An Analysis of Popular Songs on TikTok and Spotify

Josie Rich, Erin Briggs, Anna Franz, and Katie Nuchols

Abstract—This paper delves into the realm of music trends in 2022, aiming to identify popular songs and understand the factors contributing to their widespread acclaim. Focusing on the intersection of media platforms, it presents a comprehensive analysis of TikTok's influence on Spotify trends, exploring the dynamic relationship of music consumption and user preferences within the Spotify platform. The study is to identify TikTok's influence on Spotify music trends specifically in the year 2022. We will examine two datasets to identify whether TikTok has an influence on everyday music streaming trends.

1 Objective

In the dynamic landscape of the digital music industry, platforms like TikTok and Spotify have emerged as the two most popular platforms to discover new music and shape the popularity and success of contemporary music. TikTok, with its short-form video format and viral challenges, has become a cultural phenomenon, catapulting songs to global stardom in a matter of days. On the other hand, Spotify, one of the leading music streaming platforms provides users with an extensive library of songs to discover and enjoy.

This project aims to delve into the intricate features of popular songs on TikTok and Spotify during the year 2022. By leveraging datasets sourced from Kaggle, we seek to uncover key patterns, similarities, and differences that contribute to the popularity of a song on these distinct platforms. We then try to apply these patterns in a machine learning model to try to create a tool that would help artists tailor their music to each platform. By exploring these two datasets, our project strives to contribute valuable knowledge to artists, music industry professionals, and enthusiasts, shedding light on the multifaceted factors influencing songs popularity across diverse online platforms.

2 Data

Our datasets originate from Kaggle, titled "Spotify Top Chart Songs 2022" and "TikTok Popular Songs 2022." We carefully selected datasets that offered a recent and accurate representation popular songs on both platforms throughout the entirety of 2022. Upon reading in both datasets, our initial focus was cleaning the data. We systematically examined all features from both datasets, addressing any instances of null values. Once we eliminated all null variables, we shifted attention to the fields, determining like characteristics from both datasets.

Regarding the Spotify dataset, we identified 646 songs and 17 fields, while the TikTok dataset was comprised of 263 songs and 18 fields. To conduct a meaningful comparison, we carefully selected the top 200 songs from each platform with their respective measures of popularity.

Spotify's measure of ranking was a feature called "weeks_on_chart", representing the number of weeks a song spent on the top of charts. In contrast, TikTok measured its popularity with a feature called track_pop which is a popularity score ranging from 0 to 100. Examining the top 200 songs from each set, we concluded there were 46 songs in common. Subsequently, we highlighted and printed the five most popular from each dataset.

Further analysis led us to categorize features into two main groups: mood descriptions and acoustic descriptions. Mood descriptions consisted of features such as dancebility, energy, speechiness, acousticness, liveness, and instrumentalness. On the other hand, acoustic descriptions consisted of features such as loudness, mode, key, tempo, and time signature. This refined categorization facilitated a deeper understanding of the distinct musical attributes to the popularity of songs on Spotify and TikTok.

3 Models

In this project, our analytical approach involved employing various models and algorithms to discern the similarities and differences between popular songs on Spotify and TikTok in 2022. We coded in a Jupyter Notebook using popular python libraries such as pandas and numpy to visualize and analyze the datasets. We utilized correlation matrices to investigate the relationship between different features and the popularity of songs on both platforms. This statistical method allowed us to quantify the degree of association between variables, helping identify mood and acoustic descriptions were most strongly correlated with a song's popularity on Spotify and TikTok.

We used linear regression to analyze the relationship

between song duration and popularity. This statistical technique enabled us to model and quantify the linear relationship between the duration of songs and their popularity to gain knowledge about whether song duration played a significant role in influencing popularity on each platform.

The decision tree algorithm was employed to leverage the correlated features identified in the datasets. by using these features, we aimed to develop a predictive model capable of determining the likelihood of a song becoming popular on either Spotify, TikTok, or both platforms simultaneously. Decision trees provide a structured and interpretable framework for decision-making, making them suitable for discerning patterns in complex datasets and predicting outcomes based on input features.

4 Results

We present a comprehensive analysis of the correlation matrices, linear regression models, and a decision tree classifier applied to the Spotify and TikTok dataset. We explored the intricate relationships between song features and popularity on these platforms. Figure 1-4 reveal distinct patterns in the correlation between mood and acoustic features with song popularity for the two platforms. Figures 5-6 uncover the subtle difference in average durations between popular songs on Spotify and TikTok. Finally, the application of a decision tree classifier in Figure 7 merges the datasets to provide a glimpse into the predictive capabilities of the model. Together, these findings contribute to a nuanced understanding of the diverse factors shaping song popularity on Spotify and TikTok in 2022.

