



Final Report

Arranged by:

Grace Chen

Patrick Cole

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

| | |
|-------------------|----|
| The Team | 01 |
| Executive Summary | 03 |

02 Full Report

| | |
|-----------------------|----|
| Project Review | 20 |
| Data Sources | 21 |
| Models | 22 |
| Application | 31 |
| Mobile Application | 41 |
| Future Enhancements | 49 |
| Financial Projections | 50 |

03 Closing

| | |
|-----------------|----|
| Closing Remarks | 52 |
| References | 53 |

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 Cole has over twenty years of experience within the United States Department of Defense in a variety of roles that have included; disarming bombs, presidential protection, leading a rapid prototyping laboratory, and serving as a data scientist and data engineer. He has led high-stakes projects that garner national level exposure where failure is not an option. He is comfortable translating technical subjects into digestible information and briefing findings to decision makers at all levels. He finds creative and innovative solutions by partnering with different skill sets to enhance project outcomes.



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.



LISTR

Executive Summary

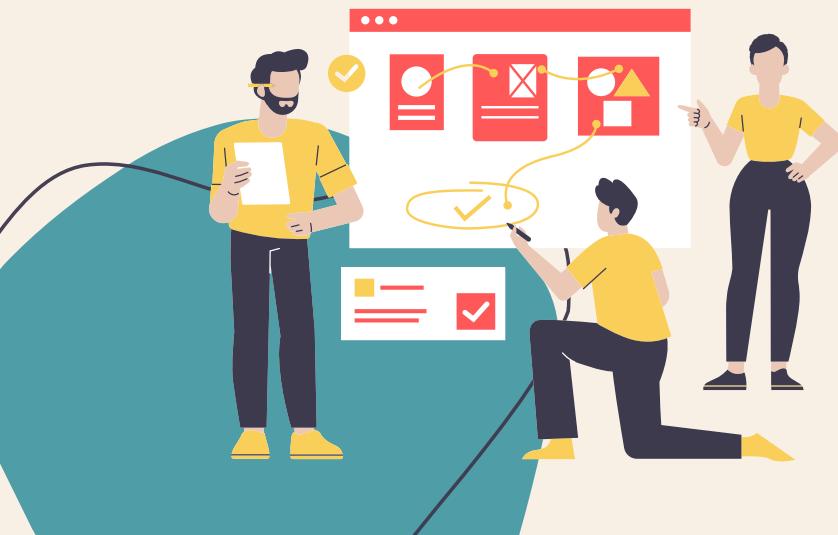


Overview

This final report outlines the journey of our AI music streaming platform project, now branded as **Listr**, The AI Music Streaming Add-On App. Originally envisioned as a standalone rival to Spotify, we pivoted our strategy to create a supplementary add-on app for Spotify users. Listr leverages AI personalization to offer unique playlists and a social user experience within Spotify's ecosystem. This report provides an overview of our objectives, achievements, data sources, challenges faced, financial impact, and future recommendations.

Project Objectives

- Develop Listr as a web and mobile application for users to receive music reports and create customized playlists.
- Finalize the UI/UX design for Listr to seamlessly integrate with the Spotify app.
- Prototype Listr for testing and user interface improvements.
- Obtain approval for live beta implementation of Listr within Spotify in Q4 2023 and full-scale deployment by Q1 2024.
- Establish a backlog of future Listr improvements for continuous enhancements.



Data Sources

Spotify API¹

Spotify's API was used to access music data, user playlists, and recommendations. We extracted data for unique audio features such as danceability, energy, loudness, and valence.

The API is also used to seamlessly connect to users' accounts to extract their user data for personalized reports.

Genius API²

The Genius API was used to pull lyrics from the Genius site for song lyrics. Genius has the largest collection of song lyrics in the world

Data Issues

In an earlier report, The Million Song Dataset³ was initially scoped out to be 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. Because of this, it was decided to promote the Spotify dataset to the primary source.

MusixMatch⁴ 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.

[1] "Spotify API," Web API | Spotify for Developers, accessed July 8, 2023, <https://developer.spotify.com/documentation/web-api/>.

[2] "Genius API Documentation," Genius API, accessed July 12, 2023, <https://docs.genius.com/>.

[3] Million Song Dataset, accessed July 8, 2023, <http://millionsongdataset.com/>.

[4] "Build with Lyrics." Musixmatch Developer. Accessed July 8, 2023. <https://developer.musixmatch.com/>.

