LISTR

Initial Findings

Arranged by:

Grace Chen
Patrick C
Nicholas Drake
Olushola Durojaiye
Lena Lu

Office:
633 Clark St.
Evanston, IL 60208

+1 847 491-3741
project@listr.com
www.listr.com

Table of Contents

01 Overview

02 Project Plan

03 Closing

The Team

Grace Chen currently is working as a data science developer within a large insurance firm by creating various pricing optimization tools across multiple lines of business for internal stakeholders. She has led teams of software developers as well as spearheaded high priority projects to produce explicit results and immediate benefits to the organization. Prior to her current position, she has experience with various roles involving production support, model building, and data engineering.

Patrick C has over twenty years of experience ... and the rest of his biography had to be redacted.

Nicholas Drake is a Director of Data Solution Strategy responsible for developing and executing ESG data-driven strategies to solve Responsible Investing business problems and drive organizational growth. He has expertise in various areas of data analytics, including data warehousing, data mining, data visualization, and predictive modeling. Nicholas works closely with data scientists, engineers, and other stakeholders to identify and prioritize data-driven solutions that align with business goals.

Olushola Durojaiye is an Analytics Developer with several years of experience in system analysis, data modeling, database design, SQL programming, data analysis, and metrics reporting. He possesses extensive expertise in ETL processes and business reporting, utilizing the Microsoft Business Intelligence Stack and Tableau suite of products. Additionally, he demonstrates strong leadership, organizational, oral/written communication, interpersonal, analytical, and problem-solving skills. He excels in both independent and collaborative work environments

Lena Lu is a Senior Insight Analyst and has five years of experience working with data initiatives in retail and supply chain. She currently works with a large U.S. grocery retailer client on various data driven projects. She has led end-to-end data projects and has worked alongside data engineers, data scientists, analysts, and key business stakeholders on those projects. She is able to work cross functionally with different teams and audiences while being an individual contributor.

Executive Summary

Originally our scope was to create a music streaming platform to rival the likes of Spotify by offering AI personalization for playlists beyond what is available on the market currently. However, after scoping out the initial requirements needed to produce such a product, we discovered that it would require substantial overhead on a virtual machine to host and process the quantity of data needed. In our initial investigation, we were able to produce a minimum viable product to complement Spotify rather than compete.

We are proposing a scope change from a fully independent streaming service to a supplementary add-on application for Spotify users.

This will remove the overhead storage costs from our platform and utilize the existing Spotify ecosystem. It will also give Spotify competitive advantage for them to remain relevant in the market and compete against contenders like Amazon Music and YouTube Music who increased their subscriber counts by 25% and 50% respectively in 2021 while Spotify's market share steadily decreased by 100 basis points every year since 2019 where it was at 34% to 32%.[1] Our app will help Spotify transcend beyond a functional streaming service to one that is social and user driven.

The project is currently on track to deliver on August 13, 2023 with this new proposed scope.

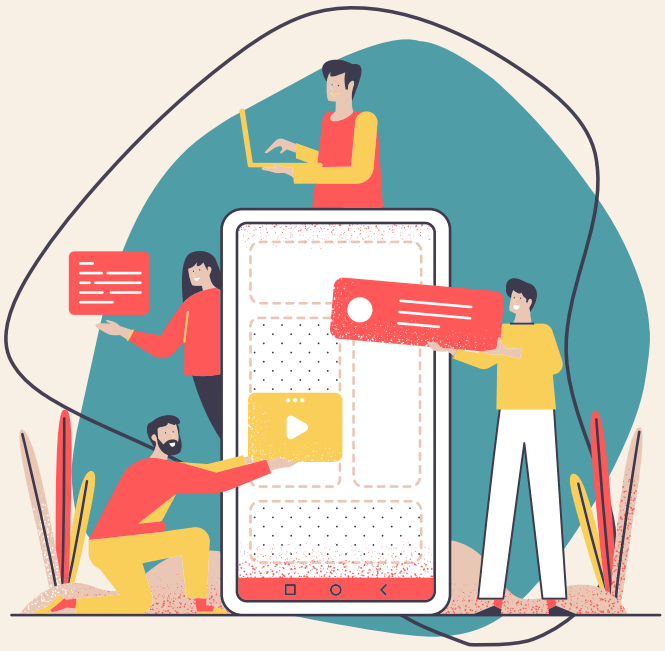
[1] "Music Streaming Market Share and Revenue Statistics: Details on the Biggest Music Streaming Services," SiriusXM Music for Business, January 27, 2023, https://sxmbusiness.com/music-streaming-market-share-and-revenue-statistics.

