Aesthetic Preference in Music

based on a dataset from

Prepared for

Data Science 2: Statistics for Data Science

by

Group 11

Benhamida, Jamil

Laudrum, Eric

Shah, Akash

Treshan, Waas

August 2, 2021

Introduction

The dataset we decided to investigate possessed a plethora of musical feature, song metadata, and engineered variables obtained via NLP that describe the emotions present in a given song in the top 200 list of songs from a collection of 34 countries around the globe in the past 3 years. Additionally, corresponding data from the global top charts was available. A significant amount of data was available for analysis; hence, our group was interested in determining if global tastes in music could be quantified and compared based on differing geographic regions.

Data Source

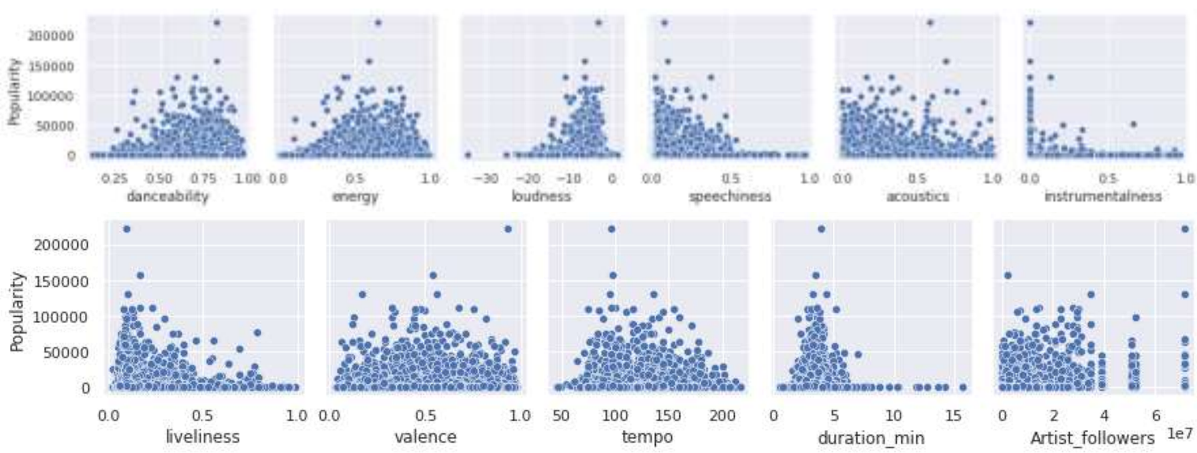
https://www.kaggle.com/pepepython/spotify-huge-database-daily-charts-over-3-years

Data preparation

Most engineered variables through natural language processing (generated by the data set creators) were only applicable to English-language songs; as such, for the purpose of our analysis and model generation we did not consider these factors in our analysis. This allowed us to access a considerable amount of information for hypothesis testing in different geographic regions compared to the global average. By extension we attempted to accurately predict the popularity of a song based on the following basic musical features:

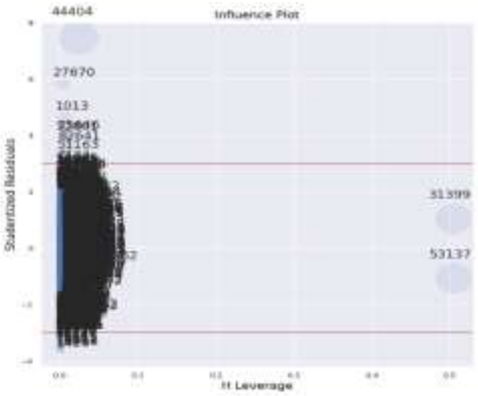
1. Popularity - A popularity grade generated by the team that pre-processed raw Spotify data through NLP and longevity of specific titles remaining on the Top 200 lists of their respective countries.
2. Danceability - A combined variable generated by the dataset creators that incorporated tempo, rhythmic stability, and beat strength to assess how suitable a track is to dancing. This is measured between 0 and 1.
3. Energy - A perceptual measure of the intensity of a track based on dynamic range, perceived loudness, timbre, onset rate and general entropy. This is measured between 0 and 1.
4. Loudness - Overall loudness of the track measured in decibels.
5. 'Speechiness' - The presence of spoken words in a track. This is limited between 0.0 to 1.
6. Acoustics - A subjective measure of the 'acousticness' of a track.
7. 'Instrumentalness' - A measure of a track's lack of vocals, this is limited between 0.0 and 1.
8. Liveliness - A measure of audience sound predominantly used to discern live recordings.
9. Valence - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)
10. Tempo - Overall estimated paces of the song measured in BPM (beats per minute)
11. Duration_min - The song length in minutes

Through visual inspection of the bevy of musical features available compared to the engineered popularity score we were unable to draw meaningful correlations in the data and were ultimately unsuccessful in generating a predictive model for this variable based on these factors. The only factors that appeared to have a clear relationship with the popularity score were metrics that tracked the length of time that each song was on the top 200 or top 50 list of their respective country. These were considered meaningless correlations as the popularity scores were generated from these variables by the dataset creators. We used a combination of a scatter matrix and Spearman's correlation data to come to these conclusions.