Progress Review

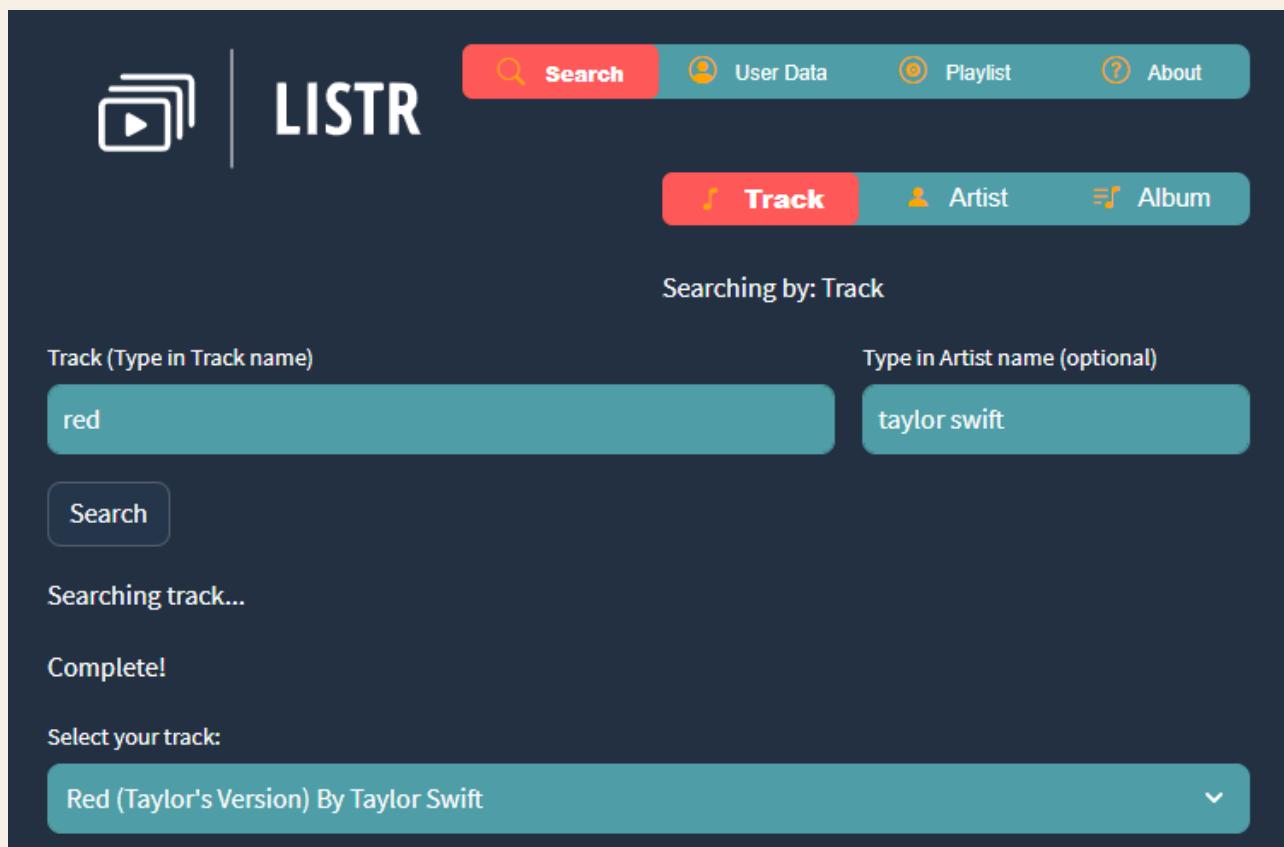
| | |
|----|---|
| 01 | <p>Created a web-based application, compatible with both desktop and mobile browsers</p> |
| 02 | <p>Created a search function that allows users to search by track, artist, and album</p> |
| 03 | <p>Populated the track search pages with: album art, track features visualizations, recommendations of similar songs, full lyrics, and a word cloud visual of the lyrics</p> |
| 04 | <p>Populated the artist search pages with their albums' names, release dates, and song count and their top tracks of all time</p> |
| 05 | <p>Populated the album search pages with each song's name, explicit rating, features, Spotify link, and duration along with visualizations giving an overview of the album's energy, danceability, and valence</p> |
| 06 | <p>Created a playlist comparison feature on the application that finds mutual songs from two playlists and produces visualizations based on those songs. Users will eventually be able to export the overlap between two playlists as a new playlist directly into Spotify</p> |

Application

The app provides 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, they are able to search by track and artist (optional). The app then shows the details of the selected track and lyrics along with visualizations characterizing the song.



Select your track:

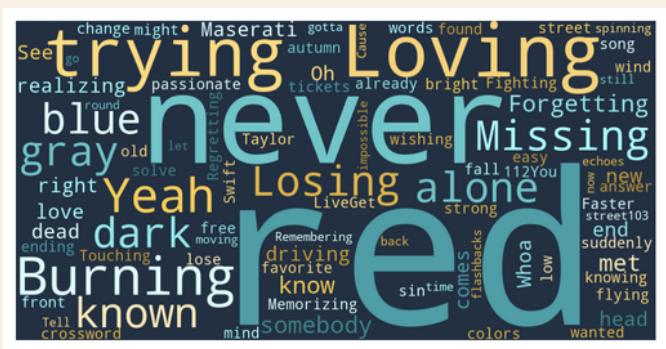
Red (Taylor's Version) By Taylor Swift

Please select track option:

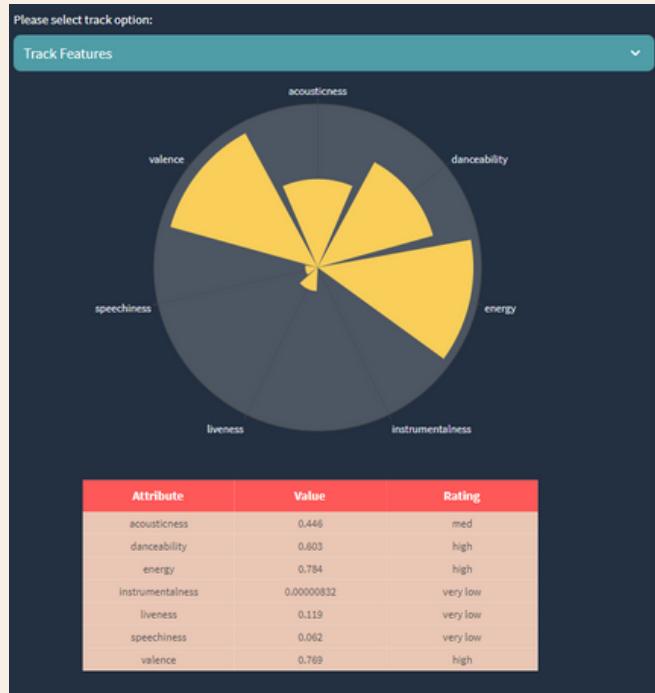
Track Features

| | acousticness | danceability | energy | instrumentalness | liveness | speechiness | valence |
|---|--------------|--------------|--------|------------------|----------|-------------|---------|
| 0 | 0.001 | 0.519 | 0.783 | 0 | 0.0776 | 0.0398 | 0.421 |