Scope

Our app will utilize the Spotify API's built-in functionality of song recommendations and supplement with additional recommendations generated from our model outputs.

We will generate recommendations based on similar song musicality and lyrical similarities, i.e., similar topics and moods. Our app will also have a dashboard view of each song which will include metrics such as song musicality, topic, emotion on top of a similar songs recommendation playlist. Users will be able to add recommended songs into a tentative playlist and then a custom playlist art will be generated from the songs in the list. Users will also be able to generate overlapping playlists with their friends by identifying the features contained within two playlists and outputting a mutual playlist of songs that both users would enjoy.

Users will also be able to connect their own Spotify accounts to the app to get a report of their listening habits similar to the currently existing Spotify Unwrapped report that Spotify produces at the end of the year. The Spotify Unwrapped functionality is well received every year, but it only gets produced once a year, with our app, users can see in real[1] time what their habits are and see additional features such as mood.



Objectives

01	Create an application for both web and mobile that will allow users to get reports on searched songs and user habits, add similar songs to a playlist, create mutual interest playlists with friends, and generate custom playlist album covers
02	Determine core capability and functionality for the application's first release
03	Finalize the UI/UX design for the application
04	Prototype a web and mobile application that will allow testers, decision-makers, and customers to create custom playlists, test model performance, showcase functionality, and drive user interface adaptations prior to launch
05	Obtain approval from the company CEO and board for a full-scale application deployment by the end of quarter 1 of 2024
06	Create a backlog of application improvements for future release

Data Sources

Spotify API[2]	**As we will be focusing on personalization within Spotify, we will be utilizing Spotify's API to access music data, user playlists, and recommendations. Spotify's API allows us to search for tracks, albums, artists, and playlists and pull back unique audio features such as danceability, energy, loudness, and valence. The API also allows Spotify users to interact directly with their account, in which we can programmatically create playlists based on our model's suggestion**
Genius API[3]	**We will be using the Genius API to pull lyrics from the Genius site as our source of truth for song lyrics. Genius has the largest collection of song lyrics in the world**

Data Issues

The Million Song Dataset[4] was initially the primary choice due to its number of tracks and robust data features. While this dataset initially provided great promise, it ended up being shelved due to data integrity issues. It was found that many of the fields which were of interest were corrupted. Fields which were meant to be Float values, normalized between 0 and 1, were converted to integer values and rounded down to 0 within the dataset, rendering the values unusable. This error was present in 17 of the 30 possible fields and due to the substantial impacts on dataset usability, it was decided to promote the Spotify dataset to the primary source.

MusixMatch[5] was originally chosen to be used alongside the Million Song Dataset, but upon further investigation, the lyrics are not full texts but rather a bag-of-words which is not usable for our intended purpose.

[2] "Spotify API," Web API | Spotify for Developers, accessed July 8, 2023, https://developer.spotify.com/documentation/web-api.
[3] "Genius API Documentation," Genius API, accessed July 12, 2023, https://docs.genius.com/.
[4] Million Song Dataset, accessed July 8, 2023, http://millionsongdataset.com/.
[5] "Build with Lyrics." Musixmatch Developer. Accessed July 8, 2023. https://developer.musixmatch.com/.

Models

Topic

Determining the main topic of the song through topic analysis of the lyrics

Emotion

Determining the primary emotion of the song through sentiment analysis on the lyrics

Activity

Determining best songs for common activities such as working out, dancing, meditation, etc. through their musical features

Topic Modeling

To extract the topics, we used the Lyric Genius API to retrieve the lyrics of specific songs. We approached the lyric extraction process by exploring two built-in functions of either specifying the song directly to retrieve the corresponding lyrics or identifying the artist and dictating how many songs the user would like to use from the artist's repertoire. For optimal results, we chose to exclude duplicated live versions and remixes of original songs.

