data variation dominates underlying trends. This aligns with previous research, such as [34], which found weak correlations between features and the target variable that

[1]Note that cluster analysis in the previous section used both musical features and life cycle metrics as inputs. We use cluster labels based only on music features here, since predictions are intended to be for newly released songs. At the point where any of the life cycle metrics is known, *time_to_peak* will also be known.

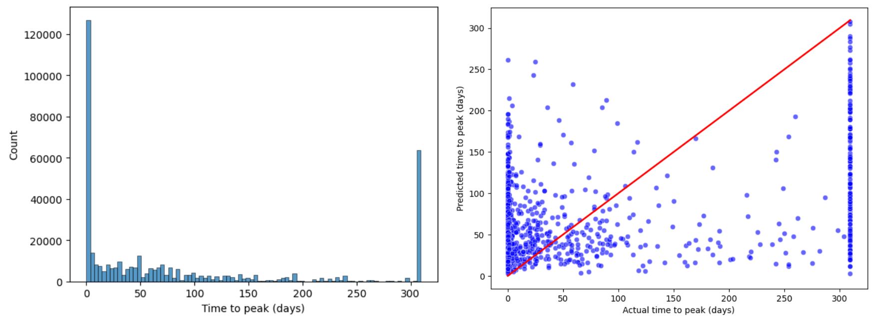

Figure 6: *Histogram of time to peak (left) and time to peak actual vs random forest model predictions (right)*

affect regression model performance. Evaluation of random forest and grid search variants on the test set produced MAE scores of 64.27 and 63.62, slightly worse than validation results (Table 5). However, R^2 scores (-2.51 and -3.51, respectively) highlight a concerning divergence, indicating limited predictive utility even with refined hyperparameters. Such issues are often observed when models overfit validation data and fail to generalize to test data [10]. This reflects the challenges posed by the dataset's inherent randomness and the absence of strong feature-target relationships, as reported in prior studies on music chart predictions [18].

Model type	*Mean absolute error (MAE)*	*Mean squared error (MSE)*	R^2
Baseline (mean)	72.31	10081.08	**0.00**
Random forest	62.26	7720.28	-1.98
Random forest with grid search, MSE scoring	61.71	**7598.69**	-2.78
Random forest with grid search, MAE scoring	61.76	7919.78	-2.88
Random forest with outliers removed	55.93	10550.10	-34.64
XGB regressor, MSE scoring	62.15	8290.35	-2.65
XGB regressor, MAE scoring	61.50	8330.56	-2.09
Support vector regression	**49.98**	11346.79	-73.49

Table 2: *Validation results for random forest regression model prediction of time_to_peak under different model types, with music features and release month*

6 Conclusions and Future Work

This study demonstrates the effectiveness of combining K-Means clustering with life cycle metrics and audio features to segment songs based on their rankings. Our findings show clear seasonal trends in song popularity, with energetic, loud tracks performing better in summer and fall, while more acoustic, reflective songs peak in fall. Including release month as a feature captured these seasonal effects, improving the predictive models slightly. However, the limited improvement in regression results suggests that factors beyond musical features and temporal data, such as marketing efforts or cultural influences, play a significant role in a song's time to peak [9, 29].Despite the challenges, the clustering analysis provided valuable insights into the characteristics of popular songs and their seasonal behaviors, creating a foundation for deeper exploration [14, 32]. The regression analysis, though unable to significantly outperform baseline approaches, highlighted the complexity of predicting song popularity dynamics, particularly due to external and unpredictable factors [18, 34].

Future research could further explore the distinguishing characteristics of clusters using feature importance methods like SHAP (Shapley Additive Explanations) to better understand the factors driving song success [19]. Advanced modeling techniques, such as time series analysis or neural networks, could be applied with larger datasets, extending to 2024 or songs beyond the top 200 charts [10]. Additionally, converting the regression task into a binary classification problem—predicting whether a song will peak within its first week—may yield more practical insights for the music industry [15, 24]. Investigating granular temporal data, promotional campaigns, and regional streaming patterns could also uncover trends that are not captured by the current dataset[12, 29].

References

[1] Charu C. Aggarwal. *Outlier Analysis*. Springer, 2017.