The initial results from track search



Word Cloud of the lyrics



Visual of the song features

Lyrics for 'Red' by Taylor Swift

Lyrics

Loving him is like driving a new Maserati down a dead-end street
 Faster than the wind, passionate as sin, ending so suddenly
 Loving him is like trying to change your mind once you're already flying through the free fall
 Like the colors in autumn, so bright just before they lose it all

Losing him was blue like I'd never known
 Missing him was dark gray, all alone
 Forgetting him was like trying to know somebody you've never met
 But loving him was red

(Red, red)
 (Red, red)
 Loving him was red
 (Red, red)
 (Red, red)

Touching him was like realizing all you ever wanted was right there in front of you

Full lyrics from Genius

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 Tracks Recommendation

| | name | artist | explicit | track duration | popularity |
|---|------------------------------|----------------------|-------------------------------------|----------------|------------|
| 0 | Same Boat | Lizzy McAlpine | <input type="checkbox"/> | 00:03:22 | 58 |
| 1 | I'm In Love With You | The 1975 | <input checked="" type="checkbox"/> | 00:04:22 | 74 |
| 2 | Best Friend Breakup | Lauren Spencer Smith | <input type="checkbox"/> | 00:02:21 | 66 |
| 3 | Summertime Sadness | Lana Del Rey | <input type="checkbox"/> | 00:04:24 | 65 |
| 4 | Liability | Lorde | <input type="checkbox"/> | 00:02:51 | 75 |
| 5 | don't come back | Tate McRae | <input type="checkbox"/> | 00:02:32 | 67 |
| 6 | this is me trying | Taylor Swift | <input type="checkbox"/> | 00:03:15 | 79 |
| 7 | Take A Chance With Me | NIKI | <input type="checkbox"/> | 00:05:03 | 59 |
| 8 | Music For a Sushi Restaurant | Harry Styles | <input type="checkbox"/> | 00:03:13 | 82 |
| 9 | Kids Are Born Stars | Laув | <input type="checkbox"/> | 00:03:07 | 64 |

Song Recommendations for "Red" by Taylor Swift



Artist Search

If the user chooses the "Artist" option, the app displays all the albums created by that artist along with the release date and song count. There is also an option to see that artist's top tracks.

Select your artist:

Taylor Swift

Start artist search...

Complete!

Select artist choice

Albums

| album type | album name | album release | track count |
|------------|-------------------------------------|---------------|-------------|
| album | Speak Now (Taylor's Version) | 2023-07-07 | 22 |
| album | Midnights (The Til Dawn Edition) | 2023-05-26 | 23 |
| album | Midnights (3am Edition) | 2022-10-22 | 20 |
| album | Midnights | 2022-10-21 | 13 |
| album | Red (Taylor's Version) | 2021-11-12 | 30 |
| album | Fearless (Taylor's Version) | 2021-04-09 | 26 |
| album | evermore (deluxe version) | 2021-01-07 | 17 |
| album | evermore | 2020-12-11 | 15 |
| album | folklore: the long pond studio sess | 2020-11-25 | 34 |
| album | folklore (deluxe version) | 2020-08-18 | 17 |

Select artist choice

Top Tracks

| | track name | track duration | danceability | energy | key | loudness | mode | speechiness | acou |
|---|-------------------|----------------|--------------|--------|-----|----------|------|-------------|------|
| 0 | Cruel Summer | 00:02:58 | 0.552 | 0.702 | 9 | -5.707 | 1 | 0.157 | |
| 1 | Anti-Hero | 00:03:20 | 0.637 | 0.643 | 4 | -6.571 | 1 | 0.0519 | |
| 2 | Blank Space | 00:03:51 | 0.753 | 0.678 | 5 | -5.421 | 1 | 0.0644 | |
| 3 | Style | 00:03:51 | 0.598 | 0.786 | 2 | -5.572 | 1 | 0.0383 | |
| 4 | august | 00:04:21 | 0.532 | 0.623 | 5 | -9.208 | 1 | 0.0331 | |
| 5 | cardigan | 00:03:59 | 0.613 | 0.581 | 0 | -8.588 | 0 | 0.0424 | |
| 6 | I Can See You (Ta | 00:04:33 | 0.694 | 0.764 | 6 | -4.893 | 1 | 0.0337 | |
| 7 | Karma | 00:03:24 | 0.642 | 0.62 | 8 | -7.091 | 1 | 0.0666 | |
| 8 | Don't Blame Me | 00:03:56 | 0.615 | 0.534 | 9 | -6.719 | 0 | 0.0386 | |
| 9 | Lover | 00:03:41 | 0.359 | 0.543 | 7 | -7.582 | 1 | 0.0919 | |

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.

Select your album:

folklore By Taylor Swift

Collecting all the tracks for the album: folklore

| | track name | explicit | track duration | danceability | energy | key | loudness | mode | spee |
|---|-------------------------|--------------------------|----------------|--------------|--------|-----|----------|------|------|
| 0 | the 1 | <input type="checkbox"/> | 00:03:30 | 0.765 | 0.348 | 0 | -6.963 | 1 | |
| 1 | cardigan | <input type="checkbox"/> | 00:03:59 | 0.613 | 0.581 | 0 | -8.588 | 0 | |
| 2 | the last great american | <input type="checkbox"/> | 00:03:50 | 0.684 | 0.661 | 7 | -8.507 | 1 | |
| 3 | exile (feat. Bon Iver) | <input type="checkbox"/> | 00:04:45 | 0.298 | 0.38 | 6 | -8.426 | 1 | |
| 4 | my tears ricochet | <input type="checkbox"/> | 00:04:15 | 0.456 | 0.263 | 0 | -10.63 | 1 | |
| 5 | mirrorball | <input type="checkbox"/> | 00:03:28 | 0.553 | 0.413 | 2 | -10.037 | 1 | |
| 6 | seven | <input type="checkbox"/> | 00:03:28 | 0.596 | 0.488 | 4 | -10.769 | 1 | |
| 7 | august | <input type="checkbox"/> | 00:04:21 | 0.532 | 0.623 | 5 | -9.208 | 1 | |
| 8 | this is me trying | <input type="checkbox"/> | 00:03:15 | 0.515 | 0.476 | 9 | -9.843 | 1 | |
| 9 | illicit affairs | <input type="checkbox"/> | 00:03:10 | 0.559 | 0.31 | 10 | -10.445 | 1 | |



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.

