Predicting the Popularity of Billboard Top Songs on Spotify: A Regression-Based Analysis

Hadley Browder, University of Florida Department of Statistics

ABSTRACT

The idea of ranking songs has become increasingly popular as listeners request personalized playlists and song recommendations. One such company that aims to rank popular music is Billboard. Music streaming platforms such as Spotify use these rankings, along with songs' metadata features, to help tailor user-targeted playlists and song recommendations. This study aims to determine if a song's popularity can be predicted from its metadata features. The results of this study should help others interested in predicting song popularity to determine if popularity can be predicted on the sole basis of metadata features. 598 Billboard global top songs from 2010-2019, each with 11 listed features, were used to create a multiple regression model to predict popularity score. The model suggested that the included predictors explained very little variation in popularity scores of the top songs with relatively high error in the predictions. Ultimately, the model has low value in predicting a song's popularity score. Predicting song popularity on the basis of metadata fails to consider the elements of human emotion and behavior that influence how popular a song will be.

INTRODUCTION

Since its platform launched in 2008, the music streaming service Spotify has risen in popularity. With over 50 million tracks to choose from, Spotify has proven to be a major source of popular music. As such, Spotify closely monitors the rankings of songs from the Billboard Hot 100 standard record charts, updating their playlists accordingly. While Spotify provides access to metadata regarding these songs, they do not offer a clear explanation as to how those descriptors relate to a song's popularity in the Billboard rankings. This makes it difficult to determine why a song is popular and if it's possible to predict the popularity of a given song. Making predictions regarding song popularity holds great power in the arena of customized song suggestions for listeners and targeted playlists for certain audiences. This field of "hit song science" piques the interest of many private companies, however, the success of these companies' popularity-prediction algorithms is unknown (Pham et. al., 2015). Pham, Kyuak, and Park explored the most influential features for a song's popularity through the creation of various stepwise-selected models but did not definitively offer a final model for song prediction. Patchet and Roy concluded that "state-of-the-art machine learning" could not be used for predicting song popularity. However, the model used to form this conclusion worked with a set of over 600 predictors, suggesting that it may have suffered from overfitting.

The purpose of this study is to determine if a song's popularity can be predicted from its Spotify-classified features. The study is conducted using Billboard global top songs with descriptors listed by Spotify. The findings of this study are expected to help determine if a song's popularity can be predicted on the basis of its features. The remainder of this paper presents research methodology, followed by diagnostics and results, and concludes with a discussion on findings and potential improvements for further research.

METHODOLOGY

The target population for this study included the Billboard global top songs from 2010-2019, made available on Spotify. The unit of analysis was an individual top song. The data set used in this study was obtained from Kaggle user Leonardo Henrique. Henrique extracted the data from the Spotify service Organize Your Music. The data set includes 603 Billboard top songs from 2010-2019, each with the following variables listed: genre, beats per minute, energy, danceability, loudness (dB), liveness, valence, length, acousticness, speechiness, duration, and popularity score. In order to properly account for the qualitative variable of genre, coded variables were utilized. Song genre was grouped into two categories: 0 for songs that did not fit into a pop-type genre, and 1 for songs that did. The other predictors were measured quantitatively. 5 songs that had a reported popularity score of 0 were removed from the dataset, as these measurements were deemed to be errors in the data collection process.

As our variable of interest was popularity, a score measured on a scale of 0 to 100 with 100 being most popular, we aimed to find a multiple regression model that held the most predictive power from the features given. In order to identify the best-fitting regression model for the data, we utilized a stepwise Akaike Information Criterion (AIC) selection process. Stepwise AIC selection performs forward and backward selection, utilizing a formula that includes the number of model parameters and the log-likelihood of a model to find the best fitting model with the lowest AIC value. To this end, the general model form we fit is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

Our use of the multiple regression model is accompanied by the following model assumptions, which were used to evaluate the appropriateness of our determined multiple regression model:

- The regression function is linear.
- Error terms are normally distributed.
- Error terms have constant variance.
- Error terms are independent.

RESULTS

Our stepwise AIC selection process indicated that the features that should be included in the multiple regression model for predicting a song's popularity score are danceability, song duration, energy, and loudness.

Table 1: Significant Predictor Descriptions	
Danceability	Higher danceability values indicate songs that are easier to dance to.
Duration	The length of the song, as measured in seconds.
Energy	Higher energy values indicate songs that are more energetic.
Loudness (dB)	Higher dB values indicate louder songs.

The coefficients given in the multiple regression output yielded the fitted regression model:

$$\hat{Y} = 90.0467 + 0.0651dnce - 0.0422dur - 0.1695nrgy + 1.0406dB$$

Interpretations of model features are as follows:

- For every one-unit increase in danceability value, popularity score is estimated to increase by 0.0651units.
- For every one-second increase in song duration, popularity score is estimated to decrease by 0.0422 units.
- For every one-unit increase in energy value, popularity score is estimated to decrease by 0.1695 units.
- For every decibel increase in loudness, popularity score is estimated to increase by 1.0406 units.

