Hyphenova Song Engagement and Popularity Analysis

Access to Github with all notebooks/analysis: [Here]

Problem Statement and Scope

Our goal for this project is to analyze different aspects and features of songs. We closely look at both instrumentals and lyrics surrounding a song to determine various characteristics of the songs such as sentiment, valence, tempo, acousticness, and so on. Through this analysis we aim to understand **what makes a song trendy/engaging** and hit the top charts. Another objective of this analysis is to understand **which of these characteristics of a song make it suitable for a given application/occasion**, for example, what characteristics of a song make it suitable to be used in a TV commercial or a spa setting.

The motivation behind this project is, there are many content creators and artists and there are billions of songs to choose from. With the help of this project, we aim to narrow down the selection based on the clients' business use case. And more importantly, we want to determine what makes a song engaging/trendy.

Approach

Our approach is to use the following techniques to achieve our goal. We will run a **Trendy Song EDA/ Popularity Prediction** which will be an exploration to understand what makes a song trendy. We will use **Clustering** to discover musical features that affect song grouping, finding common themes/ musical features/use cases for songs within a cluster with the hope to learn something that will help to predict use case in future steps. Another approach to clustering that we will explore in future steps is starting from songs that already have business use case labels and run a separate analysis on each business use case. For example, analyze/cluster songs that are used in spas.We will conduct **Sentiment Analysis** to understand the sentiment and theme of the songs, i.e. positive, negative, neutral - what makes them positive, negative using either lyrics, musical/audio features or a combination of both. Finally, we will perform **Feature Extraction/Machine Learning** to extract musical features using music libraries given the audio track of the song then move to predicting if a song will be trendy using popularity metrics (streams, spotify charts, apple music charts, etc.).

Initial data analysis and expected findings

Trendy Song EDA

We performed an EDA on this dataset (paste link) to identify what makes a song trendy. This was done to discover insights that would allow us to predict whether a song will become popular. For the purpose of this initial/preliminary analysis, streams were used as an indicator of a song's trendiness. Future analyses will look into other measures of trendiness such as spotify charts. We got the following preliminary insights:No single audio feature strongly correlates with streams, indicating complex factors influence popularity. **(Figure 5)**.Danceable, energetic, and positive songs may have a slight edge in becoming trendy.**(Figure 9 & Figure 10)**. Artist popularity (e.g., Taylor Swift, The Weeknd, Bad Bunny) likely plays a significant role in a song's success. **(Figure 8)**. Playlist inclusion, especially on Spotify, may be crucial for increasing streams. **(Figure 6)**

Overall, these insights suggest that to predict and promote song popularity, one should focus on:Creating energetic, danceable tracks with positive sentiment, leveraging the reputation of established artists, securing placement on Spotify playlists, recognizing that while many songs achieve moderate success, breakout hits are rare but impactful, understanding that the streaming landscape is dynamic, with performance varying by year.

Clustering

We performed clustering on a Kaggle Dataset that contains Top Global rated Spotify songs in 2023. We used the following musical features to perform the clustering: **Acousticness, Danceability, Energy, Instrumentalness, Liveness, Speechiness, and Valence** available on Spotify API **.**We experimented with 3 clustering techniques, **Kmeans**, **Hierarchical clustering**, and **DBSCAN**. For each clustering technique, we ran 2 models (one model using all features, and another model with features that highly influence the clustering as predetermined by an ML model of our choice or other methods). Below is a summary of the insights we got from clustering:

Hierarchical Clustering

Using the model with all features, we discovered that the features that influence the grouping of songs the most are speechiness and accousticness, followed by energy and other features as shown in **(Figure 1).** This model gave us 4 clusters.

