Basic Info

The title of our project is SpotifyMBd.

Our team:

Steffen Dreesen	C10630953	sdreese@clemson.edu
Nick Macris	C13084809	nmacris@clemson.edu
Jacob Madsen	C14645719	jmmadse@clemson.edu

Project repo: https://github.com/steffendreesen/SpotifyMBd

Background and Motivation

Our project is motivated by our collective interest in music and film. Having interests in both of these areas, we naturally gravitated towards studying the intersection of these two topics: movie soundtracks. The soundtrack of a movie can have a significant impact on the viewer's reception of the film, which in turn determines its financial success. For this reason, we believe our study has potential for genuinely valuable discovery. We also knew of publicly available databases and subsequent APIs (IMDb and Spotify) which could provide us with a large and interesting dataset. For all the above reasons, we decided that film soundtracks would be a promising dataset to study.

Project Objectives

Overall, our main objective in completing this project is to study the relationship between film genre and soundtrack selection. More specifically, we want to create meaningful visualizations that allow the viewer to easily observe relationships and trends between movie genre and song selection. For example, how do the soundtracks of Comedies differ from Horror films? We want to create a set of visualizations that clearly displays the soundtrack differences in these two movie genres.

Finally, we would like to provide the user an interactive experience that allows them to study the relationship between film genre and soundtrack selection in real time. Our main objective in this regard is a film genre recommendation service. This service will take a user-submitted song and generate a visualization summarizing which film genres are most likely to have a soundtrack containing that song. Our goal for this objective is to provide the user with a fun mechanism to study the relationship between a film's genre and its soundtrack.

Data

Movie titles, genres and soundtracks were collected through the IMBd API. We used the Spotify API to collect attributes associated with the songs in each movie's soundtrack. These attributes include the following: track title, artist, valence, tempo, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, and liveness. A description of these attributes is provided below:

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).
acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features

	contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
instrumentalness	Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.
speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

The data was scraped using a Python script utilizing the IMBd and Spotify API. The data was then exported into a csv file for further processing. Our dataset is 1778KB large and contains all of the above information for 4871 films.

Data Processing

We do not expect to do any substantial data cleanup for our dataset. Much of the common data cleaning processes were implemented programmatically in our Python script.

We intend to make use of several derived quantities from our dataset. For each movie genre in our database, we will compute the average of each attribute available; for example, average tempo, average danceability, etc. Note that, for each film, the value for each attribute is an average of all the songs in that movie's soundtrack. We also intend to compute similarity scores between film genre soundtracks (described more below). These averages and similarity scores will be computed within the Excel file using Visual Basic.

Visualization Design

Our goal is to provide three visualizations:

1. Film Genre Soundtrack Similarity Matrix
2. Interactive Film Genre Soundtrack Attribute Analyzer
    a. User selects song attribute
3. Song to Film Genre Fitness Chart
    a. User provides song

In the first visualization, we intend to create a similarity matrix that will compare each film genre with each other based on a similarity function that we will define. The similarity function will be defined in the following manner: dissimilarity = (avg_valence_genre1 - avg_valence_genre2)^2 + (avg_tempo_genre_1 + avg_tempo_genre2)^2 + … and the smaller the value of dissimilarity the more closely related the soundtracks are between the film genres. The saturation of each cell

in the matrix will then be determined by this similarity function, with more similar genres' soundtracks resulting in a cell with greater saturation. The similarity matrix and use of saturation will allow viewers to quickly spatially and comparatively recognize which film genres typically make use of similar soundtracks and which do not.

The second visualization will be an interactive bar chart. The user will be provided a dropdown menu allowing them to select a song attribute: valence, tempo, energy, etc. The x-axis will be labeled with the film genre, and the y-axis will provide a scale for the selected attribute. The user will be able to switch between the provided song attributes and observe visually how the typical soundtrack for each film genre makes use of the selected song attribute. The difference in heights of each bar will allow the user to quickly discern whether film genre_1 typically contains more happy songs in its soundtrack than genre_2, for example.