The model's adjusted R-squared value was 0.03282, suggesting that 3.282% of the variability in song popularity score can be explained by the model. The model had a residual standard error (RSE) value of 13 on 593 degrees of freedom, meaning that the actual song popularity scores differ from the regression line scores by 13 units on average, a high value given the scale of popularity scores (0-100).

We evaluate the appropriateness of our model using the assumptions aforementioned in the methodology section. To satisfy the assumption of linearity, we examine the four variables from our model separately in relation to the response, popularity score. Based on the plots in Figure 1, we see that each of the four predictors has a linear, but weak, correlation with popularity score.

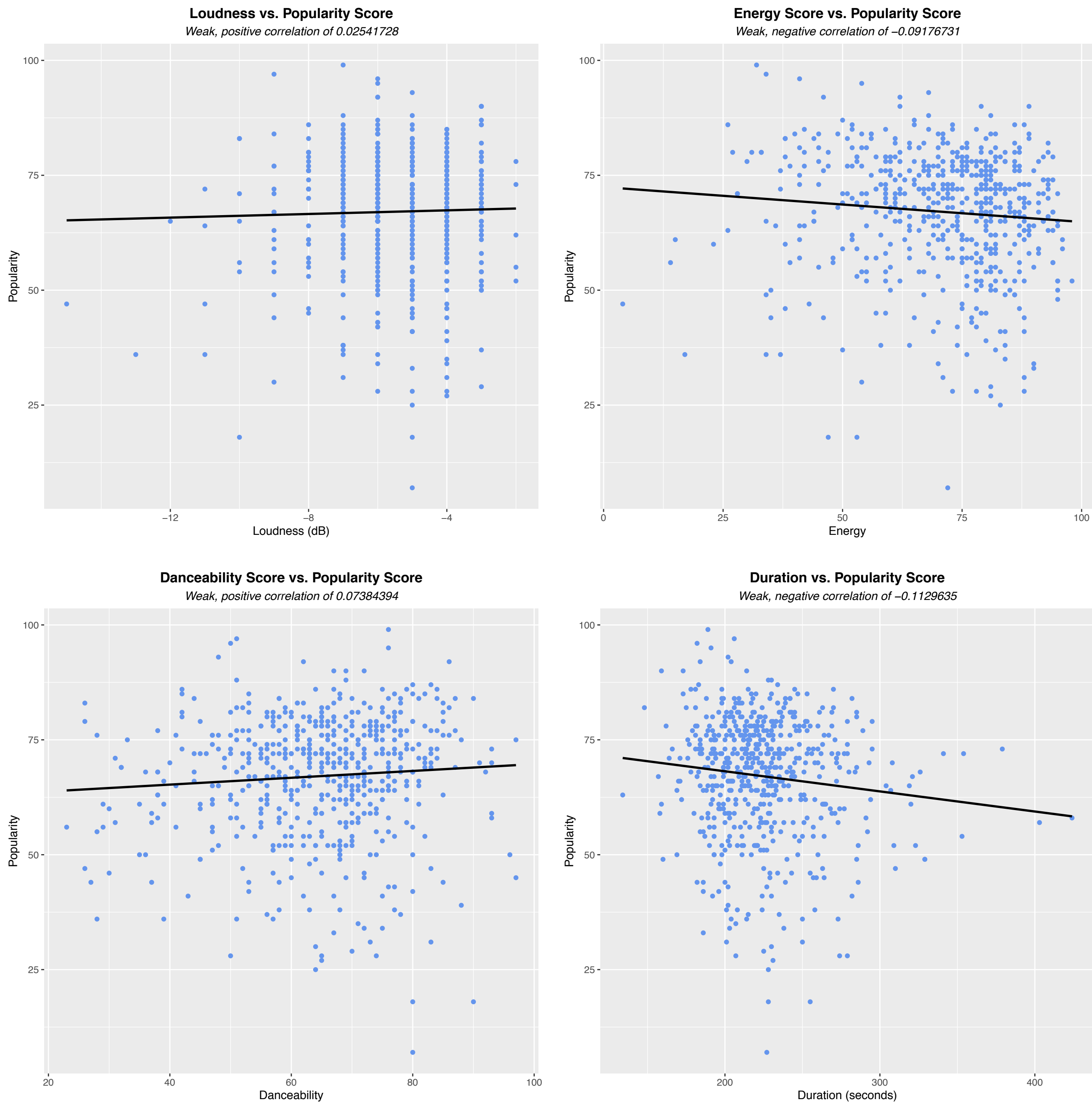

Figure 1: Individual Predictor Linearity Plots

We now proceed with checking error-related assumptions. Because our analysis did not take into consideration a specific order of songs, we will assume that the residuals are independent of one another, satisfying independence of error terms. Based on the residual analysis plots in Figure 2, we can assume that constancy of variance and normality of error terms are satisfied. The residuals appear to have a random scatter in the residuals versus predicted values plot. While the tail ends of our residuals in the normal probability plot are lower than would be expected under normality, these are not extreme deviations. With our assumptions satisfied, we determine that our multiple regression model fit is appropriate for the data.