Although this feature is not live in the deployed app, there is a demo sample to showcase the functionality.

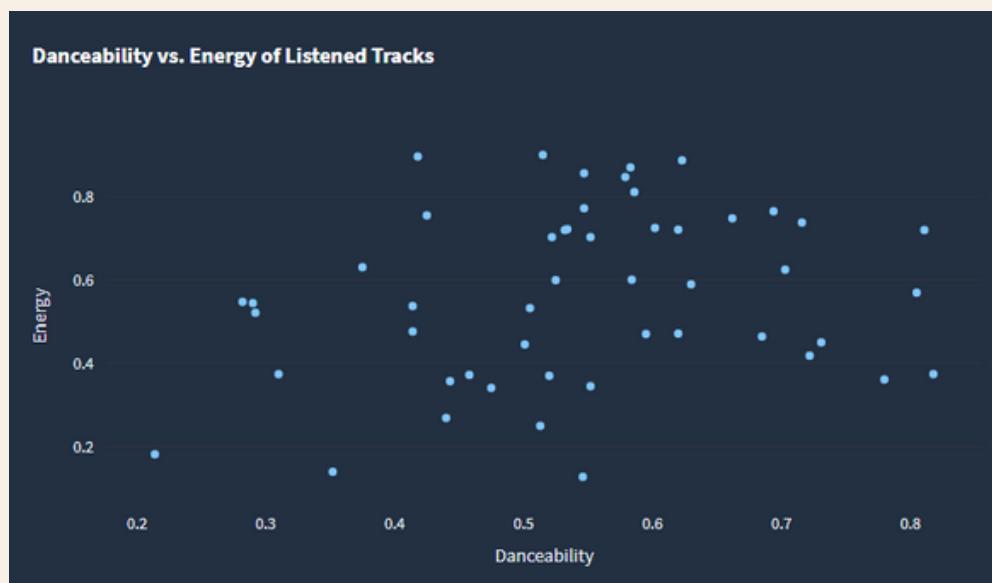
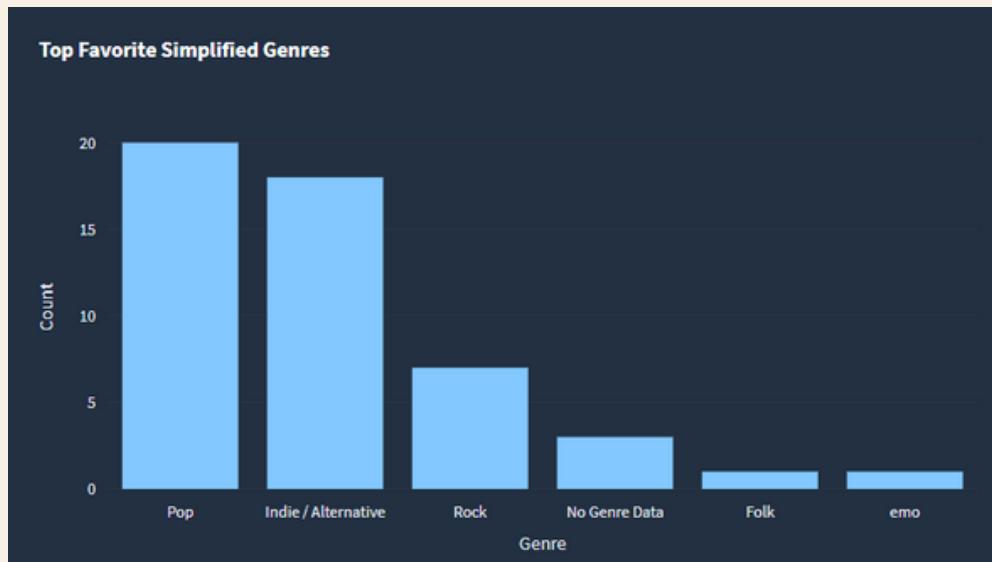
Your Spotify Listening Data

This feature is on its way and will be available shortly! In the meantime, feel free to explore the dashboard using the sample data provided by our team. Stay tuned for updates!