After extracting the lyrics correctly using the API, some preprocessing steps were necessary to clean and normalize the textual data. These steps included the following:

- Remove punctuations, HTML tags, and special characters and digits
- Convert all text to lowercase
- Remove URLs
- Remove line breaks
- Replace more than one white space with only one white space
- Remove stop words (including newly added words such as 'ooh', 'yeah', 'hey', 'whoa')
- Apply lemmatization
- Tokenize the text

After preprocessing the lyrics, we created a word cloud to visually represent the tokens that were the most common throughout the song that may correlate to main topics throughout the song.

Word Cloud for "All Too Well" by Taylor Swift

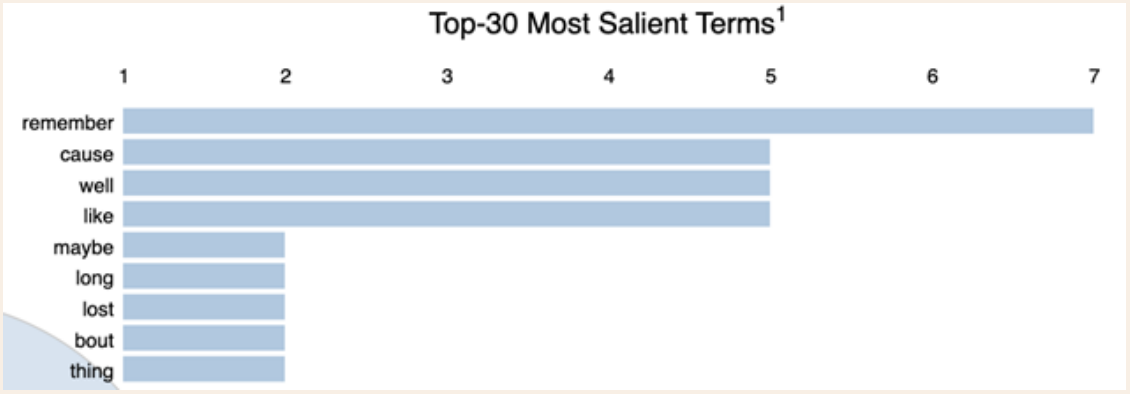
To identify the total count of each token present in the word cloud and its associated frequency throughout the song, we also output a table with the total count of each word.

Token	Frequency
remember	9
well	7
Cause	7
Oh	4
bout	3
thing	3
lost	3
maybe	3
long	3
forget	2
days	2
hair	2
old	2
gone	2
know	2
might	2
little	2
piece	2
red	2
back	2
Wind	2

We also calculated the mean term frequency-inverse document frequency (tf-idf) scores of every token in the song to identify the relevance of each token within the song with respect to the other tokens present. This metric also takes into account the frequency of each token within the song corpus.

	Mean TF-IDF
remember	0.18209
well	0.16613
like	0.14955
cause	0.14955
thing	0.07465
maybe	0.07465
lost	0.07465
long	0.07465
bout	0.07465
know	0.05269

For the topic modeling process, we utilized the Latent Dirichlet allocation (LDA) model to identify latent topics and salient terms within the song.



By comparing the results from the frequency count, mean tf-idf scores, and LDA model, it is clear that some tokens play a more significant role and contribute more meaning within the specified song relative to their counterparts.

Sentiment Analysis

To classify the emotion of songs, we sourced a pre-trained Natural Language Processing model on Hugging Face by Manuel Romero[6] that is built off of Google's T5 (Text-To-Text Transfer Transformer) model. This model is able to classify text into 6 emotions: sadness, joy, love, anger, fear, and surprise with the below test set metrics.

	Precision	Recall	f1-score	Support
anger	0.93	0.92	0.93	275
fear	0.91	0.87	0.89	224
joy	0.97	0.94	0.965	695
love	0.80	0.91	0.85	159
sadness	0.97	0.97	0.97	521
surprise	0.73	0.89	0.80	66
accuracy			0.93	2000
macro avg	0.89	0.92	0.90	2000
weighted avg	0.94	0.93	0.93	2000

Lyrics are generated via the Genius Lyrics API and then tokenized to the first 512 tokens (as required by BERT transformers. One of the blockers we found while working on this model was the storage of lyrics to put through the model. We did not have enough processing power to complete this in a timely manner; with the new scope of pulling information for each song in real time, we are able to connect the request to the Genius Lyrics API and run it through the model in that timeframe to overcome this limitation.

[6] Manuel Romero, "MRM8488/T5-Base-Finetuned-Emotion · Hugging Face," mrm8488/t5-base-finetuned-emotion · Hugging Face, accessed July 29, 2023, https://huggingface.co/mrm8488/t5-base-finetuned-emotion.