We felt that those results were not too conclusive so we ran another model that uses the top two features (acousticness and speechiness) that highly influence the clustering according to Figure 1. This model gave us 3 clusters.Cluster 1 features popular, contemporary songs with mainstream appeal, ideal for creating engaging atmospheres in high-traffic areas and attracting broad audiences on streaming platforms. Examples include Taylor Swift's "Cruel Summer" and Dave and Central Cee's "Sprinter".Cluster 2 comprises reflective, emotionally driven songs with mellow tones and acoustic elements, suitable for calm environments like coffee shops and spas, and for relaxation playlists. Notable tracks include David Kushner's "Daylight" and Billie Eilish's "What Was I Made For?".Cluster 3 consists of high-energy, rhythmic tracks suitable for fitness centers, nightlife venues, and energetic playlists. Songs like Jung Kook's "Seven" and Myke Towers' "LALA" exemplify this cluster.

In terms of musical features **(Figure 2)**, Cluster 1 (186 songs) features tracks with high speechiness and slightly negative acousticness, likely representing hip-hop or rap genres. These songs are characterized by significant spoken word or rap elements and a more produced sound.

They appeal to audiences who enjoy vocal-heavy music with strong lyrical content, making them suitable for urban settings or youth-focused marketing campaigns.Cluster 2 (149 songs) comprises tracks with low speechiness and very high acousticness, suggesting folk or acoustic genres. These songs feature minimal spoken words and predominantly acoustic instruments, appealing to listeners seeking soothing, natural sounds. They are ideal for use in coffee shops, relaxation playlists, or nature-related media.Cluster 3 (618 songs) consists of tracks with moderate speechiness and low acousticness, likely representing mainstream pop. These versatile songs balance spoken words and singing, with a mix of electronic and organic elements. Their broad appeal makes them suitable for commercial use, radio play, and public spaces targeting diverse audiences.

KMeans
Using the model with all features, we discovered that the features that influence the grouping of songs the most are liveness and valence, followed by danceability and other features as shown in **(Figure 3)**. This model gave us 4 clusters.

We ran another model that uses the top three features (liveness, valence, and danceability) that highly influence the clustering according to Figure 3. This model gave us 4 clusters.Cluster 0 features a diverse mix of contemporary tracks from genres like reggaeton, trap, and modern R&B, suitable for creating upbeat atmospheres in trendy spaces and retail environments. Notable examples include Bad Bunny's "WHERE SHE GOES" and SZA's "Kill Bill".Cluster 1 comprises high-energy, contemporary pop and hip-hop tracks with mainstream appeal, ideal for energetic settings like fitness centers, retail stores, and events. Examples include Bad Bunny and Grupo Frontera's "un x100to" and Harry Styles' "As It Was".Cluster 2 consists of rhythmic and trendy tracks, often from hip-hop and reggaeton genres, perfect for high-energy environments like fitness centers, nightclubs, and social events. Songs such as Jung Kook's "Seven (feat. Latto)" and Dave and Central Cee's "Sprinter" exemplify this cluster.Cluster 3 includes emotionally charged and introspective songs, suitable for creating reflective atmospheres in spaces like cafes, bookstores, and relaxation areas. Examples include Olivia Rodrigo's "vampire" and David Kushner's "Daylight".

In terms of musical features **(Figure 4)**, **Cluster 0 (243 songs)** features tracks with moderate danceability, slightly negative liveness, and low valence. These songs have a balanced rhythmic quality, making them versatile for various environments such as commercial spaces and offices, providing engaging yet background-friendly music.**Cluster 1 (139 songs)** comprises tracks with low danceability, high liveness, and low valence. These songs have strong live performance characteristics, ideal for venues, events, or live performance atmospheres. They are also suitable for streaming services and radio stations focused on live recordings.**Cluster 2 (324 songs)** consists of tracks with high danceability, slightly negative liveness, and high valence. These energetic and rhythmic songs are perfect for high-energy environments like gyms, parties, or retail spaces. They are useful for streaming platforms and radio stations creating playlists for workouts or social events.**Cluster 3 (247 songs)** features tracks with low danceability, slightly positive liveness, and low valence. These introspective and less rhythmic songs are suitable for calm and reflective settings such as bookstores, cafes, or relaxation areas. They are also useful for streaming services curating playlists for studying or relaxation.