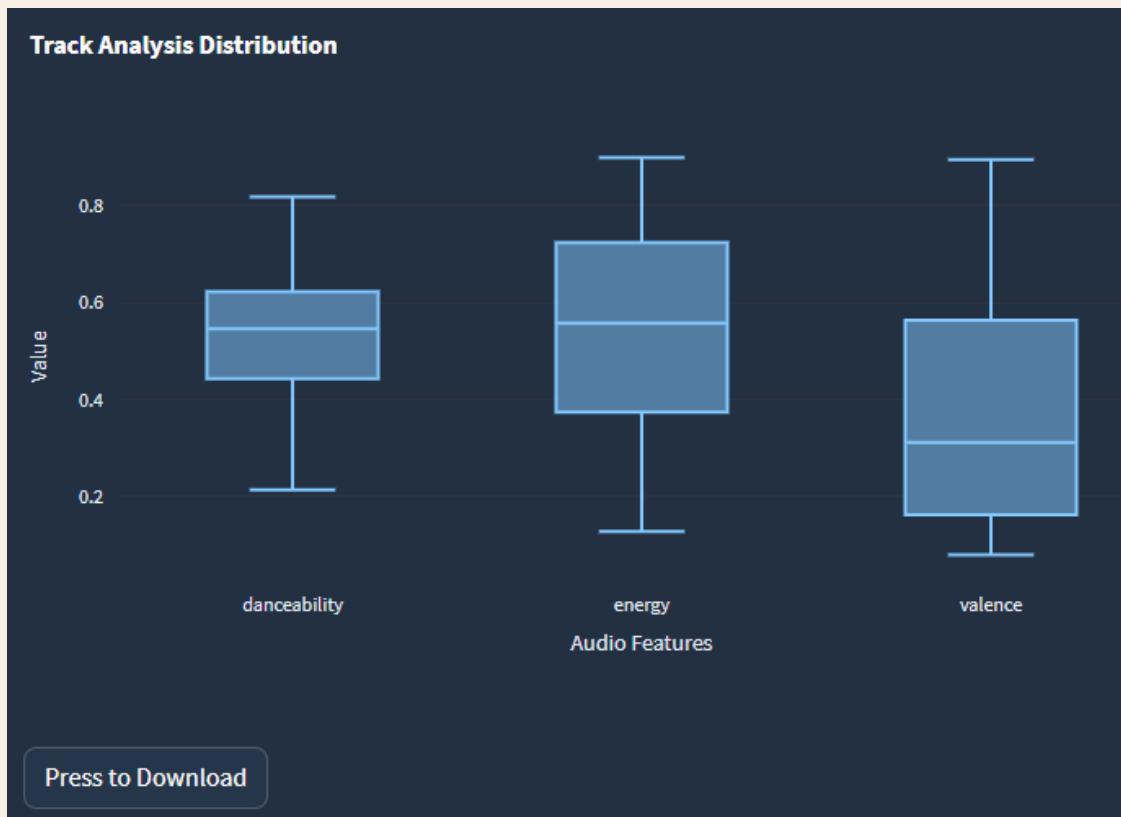
[Login to Spotify](#)

| | track name | track duration | track release | artist name | genre |
|---|--|----------------|---------------|------------------------|---------|
| 0 | Golden | 00:02:50 | 2023-04-19 | Barns Courtney | Rock |
| 1 | Shut Up Kiss Me | 00:03:22 | 2016-09-02 | Angel Olsen | Indie / |
| 2 | Mermaids | 00:04:35 | 2023-04-21 | Florence + The Machine | Pop |
| 3 | Sad Disco (Pine Studios) | 00:03:47 | 2023-02-14 | flipturn | Indie / |
| 4 | Gold | 00:03:57 | 2023-02-10 | Ivan & Alyosha | Pop |
| 5 | Anti-Hero - Roosevelt Remix | 00:04:59 | 2022-11-11 | Taylor Swift | Pop |
| 6 | Mermaids | 00:04:35 | 2023-04-21 | Florence + The Machine | Pop |
| 7 | I Can See You (Taylor's Version) (From The | 00:04:33 | 2023-07-07 | Taylor Swift | Pop |
| 8 | Miss Americana & The Heartbreak Prince | 00:03:54 | 2019-08-23 | Taylor Swift | Pop |
| 9 | Some things Cosmic | 00:02:54 | 2011-04-11 | Angel Olsen | Indie / |

Users are able to see visualizations depicting their favorite genres and the danceability and energy spread of their songs.



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.

Playlist Comparison

This functionality is designed for friends to create playlists that would be enjoyed by both. Just enter the link of two **public** playlists to get a new playlist built from the originals.