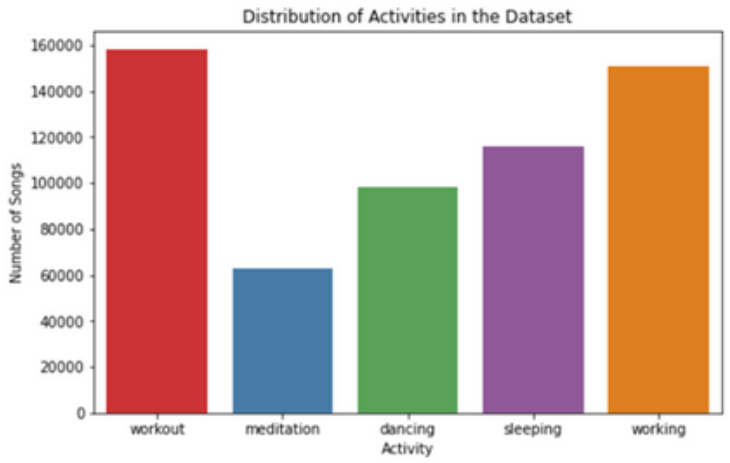
Activity Recommendations

For this section, we decided to define or generate playlists based on user activity. Some of the activities we chose were:

- Workout
- Meditation
- Dancing
- Sleeping
- Working

The thought process behind this was giving users the ability or flexibility to generate a playlist based on their current activity. For example, if a user wanted to generate a workout playlist without having to manually select or search for a song and then add to the manually created playlist, they could just select workout and the algorithm would select 10 random songs based on song tempo and recommend those as the workout playlist. Here is an example of the code output plus a bar chart showing the distribution of songs by activity.

```
Available activities: ['workout', 'meditation', 'dancing', 'sleeping', 'working']
Select an activity from the list above: meditation
Playlist Name: Meditation Playlist
Songs:
Ambon - ['Migz & Maya']
Bana Da Söyle - ['Ferdi Tayfur']
Shinig My Way - ['Jeff Stetson']
Shake it, ida, shake it - ['The Allen Brothers']
Las Playas de Invierno - ['El Barrio']
Make It Last Forever (with Jacci McGhee) - ['Keith Sweat', 'Jacci McGhee']
Skyline - ['Ak4:20']
Kings of the Party - ['Brownsville Station']
Broken And Spilled Out - ['Steve Green']
Täältä tulee yö - ['Ratsia']
```



