Ethan Alqhuist
Andrew Doud
Will Hitchcock

Pitch - What is your big picture?

In our final project we intend to use multiple music hosting and scoring databases in order to make predictions and inferences about the success of different songs based on multiple sets of factors. Some of these factors might be the use of a different scoring system, another could be musical differences in musical style, genre or time period of creation.

Specific Aims - 3-5 specific questions and potential results

1. Could we predict the review rating from the review text?
2. Can we make any inferences on how popular song preferences are over time?
    a. By looking at how features of 'well rated' songs have evolved over time, likely by genre
3. Based on song attributes can we predict what common terms we would find in the review?
4. Can we predict a song's average rating from its Spotify attributes?

Timeline

In the first week:

- We will focus on tasks we can complete with the current data at hand.
    - Predicting review rating from the text
    - We will have derived appropriate feature from the text, tf-idf, sentiment etc
- We will also acquire the other necessary audio feature data by song from the spotify api.

In the second week:

- We will have finished acquiring the audio features by title and can begin answering our other questions.
- We will complete the first iteration of model testing for predicting the review rating from text.
- We will begin model selection and feature creation for trying to predict the most common terms in a reviews for a singular title based on song attributes
    - We will have queried 5 most frequent terms for each song.
    - We will find have performed feature selection on song attributes to find what is correlated with frequently used terms

In the third week:

- We will finalize our model which predicts song rating based upon review text.
- We will have selected an appropriate model and features to use for predicting common terms, and tune hyper parameters
- We will begin to investigate how the attributes of well rated songs have evolved over time.

- We will also begin our final question by fitting a regression of all attributes to song rating.

In the fourth week:

- We will finalize our models for all questions and begin visualizations

In the Fifth week:

- We will put our final visualizations and models into a dash application to create a interactive and fun visual reference to our models

Team Outline (who does what)

- Ethan Ahlquist
    - generating sentiment data from reviews
    - Data visualizations?
- Andrew Dowd
    - API stuff
    - Model training
- William Hitchcock
    - Data preprocessing
    - Model training

Final Deliverable (website/presentation/report/etc)

By the end of week 9, we aim to deliver a presentation on our findings. We look to document all our work as we go along researching our specific questions, and intend on summarizing this work in our presentation. We will create bite-sized chunks of info on what we did in this presentation so that it is easy to understand what we did and what we found. A more in-depth look at our work can be found on the GitHub repo we will create.

rapid-api

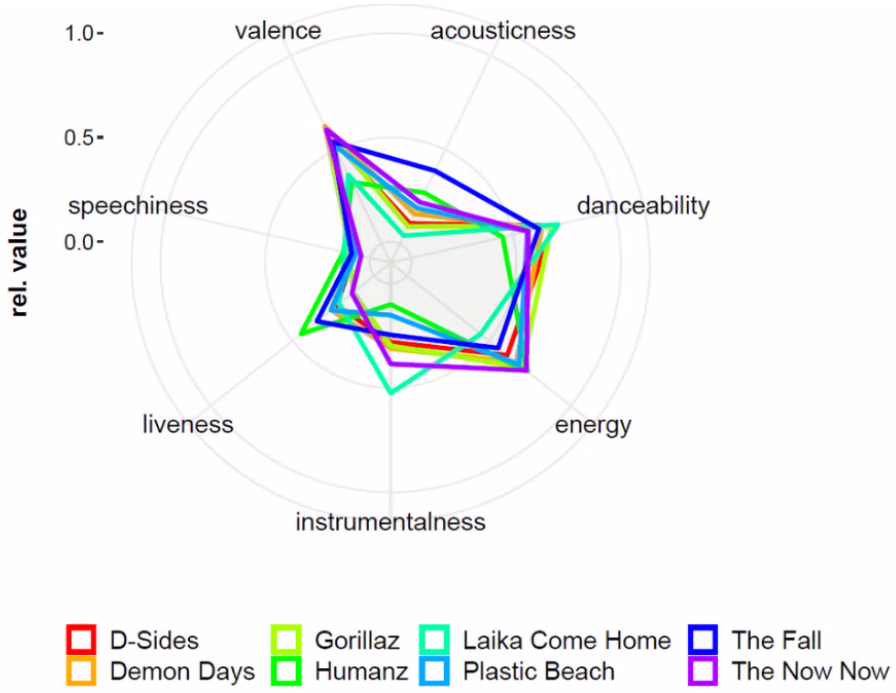
https://archive.ics.uci.edu/ml/datasets/FMA%3A+A+Dataset+For+Music+Analysis

Ideas: r/dataisbeautiful

Given the image of a house, predict the house price.

Spotify - upbeatness values. - train a model, based on a given song.
Tunebeat.com

- To make it flashy, we can display the happiness scale of popular songs.
- Given an input song how popular would it be?
    - Based on these given inputs.



includes genre info
https://www.kaggle.com/insiyeah/musicfeatures

genre is included in mp3 metadata.
https://archive.ics.uci.edu/ml/datasets/FMA%3A+A+Dataset+For+Music+Analysis#

- Might need to limit the dataset.

https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set

- Decision tree