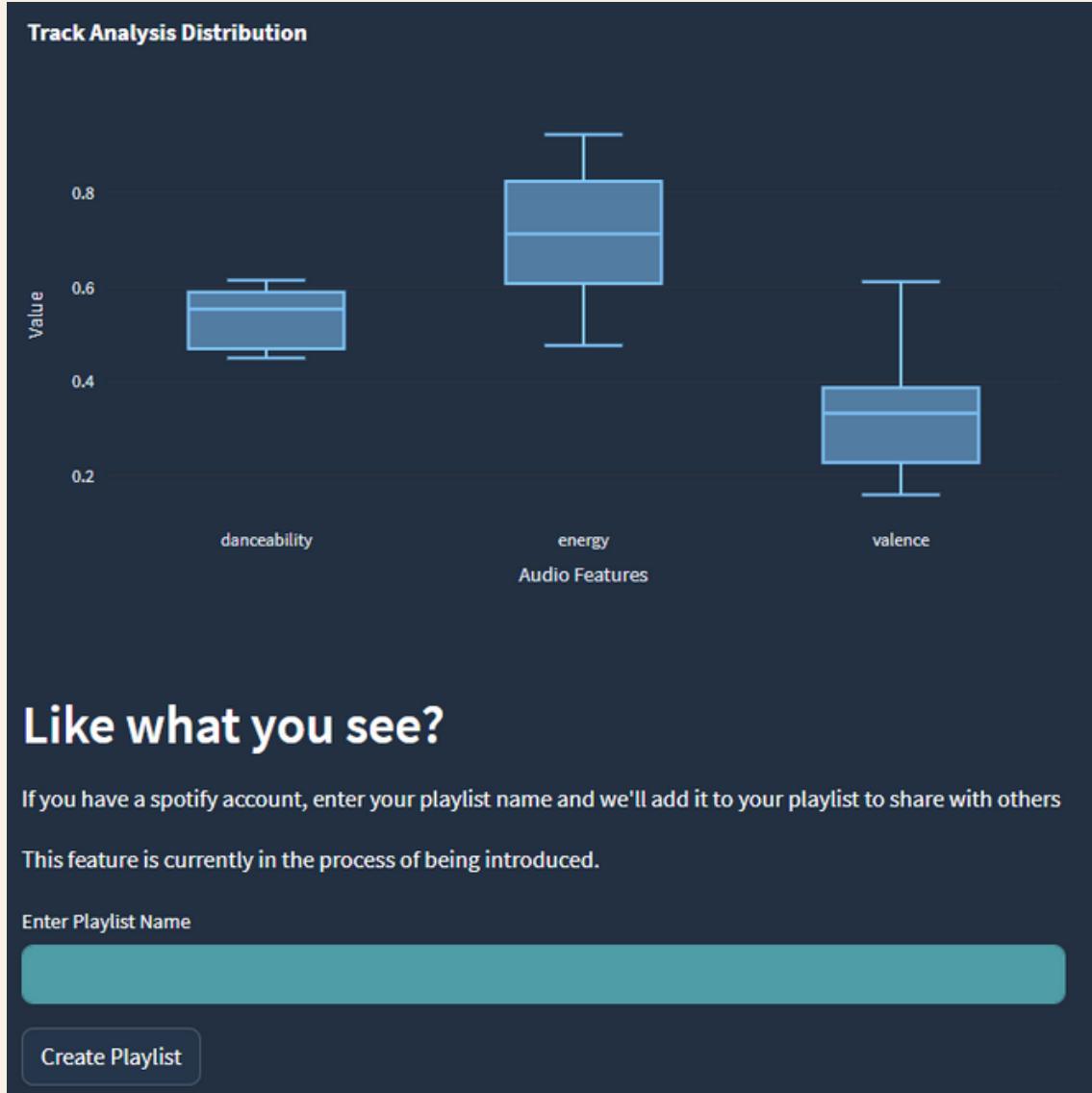


The screenshot shows the LSTR app interface with the following elements:

- Header:** LSTR logo with three play button icons, Search icon, User Data icon, and a red Playlist icon.
- Title:** Playlist Comparison
- Text:** "Let's compare two Spotify playlist of your choosing. Please make sure that they are public playlist."
- Instructions:** "In Spotify, go to the playlist and click the three '...' >> 'Share' >> 'Copy Link to Playlist'"
- Text Input:** "Enter Playlist 1 URL or Playlist ID" followed by a text input field containing "https://open.spotify.com/playlist/3XKYgDoQlczfW7WBcaluYY?si=3c41039f191845d6"
- Text Input:** "Enter Playlist 2 URL or Playlist ID" followed by a text input field containing "5Qnr7ct4jZ4ad1yDRuM1to?si=f4534b5b51a24f92"
- Button:** "Analyze Playlist" button.
- Table:** A table showing the results of the comparison, with columns: track_name, artist, and album. The data is as follows:

| track_name | artist | album |
|-----------------------------------|-----------------------|---|
| Bulletproof Heart | My Chemical Romance | Danger Days: The True Lives of the Fabulous Killjoys |
| Supernatural | Barns Courtney | Supernatural |
| State Of Grace (Taylor's Version) | Taylor Swift | Red (Taylor's Version) |
| Today | The Smashing Pumpkins | (Rotten Apples) The Smashing Pumpkins Greatest Hit |
| Graduate | Third Eye Blind | Third Eye Blind |
| Bullet With Butterfly Wings - Rem | The Smashing Pumpkins | Mellon Collie And The Infinite Sadness (Deluxe Edition) |
| Hash Pipe | Weezer | Weezer |
| 96 Quite Bitter Beings | CKY | Camp Kill Yourself, Vol.1 |
| Outshined - Remastered | Soundgarden | Badmotorfinger (25th Anniversary Remaster) |

Visualizations describing the new playlist are displayed. In future versions, users will also be able to automatically add this playlist onto Spotify to share with others.



Roadblocks

There were some experiments that were either abandoned due to infeasibility or were not ready for the deployment schedule of V1.0

User Data

Originally, the intent was to build a platform that would allow users to log into their own Spotify accounts and build reports using their data. However, that functionality had to be temporarily removed due to API limitations. The Spotify Developer platform only allows 25 registered users for free hobby projects. In a future version, we will obtain business approval to use their API.

Song Emotion Classification

1

There is a fully functional emotion classification model that ingests the lyrics and classifies them to one of six emotions: sadness, joy, love, anger, fear, and surprise. Unfortunately this functionality had to be disabled for the deployment due to its usage of PyTorch which significantly taxed the server and slowed down the application significantly.

User beta testing also found that this classification model was not good at classifying certain genres such as rap. It classified almost every rap song as anger which could be problematic to deploy, so it was disabled for now.

In addition, the classification model is catered towards western cultures and fails to take other cultures into account. When song lyrics have various languages and cultures intertwined, the results are not as accurate as expected. From the testing, it is evident that there exists bias and this is another future step to implement to become more inclusive in the long run.

Future Enhancements

Chatbot

Allows app users to ask a chatbot for recommendations based on similar songs, artists, moods, or activities

Song Recs

The current recommendation engine is solely based on Spotify's API. There are additional internal models in the works that will enhance those recommendations

Clusters

There will be clustering functionalities that will allow users to compare albums, artists, and more

User Data

Users will be able to seamlessly log into their Spotify account and extract their user data, generate user reports, and get recommendations based on their activity



Financial Projections

Listr is projected to build a 500K user base in the first year based on market research conducted. The cost per subscription will be \$1.99 monthly, so this will yield \$12M in revenue, but still -\$13M in profits due to the initial start up costs.

There was an initial investment of \$10M and an additional \$80M of funding was secured which will help over the first year's worth of overhead costs.

At 1M users, the application will break even and begin turning a profit.