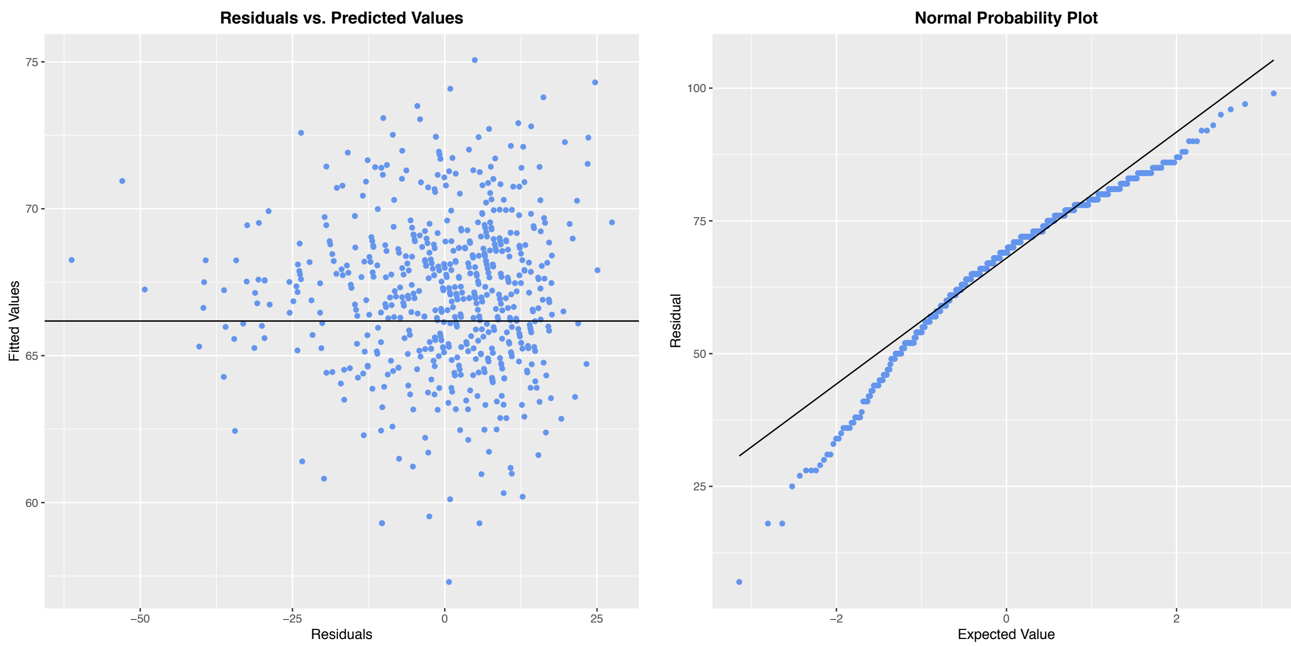

Figure 2: Residual Analysis Plots

DISCUSSION

This study investigated the use of various metadata features in predicting popularity scores of Billboard top songs on Spotify from 2010-2019. While a multiple regression model was fit and seemed appropriate for the data, the model's low adjusted R-squared value and relatively high RSE suggest that the model does not predict song popularity very well. This is consistent with previous studies that also found difficulty in predicting song popularity from metadata. The weak ability of our model to predict song popularity is unsurprising, as a song's popularity is dependent on those who enjoy it enough to listen to it. The popularity score of a song might be more related to the opinions and feelings of the people who listen to it as opposed to its inherent metadata features. It is nearly impossible to predict human emotion and behavior.

Possible limitations of this study include the relatively small number of predictors in the data set, the exclusion of year as a variable, and the broadness of our genre groupings. We selected our model from only 11 potential features, however, there are certainly more metadata features available for popular songs that may have strengthened the predictive power of our model. Our data set included songs from 2010-2019, but we did not include year as a variable in the model selection process, which may have led to skewed predictions. We used the pop genre as a means of coding song genre, but a more in-depth analysis of genre might have offered some insight into song popularity. Possible improvements to the study could mean the inclusion of the aforementioned qualities. Using the year that a song was popular on the Billboard charts as a blocking variable and creating varied groupings of song genres could lead to more valuable insights regarding song popularity. Exploration of models outside of regression might also provide a stronger means of predicting song popularity from metadata features.

WORKS CITED

Henrique, Leonardo. “Top Spotify Songs from 2010-2019 - BY YEAR.” *Kaggle*, Kaggle, 26 Dec. 2019, www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year.

Pachet, Francois & Roy, Pierre. (2008). “Hit Song Science Is Not Yet a Science”. 355-360.

Pham, James Q, et. al. (2015). “Predicting Song Popularity.”

Spotify. “Organize Your Music.” *Organize Your Music*, Spotify, 2016, organizeyourmusic.playlistmachinery.com/.