The final visualization will be a bar chart similar to the second visualization described above. However, this visualization will accept a user-provided Spotify song and generate a bar chart based on which film genre's typical soundtrack would be most likely to contain the user-provided song. The x-axis will again be labeled with each film genre, and the y-axis will provide a scale for the fitness function that we will define in the following manner: fitness = (input_song_valence - genre_valence)^2 + (input_song_tempo - genre_tempo)^2 + ... and the smaller the value of fitness the more likely the song would appear in that film genre's typical soundtrack. This fitness function will provide a fitness value for each film genre which will then determine the height of each bar in the bar chart. As before, the difference in heights of each bar will allow the user to quickly discern whether "Back in Black" by AC/DC is more likely to appear in an Action film's typical soundtrack versus a Romance, for example.

Must-Have Features

- Film Genre Soundtrack Similarity Matrix
- Interactive Film Genre Soundtrack Attribute Analyzer
    - User selects song attribute
- Song to Film Genre Fitness Chart

    - User provides song

Optional Features

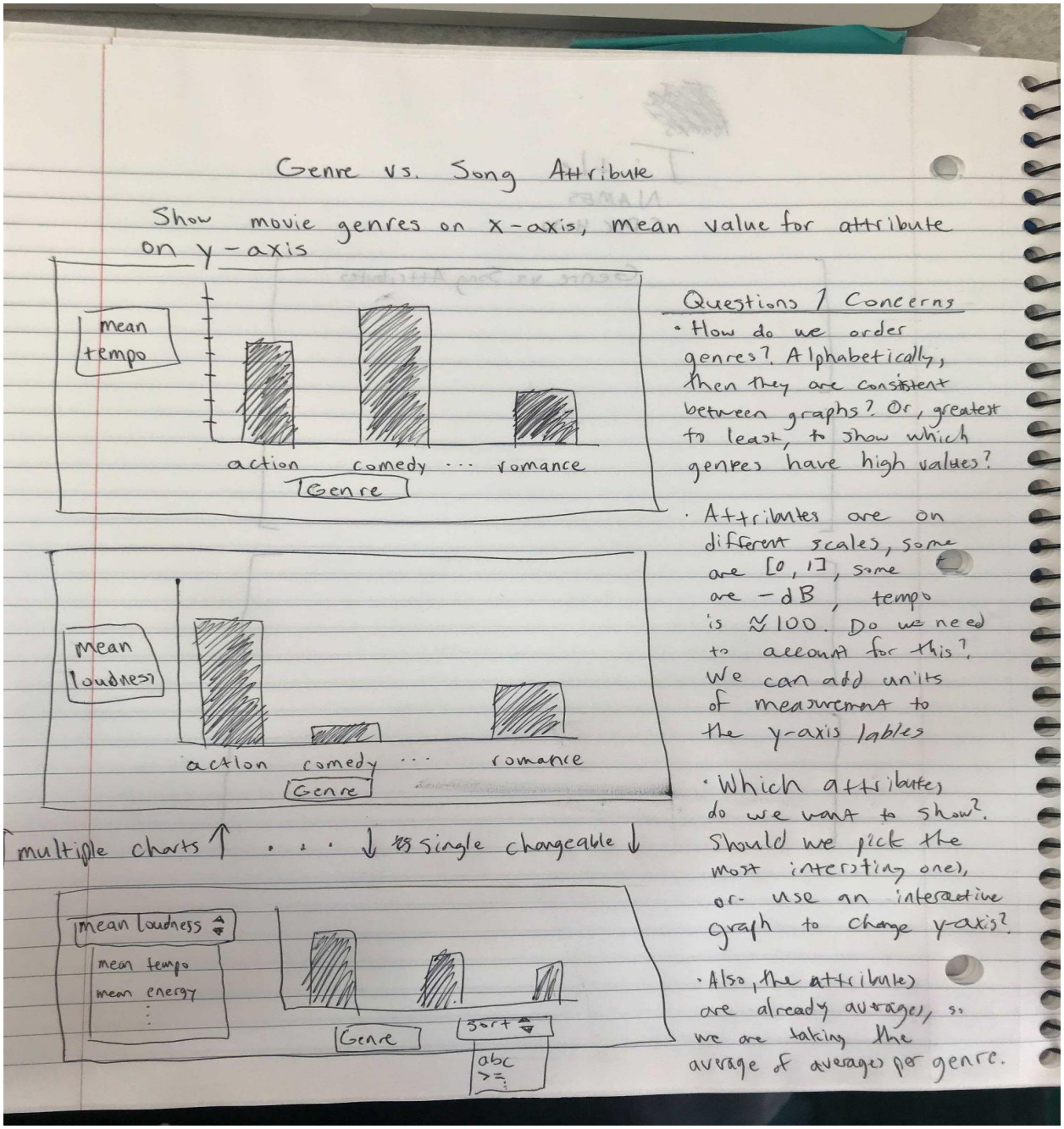
- Visualization showing top songs and top artists as well as their average attributes
- Bar chart matrix instead of Interactive Film Genre Soundtrack Attribute Analyzer