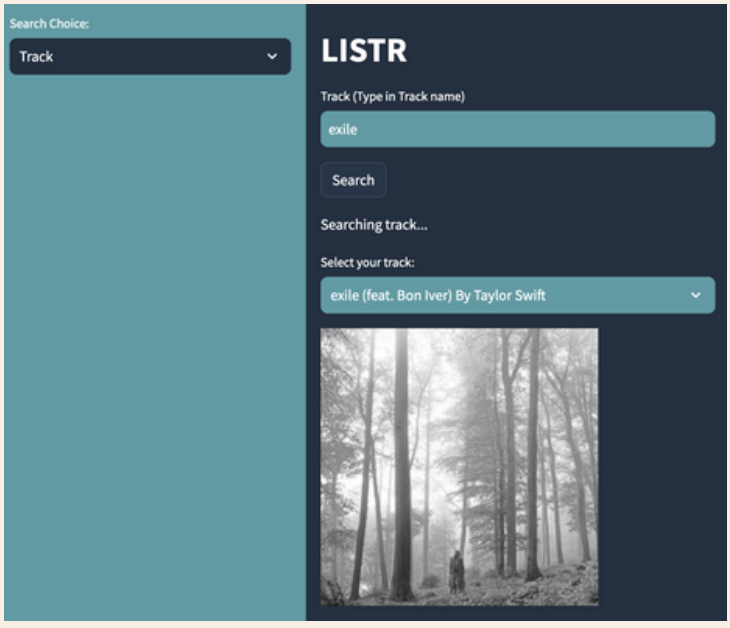
Proof of Concept

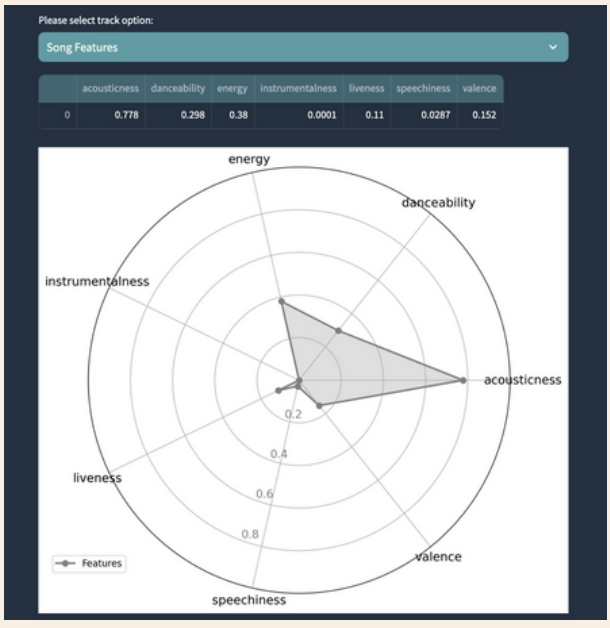
The Listr app is a user-friendly and interactive web application built using Streamlit and the Spotipy Python library to integrate with the Spotify API. Its main function is to allow users to search by tracks, artists, and albums and pull back music data through the Spotify platform.

The app provides a sidebar with options to select the type of search (Track, Artist, Album). Users can enter keywords related to their search in a text input field and click a button to trigger the search. Based on the selected search type and the entered keywords, the app fetches relevant data from the Spotify API and displays a list of search results, including tracks, artists, or albums.

Track Search

When the user selects the "Track" option, the app shows the details of the selected track, such as its name, artist, and various audio features like acousticness, danceability, energy, etc. A radar chart visualizes these audio features for improved comprehension by the user.





Recommendations

For a selected track, the app can recommend similar songs using the Spotify API. It presents the recommended tracks in a table with details like track name, artist, explicit content, duration, and popularity.

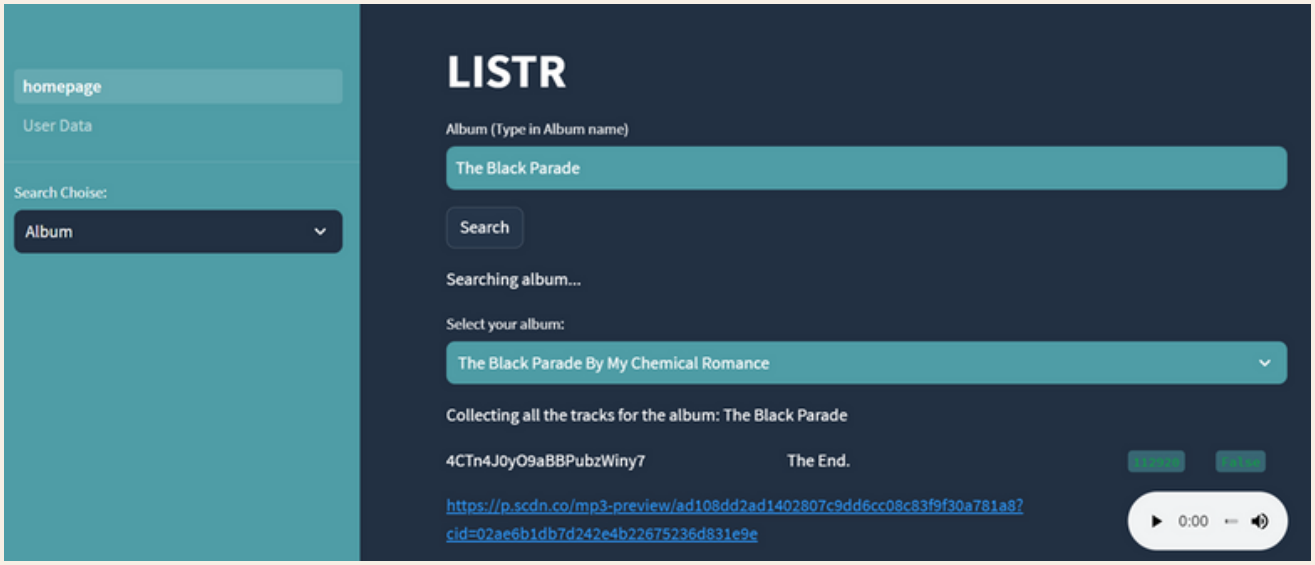
Please select track option:

Similar Songs Recommendation

	name	artist	explicit	duration_ms	popularity
0	Happier Than Ever	Billie Eilish	☑	298,899	88
1	You Let Me Down	Alessia Cara	☐	193,728	62
2	right where you left me - bonus track	Taylor Swift	☐	245,026	81
3	Helpless	Phillipa Soo	☐	249,770	71
4	Late Night Talking	Harry Styles	☐	177,954	89
5	Someone Like You	Adele	☐	285,240	48
6	Rewrite The Stars	Zac Efron	☐	217,440	76
7	Miss Movin' On	Fifth Harmony	☐	194,253	32
8	Before You	Benson Boone	☐	187,940	73
9	i hate love (feat. Steve Martin)	Kelly Clarkson	☐	213,402	59

Album Search

If the user chooses the "Album" option, the app displays all the tracks associated with the selected album. A table shows details like track ID, name, duration, explicit content, and provides a preview audio player for each track.



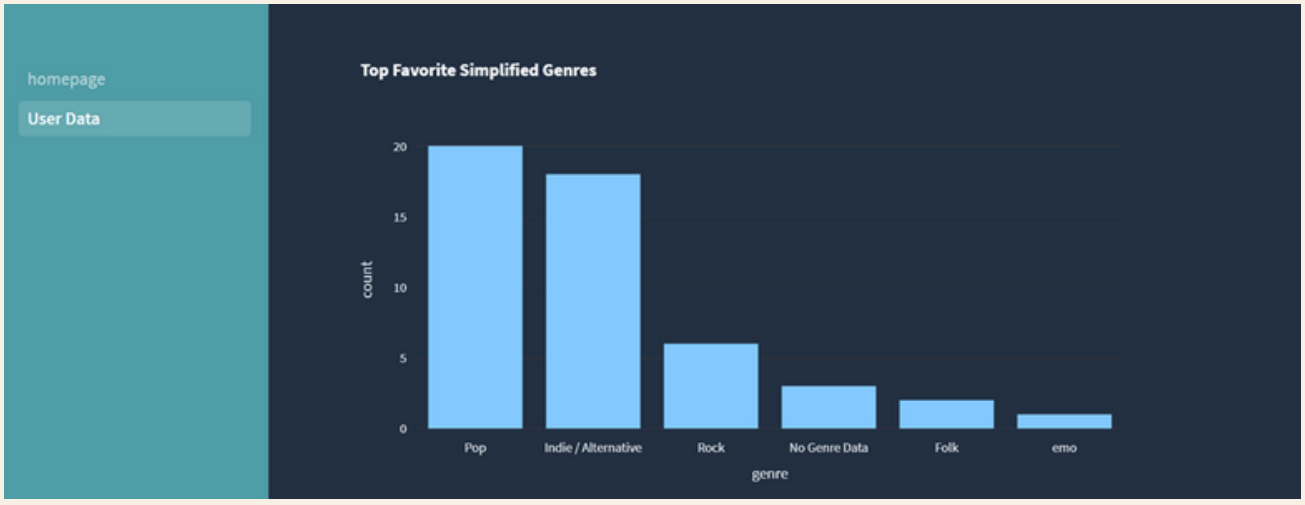
Artist Search

If the user chooses the "Artist" option, the app displays all tracks associated with the selected artist.