DBSCAN

DBSCAN was not conclusive. Despite efforts to tune the hyperparameters (see notebooks) and using a larger dataset, the algorithm kept yielding 1 cluster with all other data points classified as noise. Therefore, we will not be using it on our dataset moving forward as it may not be suitable for it.

Feature Extraction

In this stage, we aim to try on different music libraries to extract different musical features from the audio track and prepare it for further investigation in the future. We use the features provided by the Spotify API as our direction for the feature extraction. Tempo, measured in beats per minute (BPM) using `librosa`, indicates the speed of a piece, with higher tempos often correlating with more energetic songs. Energy, calculated as the mean Root Mean Square Energy (RMSE) via `librosa`, reflects the intensity of the audio signal, where higher energy suggests more dynamic and lively segments. Loudness is measured using the Short-Time Fourier Transform (STFT) in `librosa`, giving a mean amplitude in decibels (dB). Danceability, also extracted by Essentia, assesses a track's suitability for dancing based on rhythm stability, beat strength, and tempo, with higher scores indicating rhythmic and steady tracks. While for the instrumentalness, it measure the percentage of duration of song that has no vocal and only instrument, hence we make use of Spleeter which is a pre-trained convolutional neural network (CNN), to separate audio tracks into vocals and accompaniment by leveraging a that processes the input audio's spectrogram. The tool isolates the vocal components from the accompaniment, producing two separate audio tracks. To calculate the instrumentalness of a track, we first use Spleeter to obtain the vocal track and then apply a Voice Activity Detection (VAD) algorithm to determine the segments where vocals are present. By calculating the total duration of these vocal segments and the total track duration, we derive the vocal presence ratio. Instrumentalness is then calculated as one minus the ratio of vocal duration to total duration, representing the proportion of the track that is instrumental. This combined method provides a comprehensive analysis of the vocal and instrumental elements in audio files..

Challenges:

The project encountered significant challenges in aligning client expectations and understanding objectives, leading to late-stage scope adjustments. Technical difficulties arose in extracting complex features like valence, requiring advanced techniques for accurate mood prediction. Dataset limitations, including the lack of a unified dataset and absence of use case labels, resulted in fragmented analysis and difficulties in exploring song features for diverse applications. To address these issues, the team proposed clustering songs based on musical features and researching how feature combinations influence mood and usability, aiming to provide more cohesive insights and better meet the client's expectations for understanding song characteristics across various applications.

Limitation and Risks:

As we are unable to use the original Spotify songs for feature extraction, we cannot determine if the features we extract are similar to those provided by Spotify. Our results might differ significantly from Spotify's and may not be reliable for future use case analysis.
Given that we did not have a labeled dataset for feature extraction and also clustering, and that we try to label the songs by doing a clustering ourselves, the result might not be accurate and subjective to changes, as music often encompasses more than just quantifiable variables.
Given this limitation, the client wishes to label each song by use case (e.g., Spa, Trailer, Sports) for analysis manually using subjective feeling of herself in the next step, which potentially leads to another risk. This labeling process can be highly subjective. For example, while the client may find a song suitable for a spa, the general public or other listeners might disagree. Manual labeling might introduce too much subjectivity, causing the model to fall short of the project's ultimate goal: engaging the audience in different scenarios with appropriate music.

Next Steps

Our approach has been to analyze song features and determine their business applications. In the next phase, we will start with songs already categorized by application areas and analyze why they were categorized as such. We will apply these insights to understand what affects song engagement and trendiness. For **Feature Extraction:** We will investigate additional technical features, such as mel frequency cepstral coefficients (MFCCs), using a new dataset labeled by mood and genre from Pixabay. We will identify the most important features for category classification.Regarding **Clustering:** After labeling the data, we will use Spotify API features to study how different features contribute to various use cases. These features will train a supervised machine learning model to predict the use cases. For **Sentiment Analysis:** We will analyze song lyrics to assign sentiment scores (very negative, negative, neutral, positive, very positive).

