Can audio attributes determine hit songs?

Team 105:
Crystal Mejia, Isaac Afolayan, Jimmy Serrano, Kunmi Sobowale, Nelly Ruiz, Wacarra Yeomans

Highlights

- Spotify and Billboard's top tracks share the same significant audio features

- The make-up of genres changed each year

- Our logistic regression classification model to measure the predictive power of our audio features was 73% accurate

Background

Music streaming platforms, such as Spotify, have allowed the music industry to move to digital and have changed the market prioritization. In this new music ecosystem, artists and labels are challenged to determine how to best invest their time, energy, and resources to create music that will reach a broad audience.

With this in mind, our team is interested in understanding patterns and features in popular music in recent years and aims to use data science techniques to better understand the makings of a popular song.

Data

We pulled songs information from both Spotify and Billboard using their respective APIs. The data consisted of audio, artist, and album features for hit and non-hit tracks from 2017-2020.
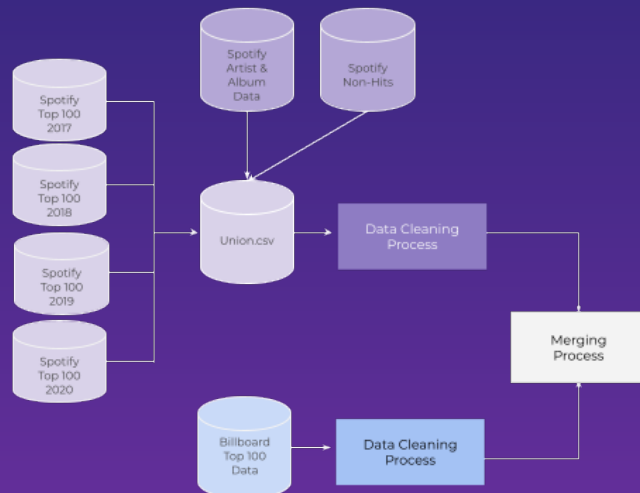
Model

From our bivariate analyses, we observed audio features that had noticeably different averages when comparing hits and non-hits. We created a logistic regression classification model to measure the predictive power of our audio features on determining if a song is a hit or not. To select the features to train our model, we performed varied T-tests to identify features with statistically significant differences in the averages. The accuracy, precision, and recall for our model were substantial. This analysis allowed us to conclude that songs can be identified as Non-Hit vs. Hit based on Spotify's few influential audio features. Furthermore, Spotify and Billboard share the same significant audio features.
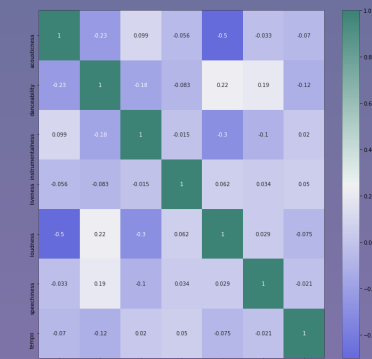
Exploratory Data Analysis

Fig. 1 Heatmap doesn't show strong positive correlation between attributes
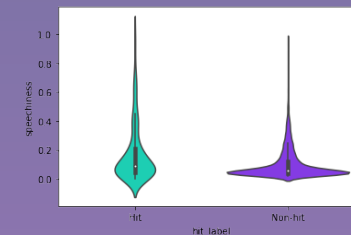

Fig. 2 Violin plot showing more speechiness in hit songs

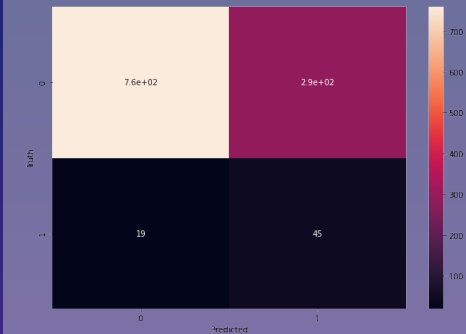### Logistic Regression Analysis
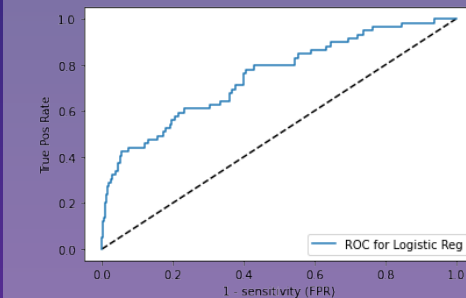

Fig. 3 Confusion Matrix within Spotify songs

Fig. 4 ROC for our model trained on Spotify to predict a hit on Billboard

Hit or non-hit songs can be determined by a few influential audio features:
Speechiness, danceability, liveness, & acoustiness.