Year 1

\$12M

*500K Users
Initial User Base*

Year 2

\$48M

*2M Users
Marketing & Product Enhancement Efforts*

Year 3

\$120M

*5M Users
International Expansion*



LISTR

Full Report

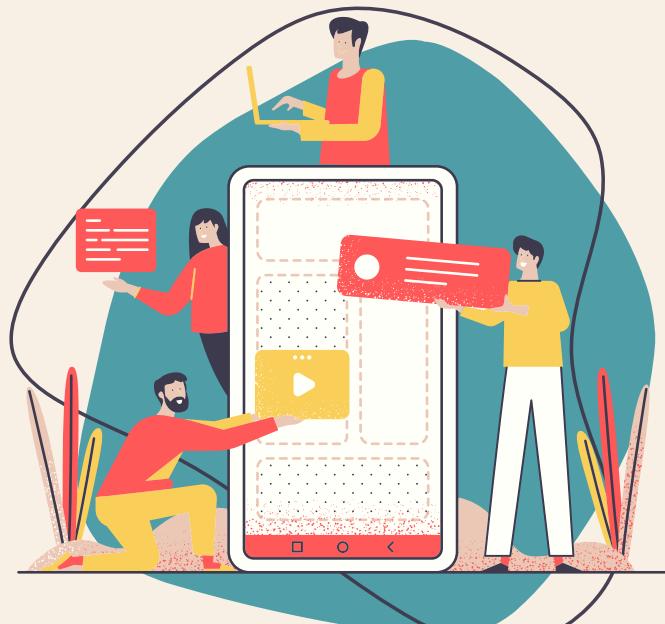


Project Review

In the last update on July 30, 2023, we proposed a scope shift from the original standalone chatbot recommender app to a more dashboard oriented app that serves as an add-on to the pre-existing Spotify ecosystem.

The app has been deployed as of August 13, 2023 with the below functionalities:

- Track Search which returns the following:
 - Album cover
 - Track feature visualizations
 - Full lyrics
 - Wordcloud of the lyrics
 - Most frequent words
- Artist Search which returns the following:
 - List of all their albums
 - Release dates of albums
 - Number of tracks in each album
 - Top Tracks
- Album Search which returns the following:
 - Songs' names, ratings, features, Spotify link, and duration
 - Visualizations describing the album's energy, danceability, and valence
- Playlist Merger which finds the commonality between two playlists and creates a subsequent new playlist of the overlaps



Data Sources

Spotify API¹

Spotify's API was used to access music data, user playlists, and recommendations. We extracted data for unique audio features such as danceability, energy, loudness, and valence.

The API is also used to seamlessly connect to users' accounts to extract their user data for personalized reports.

Genius API²

The Genius API was used to pull lyrics from the Genius site for song lyrics. Genius has the largest collection of song lyrics in the world

Data Issues

In an earlier report, The Million Song Dataset³ was initially scoped out to be 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.

MusixMatch⁴ 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.

[1] "Spotify API," Web API | Spotify for Developers, accessed July 8, 2023, <https://developer.spotify.com/documentation/web-api/>.

[2] "Genius API Documentation," Genius API, accessed July 12, 2023, <https://docs.genius.com/>.

[3] Million Song Dataset, accessed July 8, 2023, <http://millionsongdataset.com/>.

[4] "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

Playlist

Determining similarities between two playlists and generating a new playlist



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

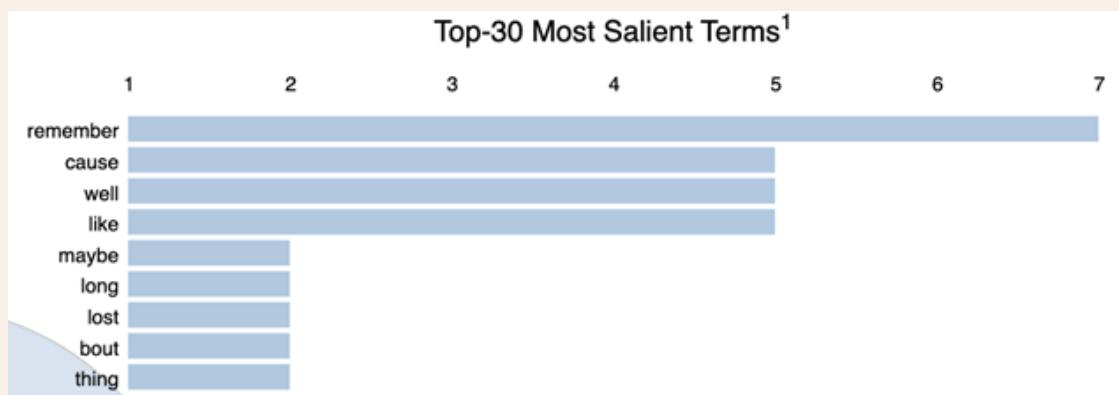
To identify the total count of each token present in the word cloud and its associated frequency throughout the song, we also output a table with the total count of each word.

| 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.