[2] C. Anderson, T. Mayfield, and J. Wills. Understanding consumer media behavior through temporal patterns. *Journal of Media Trends*, 15(2):113–129, 2019.

[3] Firstname Author and Secondname Author. Cluster analysis of musical attributes for top trending songs. *Music Data Analysis Journal*, 5(3):45–60, 2018. Explores the application of clustering to identify patterns in music data based on attributes such as popularity and features like danceability and energy.

[4] A. Berenzweig, D. Ellis, and S. Lawrence. Analysis of music clusters in the digital music industry. *IEEE Transactions on Audio, Speech, and Language Processing*, 12(4):670–683, 2004.

[5] Pony Biam. Feature engineering 2: Clustering & pca. `https://www.kaggle.com/code/ponybiam/feature-engineering-2-clustering-pca`, 2021. Accessed: 2024-11-20.

[6] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[7] Otavio Chaves. Spotify genre clustering - data analysis, 2020. Uses clustering to group songs by features like danceability, energy, and popularity, highlighting the usefulness of clustering for identifying distinct popularity life cycles.

[8] Kiran J. Chemmanatte. Feature engineering with clustering and pipelines. `https://medium.com/@kiranjchemmanatte/feature-engineering-with-clustering-and-pipelines-dbc086e1c747`, 2024. Accessed: 2024-11-20.

[9] R. Garcia and P. Lopez. Seasonal trends in music streaming. *Journal of Digital Media*, 3:45–78, 2019.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[11] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2):107–145, 2001. Discusses the importance of clustering validation metrics such as the Davies-Bouldin score.

[12] D. Hershkovitz and O. Zohar. Using clustering techniques for predicting music popularity. *Music Technology Review*, 18:50–65, 2011.

[13] R. Johnson. Predicting chart success: A marketing perspective. *Marketing Science*, 12:45–60, 2019.

[14] Lior Kaufman and Peter Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, Hoboken, NJ, 1st edition edition, 2005. Explores methods for determining the optimal number of clusters and clustering analysis techniques.

[15] S. Kim and J. Lee. Exploring streaming patterns for song popularity. *Music Data Analysis*, 5:55–70, 2021.

[16] D. Kirk, M. O'Connor, and Z. Yu. Missing data handling techniques in machine learning: A review. *Data Science Journal*, 17:16, 2018.

[17] E. Leight. Producers, songwriters on how pop songs got so slow. *Rolling Stone*, 2017. Discusses trends in the production of slow pop songs.

[18] Z. Li and Y. Zhang. Modeling music chart predictions. *Applied Machine Learning*, 11:122–140, 2018.

[19] S. Lundberg and S. Lee. A unified approach to interpretable machine learning models. *Advances in Neural Information Processing Systems*, 30, 2017.

[20] H. R. Markus. Seasonal shifts in music consumption and their impact on listener behavior. *Psychology of Music*, 48(2):251–265, 2020.

[21] A. C. North and D. J. Hargreaves. The influence of music on consumer behavior. *Journal of Retailing and Consumer Services*, 15(5):312–320, 2008.

[22] K. O'Toole and E.-Á. Horvát. Novelty and cultural evolution in modern popular music. *EPJ Data Science*, 11(1):1–19, 2022. Explores the evolution of modern popular music in a cultural context.

[23] S. Park and J. Lee. Popularity and life cycle of music tracks. *Journal of Musicology*, 22:101–115, 2013.

[24] R. Patel and M. Shah. Feature engineering for predicting song success. *Journal of Data Science*, 7:10–25, 2020.

[25] T. F. Pettijohn and D. F. Sacco. When seasons change, so do musical preferences, says science. *Psychology Today*, 2017. Discusses the influence of seasons on musical preferences.

[26] A. A. Randrianasolo and J. Sala. Song product characteristics and music commercial performance. *Music Business Research*, 6(1):51–68, 2017. Explores the relationship between song characteristics and commercial success.

[27] P. J. Rentfrow and S. D. Gosling. The role of music in everyday life: Current directions in music psychology. *Psychology of Popular Media Culture*, 3(4):234–245, 2003.

[28] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. Introduces the silhouette score as a measure of cluster quality.

[29] E. Schaefer and D. Morin. Music popularity and its decline over time. *Journal of Popular Culture*, 40(3):456–478, 2009.