Project Schedule

Week	Deadline	Goals
Oct 17 - Oct 23		- Link our html website to the github pages - Do all statistical computations - Begin D3 work on Film Genre Fitness Chart
Oct 24 - Oct 30		- Begin D3 work on Film Genre Soundtrack Similarity Matrix - Begin D3 work on Interactive Film Genre Soundtrack Attribute Analyzer
Oct 31 - Nov 6		- Iterate and come up with good design decisions such as colors, layouts, and more. - Have all 3 visualizations functional and on 1 website
Nov 7 - Nov 13	Prototype due Nov 7	- Style the

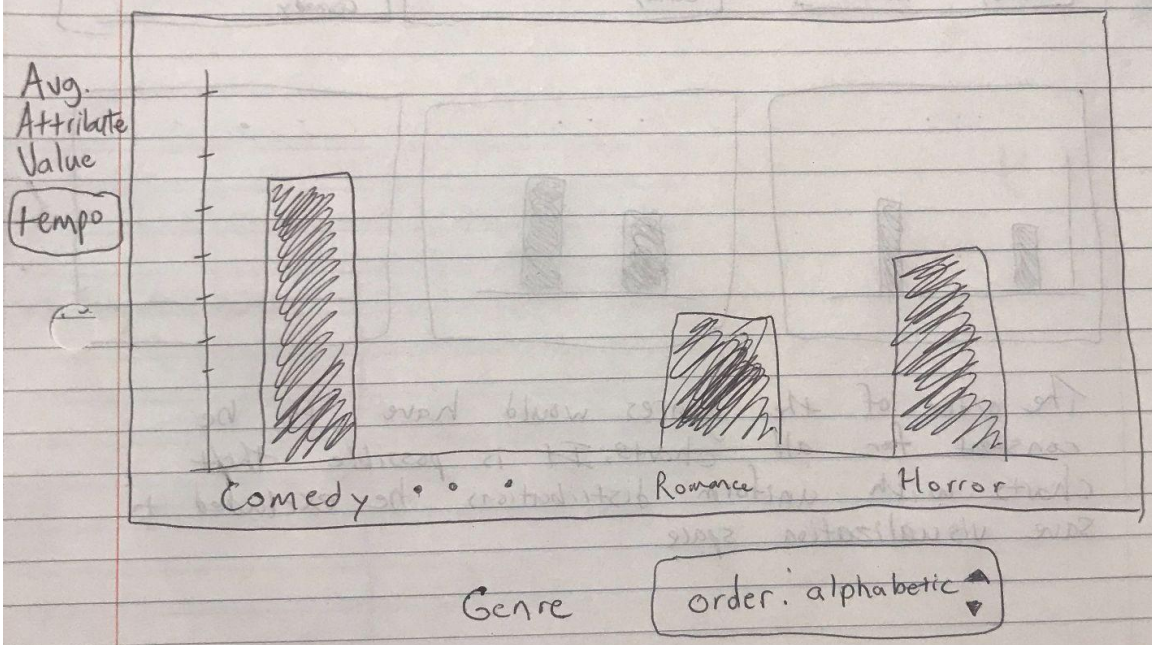
		visualizations and website to make it look nice
Nov 14 - Nov 20	Peer evaluation Nov 14	- Setup backend to support user-provided Spotify song for Film Genre Fitness Chart
Nov 21 - Nov 27		- Prepare presentation
Nov 28 - Dec 4	Oral Presentation Nov 30 / Dec 2	- Finalize all visualizations and site layout
Dec 5 - Dec 11	Peer assessment Dec 5	- Prepare for peer assessments.