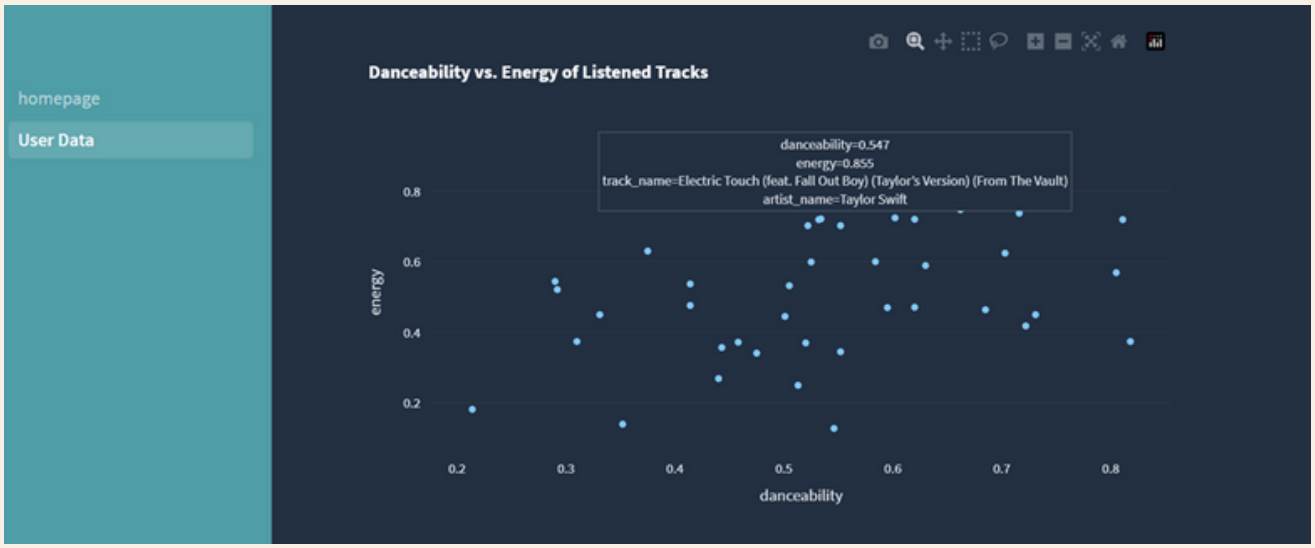
User Data

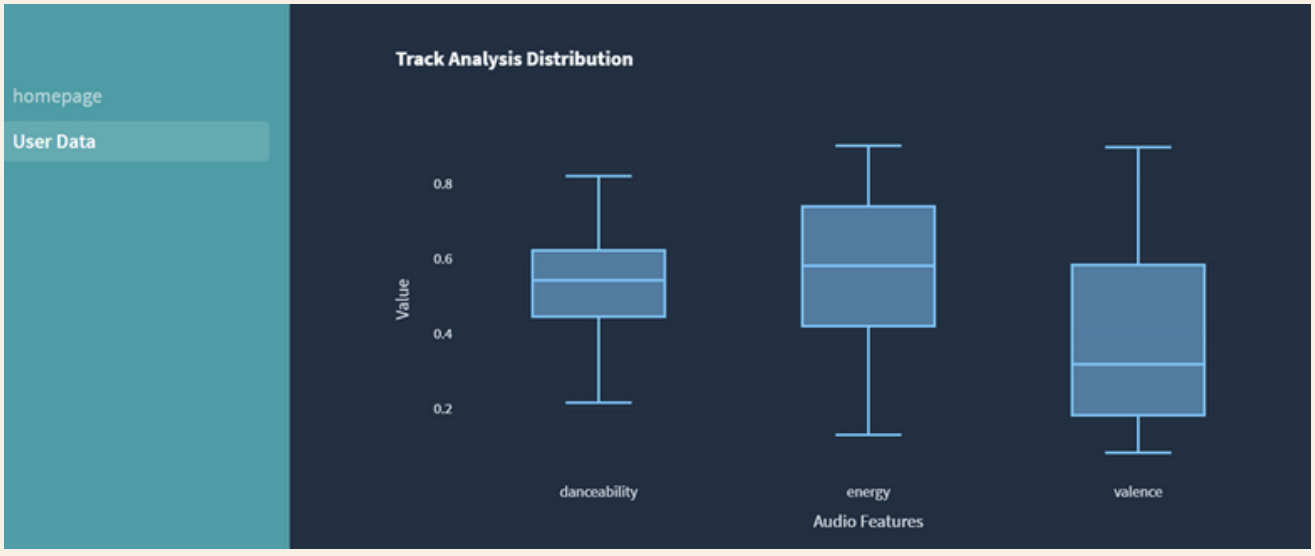
Apart from its search functionality for tracks, albums, and artists, the Listr app provides a personalized experience for music enthusiasts by enabling them to log in to their individual Spotify accounts and access their own listening data. With its intuitive interface and access to personal Spotify information, the app offers an immersive experience for music lovers to rediscover their cherished tunes. This app delivers a captivating journey into their unique musical preferences by exploring their top tracks and visualizing key discoveries.