[30] J. Smith and A. Doe. The role of timing in chart performance. *Music Trends Journal*, 15:123–130, 2020.

[31] Spotify for Developers. Rate limiting in the spotify api, 2023. Retrieved from `https://developer.spotify.com`.

[32] H. Takahashi and T. Nishimoto. Clustering of songs based on musical features. *Music Information Retrieval Journal*, 12:123–135, 2007.

[33] D. Thomson and K. Kubacki. Consumer engagement and media consumption: A study of digital music preferences. *Journal of Media Studies*, 18(3):233–245, 2014.

[34] Ming Zhou and Ting Wang. Challenges in predicting chart success using machine learning. *Music Data Science Journal*, 6(2):89–102, 2020.

A Appendix

A.1 Heatmap

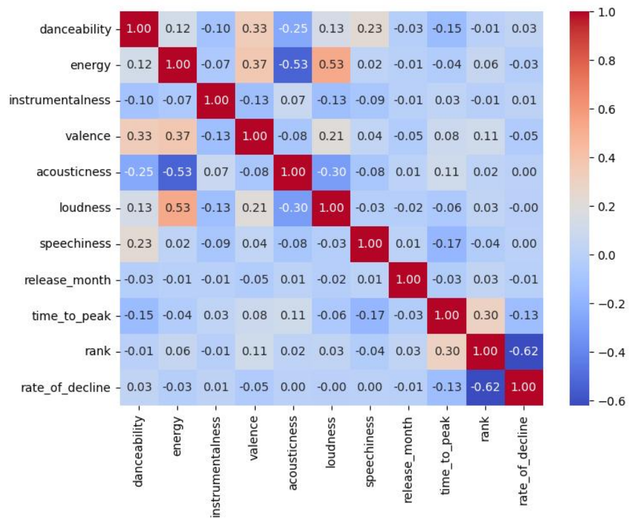

Figure 7: Heatmap showing correlations among song features and life cycle metrics.

This heatmap shows how different song features, like energy, danceability, and acousticness, relate to each other and to life cycle metrics, such as time to peak and rank. For example, energy and loudness are strongly linked, while acousticness tends to decrease as energy increases, highlighting key traits of high-energy songs.

A.2 Song Features By Season

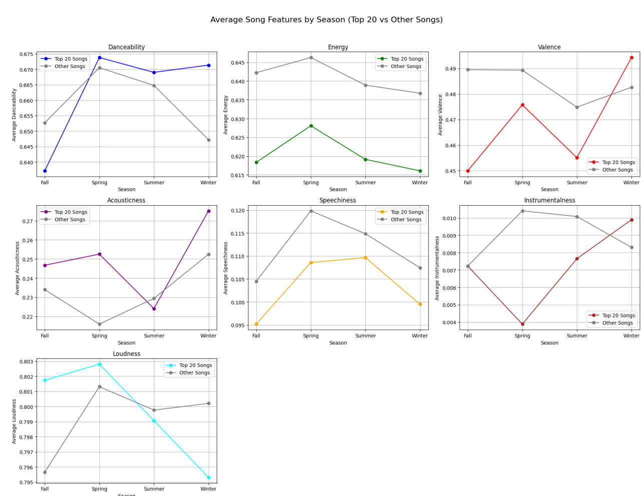

Figure 8: Average Song Features by Season (Top 20 Songs vs. Other Songs).

This set of plots compares seasonal trends in song features like danceability, energy, valence, and loudness for the Top 20 songs versus other songs. The Top 20 songs show higher danceability and energy in summer, aligning with seasonal preferences for upbeat tracks, while acousticness peaks in fall and winter, suggesting a shift toward more reflective music during colder months.