Sketches


Genre vs. Song Attribute
Show movie genres on x-axis, mean value for attribute on y-axis
mean tempo
action
comedy
romance
Genre
mean loudness
action
comedy
romance
Genre
multiple charts
single changeable
mean loudness
mean tempo
mean energy
Genre
sort
abc
Questions / Concerns
• How do we order genres? Alphabetically, then they are consistent between graphs? Or, greatest to least, to show which genres have high values?
• Attributes are on different scales, some are [0, 1], some are -dB, tempo is ~100. Do we need to account for this? We can add units of measurement to the y-axis lables
• Which attributes do we want to show? Should we pick the most interesting one), or use an interactive graph to change y-axis?
• Also, the attributes are already averages, so we are taking the average of averages per genre.

Genre / Attribute Barchart

1) An interactive chart in which the user can set the song attribute on the y-axis, and change genre order on the x-axis



other options: increasing, decreasing,

All attributes would be selectable from the y-axis

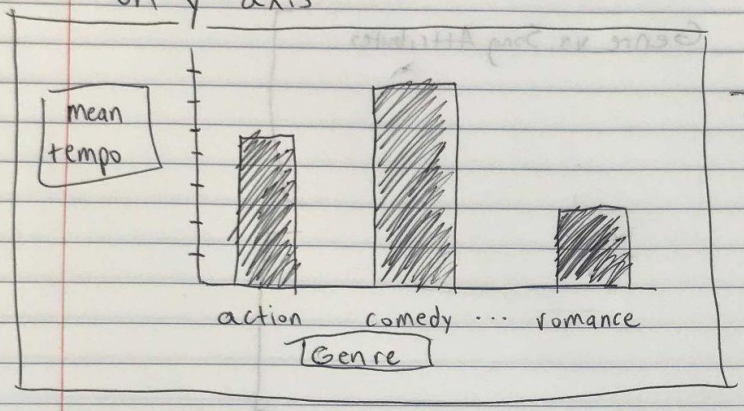
Genre / Attribute Barchart

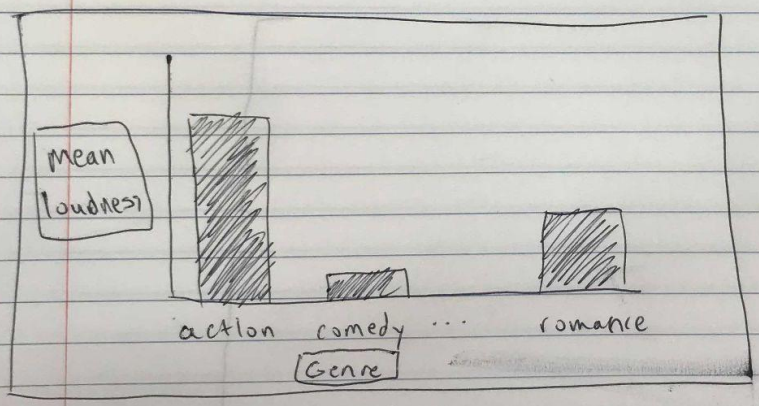
2) Have a set barchart for each attribute

The order of the genres would have to be constant for all charts. It is possible that charts with uniform distributions be excluded to save visualization space.