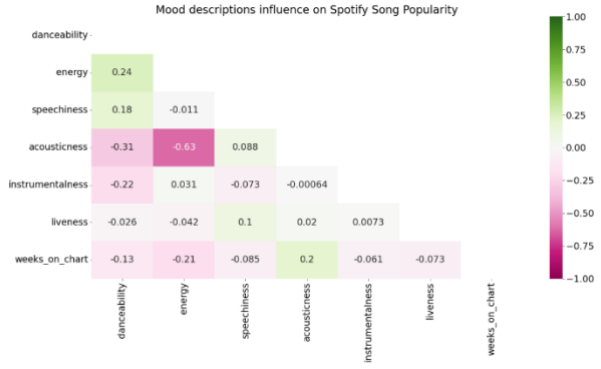

Figure 1: Mood features and popularity for Spotify

Figure 1 shows the correlation matrix for mood features and popularity for songs on Spotify. Acousticness and energy were the highest correlated with popularity with values 0.2 and -0.21 respectively. This means that as acousticness of a song increases, the popularity increases. On the other hand, as energy decreases, popularity increases.
Figure 2: Mood features and popularity for TikTok

Figure 2 shows the correlation matrix for mood features and popularity for songs on TikTok. Energy and liveness were the highest correlated with popularity with values -0.27 and -0.18 respectively. This means that as the energy of a song decreases, the popularity increases. Similarly, as liveness decreases, popularity increases.

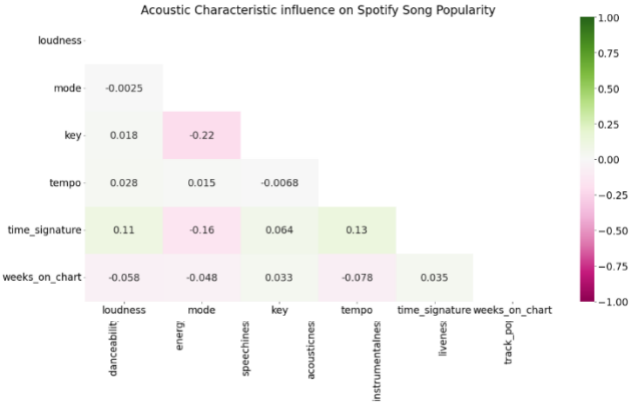

Figure 3: Acoustic features and popularity for Spotify None of these features are correlated with popularity.

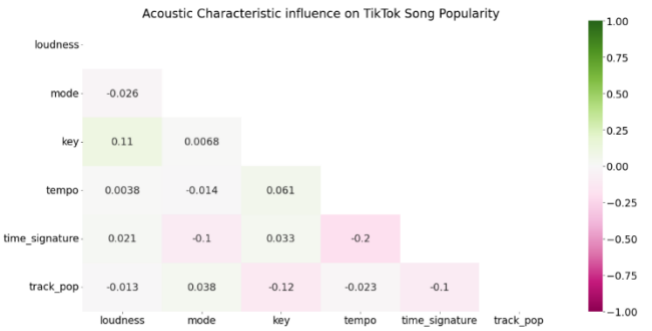

Figure 4: Acoustic features and popularity for TikTok

Figure 4 shows the correlation matrix for acoustic features and popularity for songs on TikTok. Time and keywere the highest correlated with popularity with values -0.1 and -0.12 respectively. This means that as the number of beats per bar (time) of a song decreases, the popularity increases. Similarly, as key decreases, popularity increases.

From these correlation matrices, we were able to determine that acousticnes and energy are the features most correlated with popularity for Spotify songs. For TikTok songs, energy, liveness, time, and key are most correlated with popularity.