	id	track_name	track_release	artist_name	popularity	danceability	ene
0	0woKDRMZzsFWeHppNQnGlS	Golden	2023-04-19	Barns Courtney	61	0.586	0
1	3EEr6l5PYelwkrvvKX7N0X	Shut Up Kiss Me	2016-09-02	Angel Olsen	59	0.375	0
2	1rZsn9K890Goymfl7YldgY	Mermaids	2023-04-21	Florence + The Machine	64	0.29	0.
3	2DuhJLdMNSskrd2KF8SRMr	Sad Disco (Pine Studios)	2023-02-14	flipturn	47	0.805	0.
4	3O8ZnG96gpHULyJyZ5dASk	Gold	2023-02-10	Ivan & Alyosha	27	0.62	0
5	4XiypEn8iKbeP1TbTqdfk5	Anti-Hero - Roosevelt Remix	2022-11-11	Taylor Swift	50	0.716	0.
6	39cc405lc5WqDnCvmUrP52	Mermaids	2023-04-21	Florence + The Machine	60	0.414	0.



Users also have the option to download their own data via CSV file to perform their own personal analysis.





Important notice: Please be aware that Listr is a third-party application integrated with Spotify. As part of our commitment to safeguarding user privacy and protecting personally identifiable information (PII), it is essential to clarify that Listr does not collect, store, or have access to user Spotify usernames or passwords. All authentication credentials are exclusively managed by Spotify, ensuring the confidentiality and security of user account information. Rest assured, user privacy is of utmost importance to us, and we strictly adhere to industry best practices to maintain the confidentiality of user data.

Timeline & Next Steps

Task Name	Week	Start Date	End Date
Improve Existing Capabilities	***7***	7/31/23	8/6/23
Add Additional Features	***7***	7/31/23	8/6/23
Playlist Comparison			
Song Topic & Emotion Analysis			
Custom Playlist Art			
User Interaction	***7***	7/31/23	8/6/23
Finalize Capabilities	***8***	8/7/23	8/13/23
Cloud Environment Configuration	***8***	8/7/23	8/13/23
Application Deployment	***8***	8/7/23	8/13/23
Post-Deployment Checks	***8***	8/7/23	8/13/23
Final Bug Fixes & Testing	***9***	8/14/23	8/20/23
User Experience Tuning	***9***	8/14/23	8/20/23
Submission Readiness	***9***	8/14/23	8/20/23
Presentation Content Preparation	***10***	8/21/23	8/27/23
Technical Details & Architecture Overview	***10***	8/21/23	8/27/23
Final Presentation Recording	***10***	8/21/23	8/27/23
Chatbot Functionality	***Future***		

Closing Remarks

Listr focuses on creating an AI-powered supplementary app for Spotify users, offering personalized music streaming experiences. Originally envisioned as a standalone music platform, we shifted our approach to complement Spotify's existing features due to resource constraints. The app will utilize the Spotify API for song recommendations and enhance them with AI-generated outputs based on musical and lyrical similarities. Users can create custom playlists with personalized album covers, generate mutual interest playlists with friends, and receive real-time reports on their listening habits. By enhancing Spotify's capabilities, Listr aims to provide a social and user-driven music streaming experience, empowering Spotify to remain competitive in the dynamic market.

References

"Build with Lyrics." Musixmatch Developer. Accessed July 8, 2023. https://developer.musixmatch.com/.

Curry, David. "Music Streaming App Revenue and Usage Statistics (2023)." Business of Apps, May 2, 2023. https://www.businessofapps.com/data/music-streaming-market/.

"Genius API Documentation." Genius API. Accessed July 29, 2023. https://docs.genius.com/.

Million Song Dataset. Accessed July 8, 2023. http://millionsongdataset.com/.

"Music Streaming Market Share and Revenue Statistics: Details on the Biggest Music Streaming Services." SiriusXM Music for Business, January 27, 2023. https://sxmbusiness.com/music-streaming-market-share-and-revenue-statistics/#:~:text=Spotify's%20su bscribers%20increased%20by%2020, increased%20by%20more%20than%2050%25.

Romero, Manuel. "MRM8488/T5-Base-Finetuned-Emotion · Hugging Face." mrm8488/t5-base-finetuned-emotion · Hugging Face. Accessed July 29, 2023. https://huggingface.co/mrm8488/t5-base-finetuned-emotion.

"Spotify API." Web API | Spotify for Developers. Accessed July 8, 2023. https://developer.spotify.com/documentation/web-api.