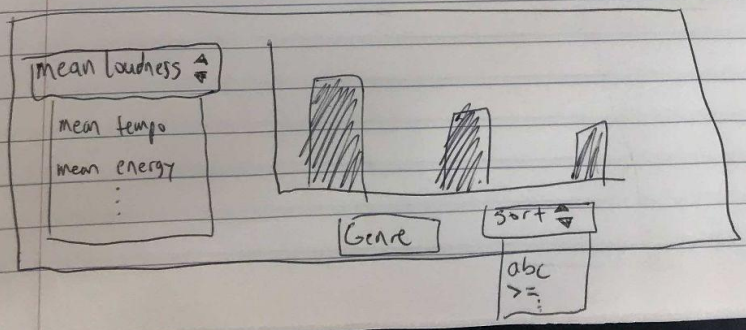
Genre vs. Song Attribute

Show movie genres on x-axis, mean value for attribute on y-axis





multiple charts ↑ . . . ↓ single changeable ↓



Questions / Concerns

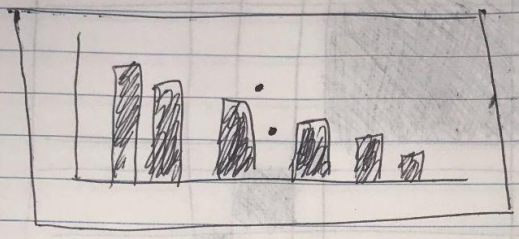
- How do we order genres? Alphabetically, then they are consistent between graphs? Or, greatest to least, to show which genres have high values?
- Attributes are on different scales, some are [0, 1], some are -dB, tempo is ≈100. Do we need to account for this? We can add units of measurement to the y-axis lables
- Which attributes do we want to show? Should we pick the most interesting ones, or use an interactive graph to change y-axis?
- Also, the attributes are already averages, so we are taking the average of averages per genre.