Conclusion

A song's popularity and engagement are influenced by multiple factors. Danceable, energetic, and positive songs have a slight edge in trendiness, but the artist's existing popularity is crucial. Inclusion in popular Spotify playlists significantly boosts streams. High-energy environments favor songs with high danceability, slightly negative liveness, and high valence. Mainstream pop tracks with moderate speechiness and low acousticness appeal broadly across various settings. Contemporary genres like reggaeton, trap, and modern R&B create engaging atmospheres in trendy spaces.

Appendix

Figure 1: Features' Importance in Hierarchical Clustering

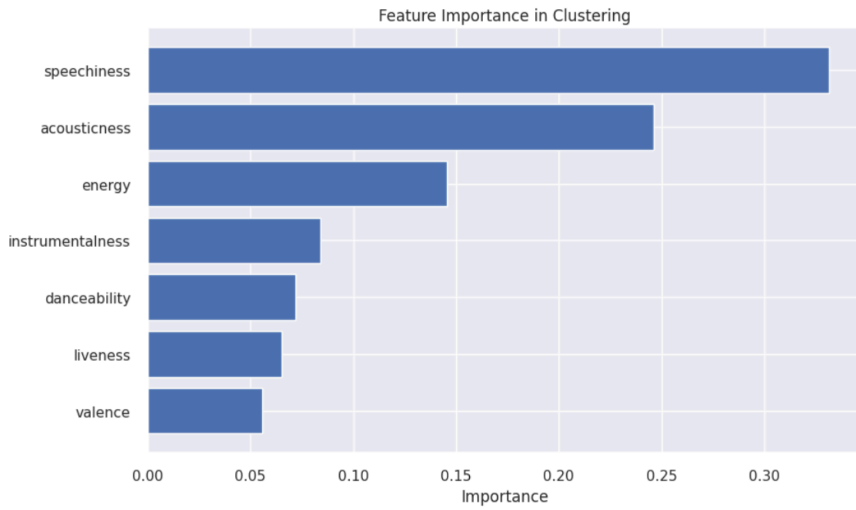


Figure 2: Clusters' Musical Features Statistics using Top 2 features (Speechiness and Acousticness) in hierarchical clustering

	cluster	index	speechiness	acousticness
0	3	count	618	618
1	3	mean	-0.415332	-0.405219
2	3	std	0.29503	0.581925
3	3	min	-0.820693	-1.04138
4	3	25%	-0.618829	-0.887435
5	3	50%	-0.517897	-0.618022
6	3	75%	-0.215102	0.0362663
7	3	max	0.592352	1.07543

	cluster	index	speechiness	acousticness
0	1	count	186	186
1	1	mean	1.79322	-0.069264
2	1	std	0.887227	0.881999
3	1	min	0.390488	-1.04138
4	1	25%	1.09701	-0.733484
5	1	50%	1.65214	-0.348609
6	1	75%	2.38389	0.421142
7	1	max	5.43708	2.30703

	cluster	index	speechiness	acousticness
0	2	count	149	149
1	2	mean	-0.515865	1.76717
2	2	std	0.250679	0.468348
3	2	min	-0.820693	0.92148
4	2	25%	-0.719761	1.34484
5	2	50%	-0.618829	1.76821
6	2	75%	-0.416966	2.15308
7	2	max	0.592352	2.69191

Figure 3: Features' Importance in Kmeans Clustering

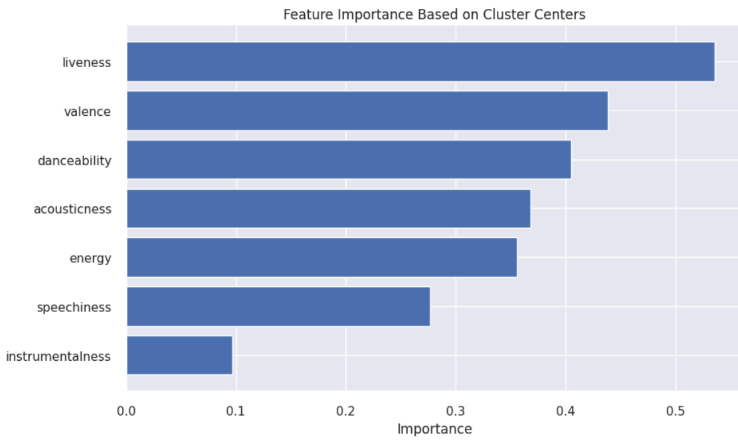


Figure 4: Clusters' Musical Features Statistics using Top 3 features (livenesss, valence, and danceability) in KMeans Clustering