Data analysis

Though there were no visible features to predict popularity we noticed a clear relationship between the energy level of the songs in the database and their respective loudness level. As such, we decided to attempt predicting the variability in energy level as a function of loudness level. The resultant model obtained was described by loudness and the genre type.



Influence plot

OLS

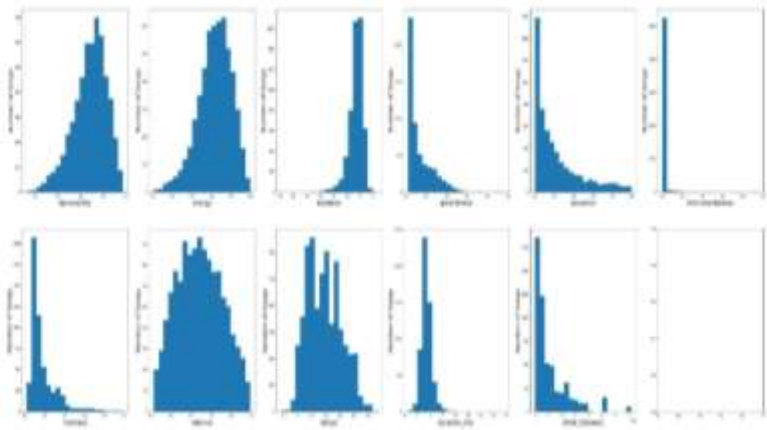
The initial model showed a weak R-squared of `0.545.` However, we used the influence plot and also feature P_Values to modify the model which possessed an R-squared of `0.637` which suggested only moderate explanation of variability. Looking at the relationships showed in scatter plots, we felt the model showed an acceptable level relationship considering the variability in features.

Random forest

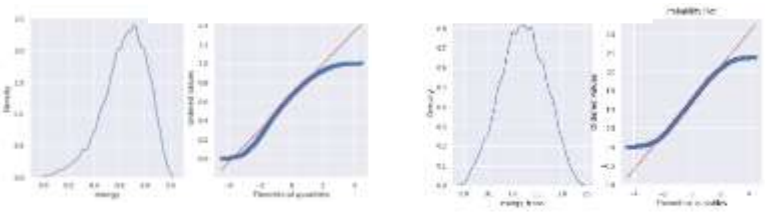
The random forest model accuracy score came to 0.589. This was expected considering the weak relationships in features. Since it was a continuous prediction we used the R2_score function to predict the accuracy.

Hypothesis testing

Along with the model development, we conducted some hypothesis testing on some of these key musical features between different ethno-linguistic regions and the global average to get a sense of variability in musical taste globally based on differing geographic regions.



Histogram view of selected features



a Yeo-Johnson power transformation

Plenty of these factors had non-normal/skewed distributions in their data and possessed negative values in certain cases. As such, a Yeo-Johnson power transformation was employed to conduct hypothesis testing (z-testing since all conditions were met following transformation) on these distributions as opposed to Box-Cox due to its added ability to normalize negative data.

From this analysis, danceability is significantly different in 'Southern Europe and Portuguese heritage' and 'English-speaking and Nordic countries' compared to the global average. Valence is significantly different in English speaking and Nordic countries. The group of 'Southern Europe and Portuguese heritage' and 'English-speaking and Nordic countries' appeared to favour danceability less than the global average. While 'English-speaking and Nordic countries' appeared to favour less positive music based on its lower valence compared to the global mean.

Conclusion

Although this dataset didn't allow us to predict popularity based on key musical indicators through OLS, it did prove its value to identify a trend between song energy. We were able to predict energy levels with loudness and the song genre. We determined that danceability is significantly different in 'Southern Europe and Portuguese heritage' and 'English-speaking and Nordic countries' compared to the global average. Valence is significantly different in English speaking and Nordic countries compared to the global average as well. Using more complex methods involving machine learning it may be feasible to utilize this information to generate models to predict popularity. Additionally, since music is art, there are so many factors, globally and socially, on a micro or a macro-scale, that could influence musical preference of different demographics in the same country. Such information was not available for this analysis.