Sample Wesite layout

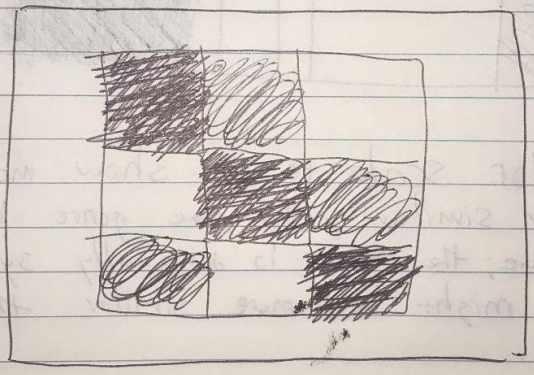
Title

Authors: Steffen, Nick, Jacob

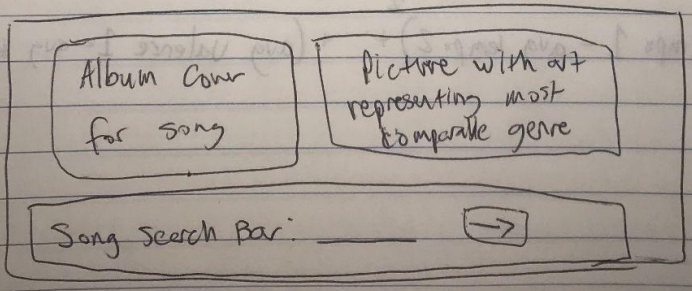
Genre vs Soundtrack Attributes

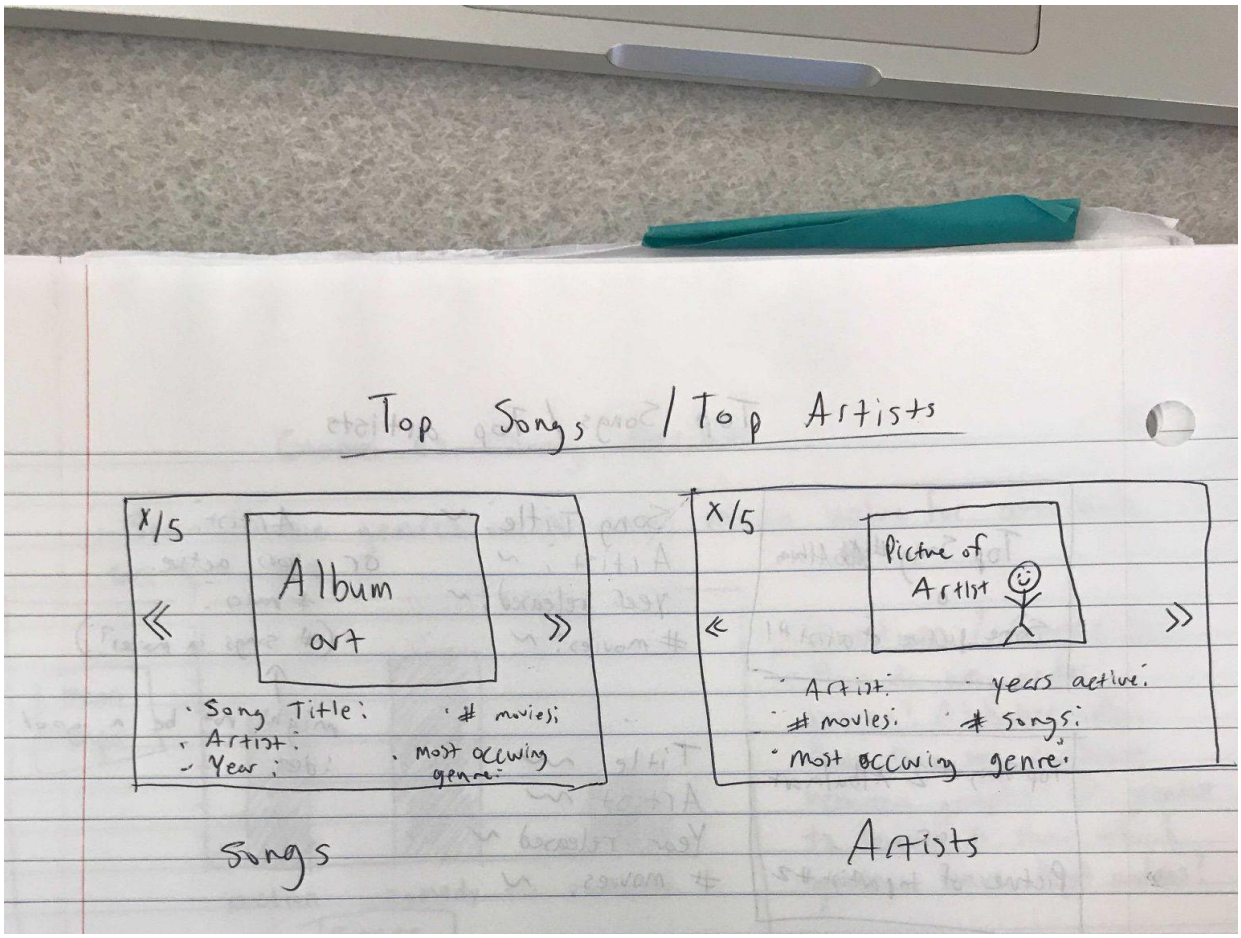


Similarity Matrix



Compare User Provided Song to Movie Genres




Top Songs / Top Artists
X/5
Album art
· Song Title:
· Artist:
· Year:
· # movies:
· most occuring genre:
Songs
X/5
Picture of Artist
· Artist:
· years active:
· # movies:
· # songs:
· most occuring genre:
Artists