	cluster	index	danceability	liveness	valence
0	0	count	243	243	243
1	0	mean	0.498229	-0.330852	-0.669636
2	0	std	0.552852	0.491636	0.576331
3	0	min	-0.339848	-1.11011	-2.02108
4	0	25%	0.00208099	-0.672285	-1.14756
5	0	50%	0.412395	-0.526342	-0.572315
6	0	75%	0.891096	-0.161486	-0.188819
7	0	max	1.91688	1.44388	0.194677

	cluster	index	danceability	liveness	valence
0	2	count	324	324	324
1	2	mean	0.622407	-0.335131	0.997572
2	2	std	0.655087	0.54788	0.50105
3	2	min	-1.36563	-1.11011	0.152066
4	2	25%	0.207238	-0.672285	0.56752
5	2	50%	0.685938	-0.526342	1.00428
6	2	75%	1.02787	-0.143243	1.34516
7	2	max	1.98527	1.58982	1.94171

	cluster	index	danceability	liveness	valence
0	1	count	139	139	139
1	1	mean	-0.157814	1.87541	0.176897
2	1	std	0.811265	1.06843	0.844005
3	1	min	-2.32303	0.641198	-1.89325
4	1	25%	-0.681776	1.152	-0.465789
5	1	50%	-0.203076	1.44388	0.194677
6	1	75%	0.412395	2.21008	0.748615
7	1	max	1.8485	5.74918	1.8991

	cluster	index	danceability	liveness	valence
0	3	count	247	247	247
1	3	mean	-1.21779	-0.290294	-0.749313
2	3	std	0.618146	0.523158	0.694928
3	3	min	-3.00689	-1.11011	-2.02108
4	3	25%	-1.57079	-0.599314	-1.2967
5	3	50%	-1.09209	-0.453371	-0.827979
6	3	75%	-0.750162	-0.161486	-0.23143
7	3	max	-0.408233	2.24656	1.04689

