

Spotify Tracks Popularity Analysis

Jongwan Kim (jk2369), Richard Lin (rl556), Lingfei Zhao (lz382)

Course code: ORIE 5741

Github link: <https://github.com/Richardlin112/Spotify-Tracks>

1. Dataset Introduction

We will be exploring a dataset of almost 90,000 tracks on Spotify spanning across 125 genres, which was retrieved from Kaggle, from the user Maharshipandya. Each track is associated with a total of 19 features such as artist, tempo, danceability, and popularity score. These features include 8 nominal variables, 1 ordinal variables, 1 discrete variable, and 9 continuous variables.

2. Problem Statement

Some questions we can answer using our dataset are:

- What features make a song more popular on Spotify?
- What attributes make a song popular within it's own genre?
- How well can we predict a song's popularity based on various features of the song?

2.1 Why are these questions important?

Music has always been a universal language that surpassed cultural boundaries, playing a vital role in human expression and connection. It has constantly evolved over time, encompassing genres from jazz and country music, to hip-hop and electronic dance music. The evolution of popular music genres over time underscores the dynamic nature of audience preferences and global trends. Identifying features that contribute to a song's popularity, on platforms like Spotify, will allow businesses involved in the creation, promotion, or distribution of music to strategically allocate their resources and effectively cater to their target audiences.

2.2 How does this data help us answer our questions?

The dataset contains information on various song attributes and the respective popularity score ranging from 0 to 100, which was calculated based on the total number of plays the track has had and how recent those plays were. Therefore, we can fit a linear regression model to gauge the impact and significance of each feature on the popularity of a track. Due to the difficulty of converting nominal variables such as artist, album, and genre to numerical values for regression, we can also use other feature selection methods for the categorical data.

To answer the second question, we can conduct cluster analysis based on genre and determine if there are trends in attributes that would classify a song into a specific genre. We can also segment the data by genre and build a regression model on each to determine if the features that make a song popular are dependent on genre.

Finally, we can build a random forest model to predict the popularity of new songs based on the features we determined to be most important. This will allow us to classify whether or not future songs are likely to be successful.

The large amount of entries present in this dataset (just under 90,000 unique tracks), as well as the well-documented and diverse range of features, will allow us to present a more robust predictive model with clear and easily comprehensible results.

3. Applications and Insights

This project provides great utility and valuable insights not only to the music producers but also to Spotify. For the music producers, this project will allow them to get an insight into the components of a song that are the most critical in gaining popularity. With this insight, the producers will have more idea of what aspects of a song to focus on for an increased chance of "success". For Spotify, this project can add more value to the existing recommendation system. The current Spotify recommendation system shows any and all songs that roughly align with the user's preferences. With this project, however, this recommendation system can be refined to have an extra layer of filtering to only suggest songs that both align with the users' preferences and have the most likelihood of being popular to the users. This improvement will lead to a better quality recommendation system, which will help enhance the users' trust in Spotify's recommendations. Due to the great benefits and insights that this project brings as mentioned above, we think that this development will prove to be meaningful and worthwhile.

4. Parting Words

We are optimistic about this development as we are equipped with a solid foundation within the robust Spotify tracks dataset and also with a strong technical skill set in our team. The plethora of our previous experiences with data analysis and predictive modeling allows us to tackle this project in a structured and organized way. Our expertise, along with the comprehensive Spotify dataset we collected, allows us to be confident in the development of this project and the potential valuable insights it delivers to artists, producers, and Spotify.

5. References

- <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>