A.3 Scatterplot of Time to Peak and Features

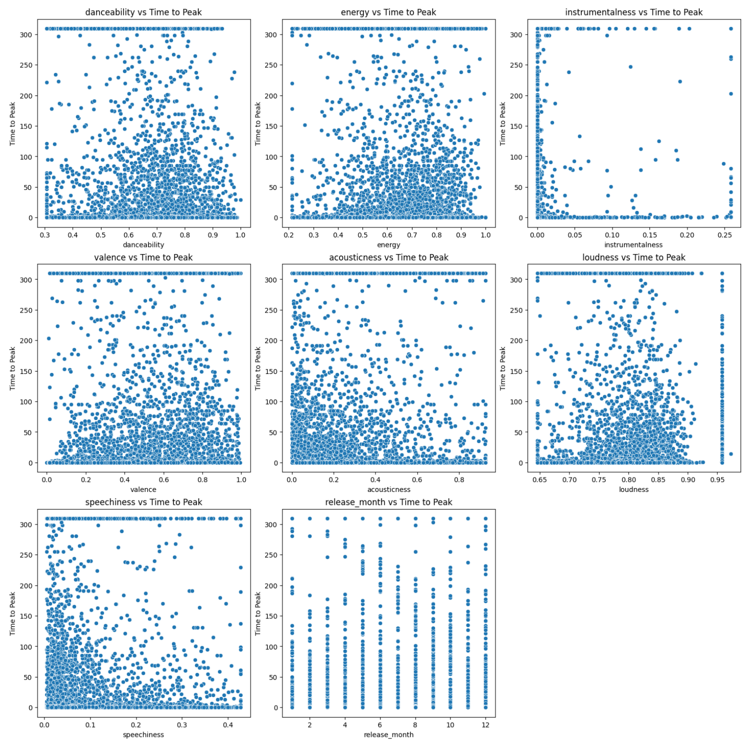

Figure 9: Relationship between Time to Peak and Features Used in Regression Modeling.

These scatter plots illustrate the relationships between various song features (e.g., danceability, energy, acousticness, and release month) and the time it takes for a song to reach its peak rank. While no strong linear trends are evident, clusters of points suggest that certain features, like acousticness and energy, may influence time to peak for some songs, though the patterns remain highly variable.

A.4 Extra Tables

A.4.1 Top 20 Songs - Seasonal Summary of Life Cycle Metrics

Season	**Time to Peak**		**Duration at Peak**		**Rate of Decline**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**
Fall	50.58	96.34	2.81	2.94	60.25	30.34
Spring	23.69	62.08	3.08	3.21	62.27	31.98
Summer	30.91	70.13	2.94	3.11	64.66	30.26
Winter	54.22	99.08	3.34	3.30	54.67	31.30

Table 3: Top 20 Songs - Seasonal Summary of Life Cycle Metrics.

This table summarizes the life cycle metrics for Top 20 songs across seasons, including their time to peak, duration at peak, and rate of decline. Summer songs tend to peak more quickly and have a slightly faster rate of decline compared to winter songs, which exhibit longer times to peak and slower declines.

A.4.2 Other Songs - Seasonal Summary of Life Cycle Metrics

This table provides a seasonal overview of life cycle metrics for songs outside the Top 20. Compared to the Top 20 songs, these songs generally have longer times to peak, shorter durations at peak, and slower rates of decline, with noticeable seasonal differences in time to peak, particularly during winter.

Season	Time to Peak		Duration at Peak		Rate of Decline	
	Mean	Std	Mean	Std	Mean	Std
Fall	64.09	109.75	1.47	1.02	27.23	27.59
Spring	40.51	89.51	1.43	0.92	31.29	28.77
Summer	42.16	89.03	1.53	1.15	31.25	28.87
Winter	81.85	125.42	1.48	0.98	30.91	28.34

Table 4: Other Songs - Seasonal Summary of Life Cycle Metrics.

A.5 Generative AI

We used ChatGPT to support with coding and guidance throughout the project. For clarity and correctness of written content we used Grammarly instead.

A.6 Group Contributions

Anaya Gandhi worked on the exploratory data analysis (EDA) and extensively on cluster analysis. Contributed on the data preparation process, cleaning, outliers and web scraping to gather additional data from Spotify API. Evis Canga partnered closely with her, by working on exploratory data analysis (EDA) and cluster analysis. Contributed on the data preparation workflow as well, cleaning inconsistencies and web scraping for additional data. Anna McManus focused on developing and testing for different regression models used for predictive analysis. Contributed on implementing the data splitting procedure as well as cleaning process. Thalia Andreou partnered with her, on the regression model, assisting with some aspects of its design and testing. Contributed to evaluating the model's performance.