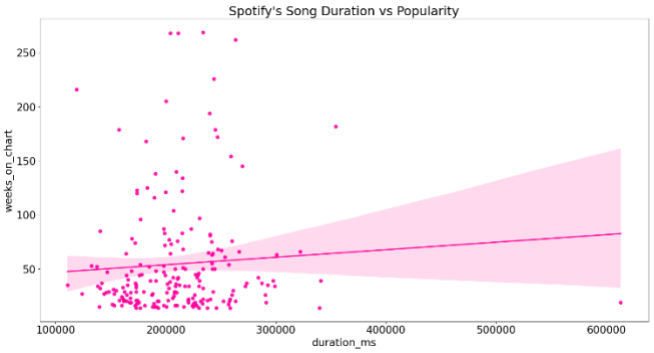

Figure 5: Linear regression of song duration and popularity for Spotify

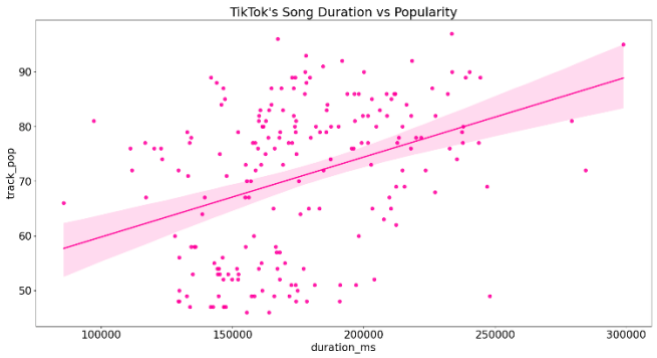

Figure 6: Linear regression of song duration and popularity for Spotify

As seen in Figure 5, the average Spotify song duration from the top 200 lies at about 200,000ms. On the other hand, in Figure 6, the average TikTok song lasts around 180,000ms. There is not a link between the duration of the track and popularity, but we did discover that popular TikTok songs are slightly shorter than popular Spotify songs.

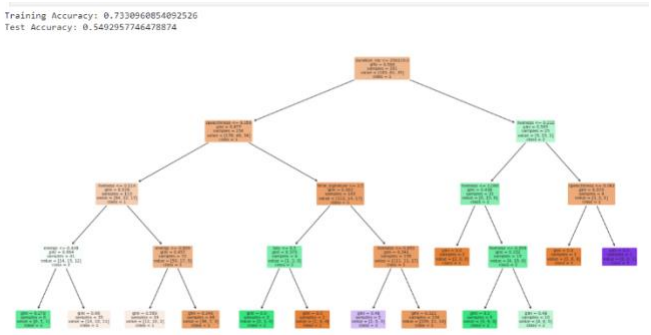

Figure 7: Decision Tree Classifier

We then applied these findings to a machine learning model. We merged the two datasets and added a popularity_label column. The value 1 would represent the song being popular on TikTok, 2 is popular on Spotify, and 3 is popular on both platforms. We were able to get a training accuracy of 0.73 and a testing accuracy of 0.55. This drastic decrease in accuracy from the training to testing is probably due to overfitting of the model to the training set and not being able to generalize to new data. The decision tree first splits on duration and then speechiness and liveness.

5 Primary Issues / Limitations

We are expecting to see a direct correlation between what was trending on TikTok and Spotify during the year 2022. TikTok is known for its videos going viral and people using the same songs repetitively, whereas Spotify is used as a music listening platform. We expect to see a decent number of the same songs to have been trending on both platforms during this time. However, our results show that there are very low correlations between the mood and acoustic features and song popularity. Our data analysis shed light on to the difficult concept of popularity in the dynamic landscape of music. Our machine learning model struggled to perform due to a couple of reasons. Our data is too small for the model to generalize well to new data. The dataset is imbalanced with only 46 songs that are popular on both platforms. The features we used are too weakly correlated to discern meaningful patterns.

6 Future Work

There are several promising directions to further enhance our understanding of song popularity dynamics on platforms like Spotify and TikTok. Firstly, a more extensive exploration into additional features, such as lyrics sentiment or genre-specific characteristics, could offer richer insights into the diverse factors influencing user preferences. Additionally, refining and optimizing the machine learning model, perhaps through advanced techniques like ensemble methods or neural networks, may improve predictive accuracy and address potential overfitting challenges observed in the decision tree classifier. Furthermore, expanding the dataset to include more recent years and incorporating user engagement metrics might yield a more comprehensive understanding of the changing landscape of music popularity on these platforms. Lastly, considering the interplay between user-generated content and song popularity, especially in the context of TikTok's user-driven trends, could provide a nuanced perspective on the symbiotic relationship between artists and their audience. Future research in these directions could deepen our insights into the multifaceted nature of music popularity.

7 Organization Chart

Katie will be responsible for importing and cleaning the data for the columns we will not be using. The only columns that will be kept are song title, artist, popularity, danceability, energy, and instrumentalness between both datasets.

Anna will be responsible for sorting the data into the top 200 songs from both TikTok and Spotify datasets. Spotify data sets popularity is determined by the

number of weeks spend on the chart. The TikTok dataset popularity is based on a scale of 0-100. She will collect the top 200 of both sets based on those values respectively.

Erin will identify the common songs and artists between the two sets based on the top 200. She will be responsible for creating graphs that show the comparisons between the rankings of the songs on each platform.

Josie will be responsible for determining what attributes make the songs popular and if there is a correlation between the platforms for popular songs. She will create visual depictions on what attributes trend between the platforms. She will also implement a machine learning model to the datasets to predict which songs will be popular on either platform.

We all contributed to the final presentation and final report for this project.