Figure 5: Correlation Matrix

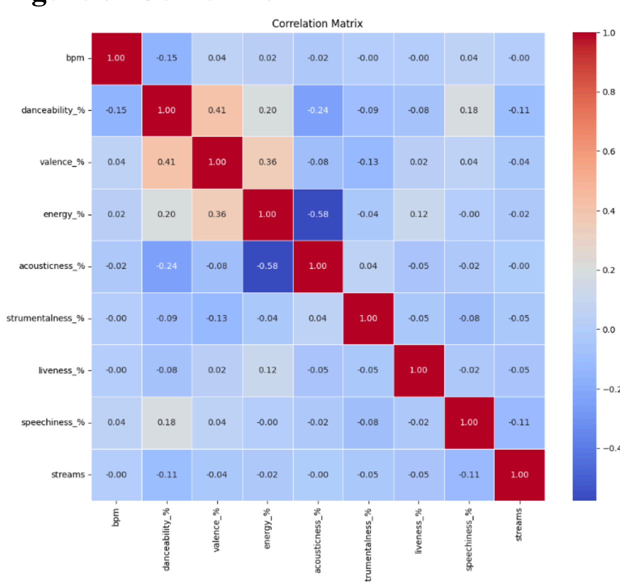


Figure 6: Proportion of songs in different platforms

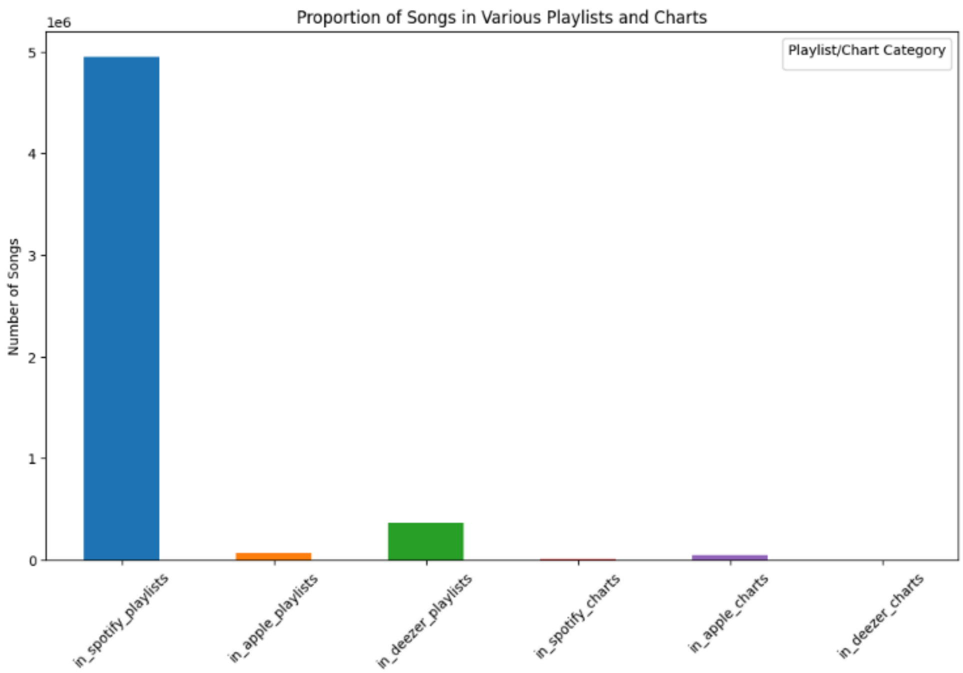


Figure 8: Artists with high stream counts

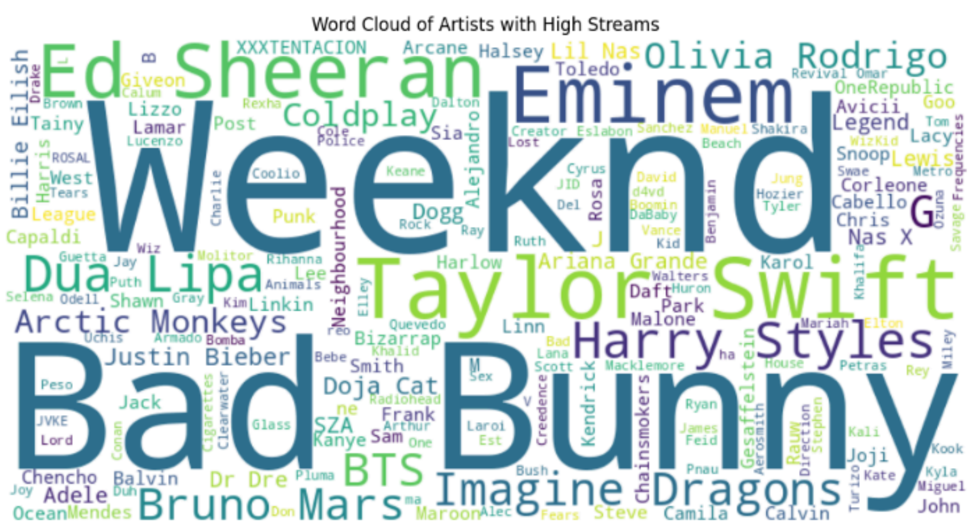


Figure 9: Average Audio features for top 10% and bottom 10% songs by stream count

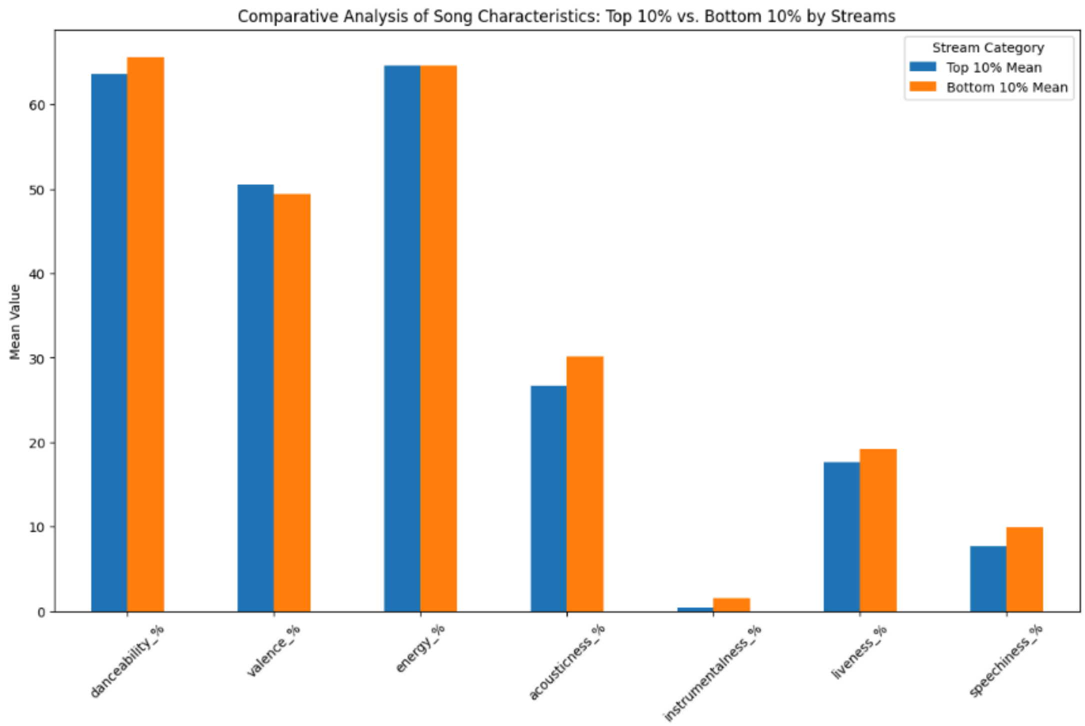


Figure 10: Song Trendiness feature importance

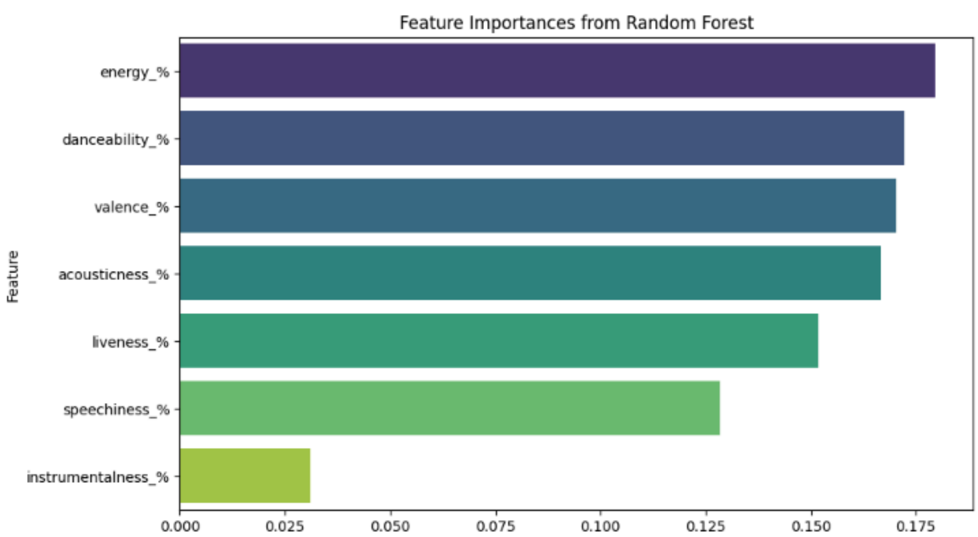
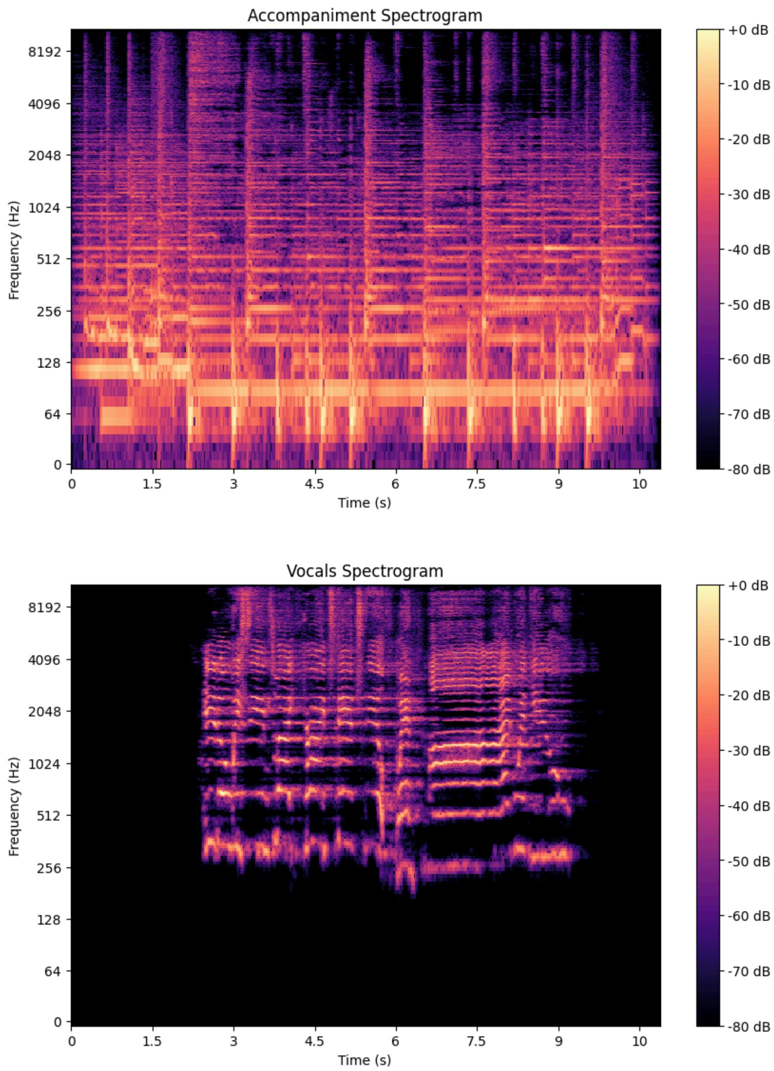


Figure 11: Frequency graph for Vocal and Accompaniment



References

Spleeter installation - https://github.com/deezer/spleeter/wiki/2.-Getting-started

Essentia library - https://essentia.upf.edu/algorithms_overview.html

Essentia library for key extraction - https://essentia.upf.edu/reference/streaming_Key.html

Spotify API documentation
-https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Sentiment Analysis Dataset-
https://www.kaggle.com/datasets/saurabhshahane/music-dataset-1950-to-2019

Feature extraction dataset -
https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data

Clustering dataset -
https://www.kaggle.com/code/nelgiriyewithana/an-introduction-to-top-spotify-songs-2023/input

Sentiment Analysis Models -
Getting Started with Sentiment Analysis using Python (huggingface.co)

Clustering Evaluation -
Clustering Evaluation strategies. Clustering is an unsupervised machine… | by Manimaran | Towards Data Science
Evaluation of clustering (stanford.edu)

Audio Classification -
https://github.com/jeffprosise/Deep-Learning/blob/master/Audio%20Classification%20(CNN).ipynb
https://towardsdatascience.com/5-websites-to-download-pre-trained-machine-learning-models-6d136d58f4e7
https://huggingface.co/tasks
https://huggingface.co/models