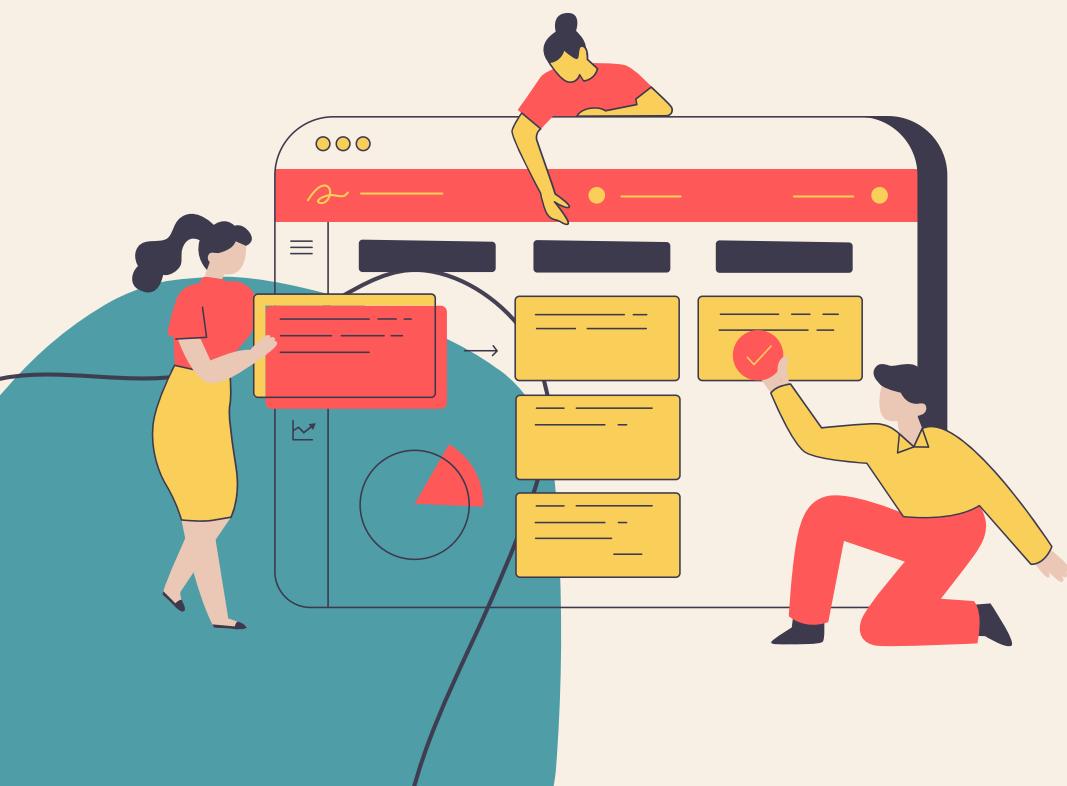
| Mean TF-IDF | |
|-------------|---------|
| remember | 0.18209 |
| well | 0.16613 |
| like | 0.14955 |
| cause | 0.14955 |
| thing | 0.07465 |
| maybe | 0.07465 |
| lost | 0.07465 |
| long | 0.07465 |
| bout | 0.07465 |
| know | 0.05269 |

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.

In V1.0, the word cloud and mean tf-idf table were deployed.



Sentiment Analysis

To classify the emotion of songs, we sourced a pre-trained Natural Language Processing model on Hugging Face by Manuel Romero⁵ 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 |
| <hr/> | | | | |
| 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).

This functionality was not deployed in V1.0 as it significantly slowed down the app due to the PyTorch library and inaccurate classification of certain genres and languages. The latency will need to be decreased in order to be deployed in a future version.

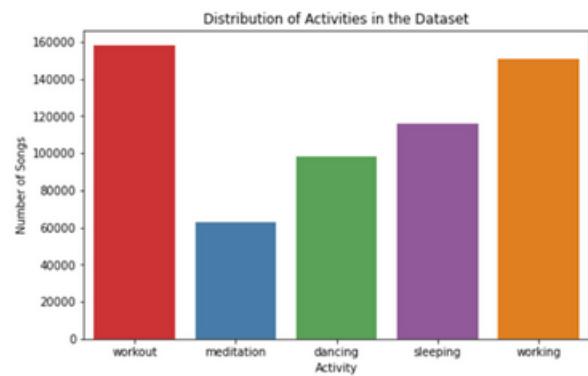
Activity Recommendations

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']
```



This model was not deployed in V1.0 as it is intended to be paired with the chatbot functionality that will be released in a future version.

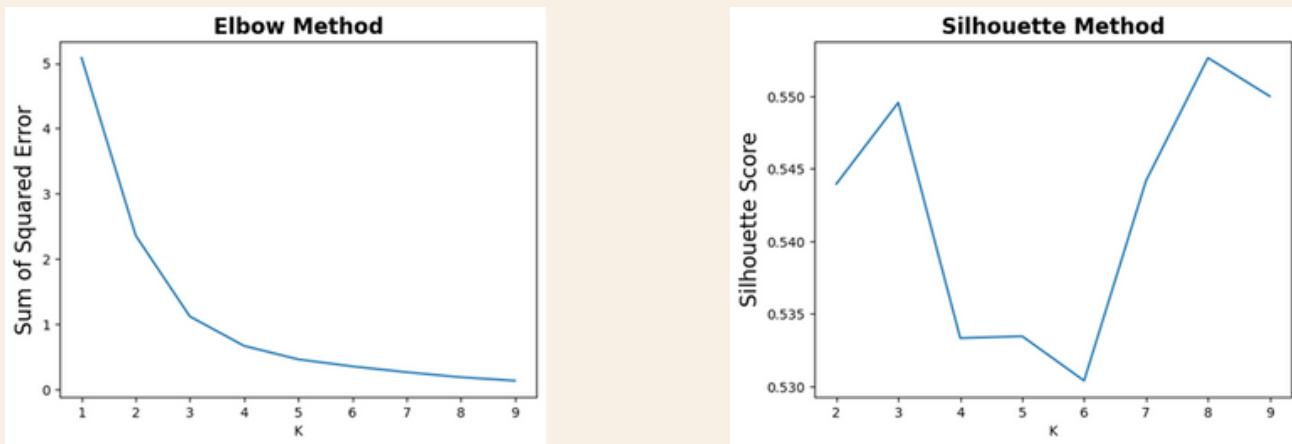
Playlist Comparison

The initial use case of this model was to find a suitable playlist for two individuals who are traveling together on a road trip and would like to find songs that both listeners would enjoy. We looked at methods to compare lists of songs against each other to determine underlying similarities that the user could not likely discern or group on their own. Clustering the playlist aggregate seemed like the logical choice for finding these answers. A cluster refers to a collection of data points aggregated together because of certain similarities. We overlaid the songs and their features from both lists and grouped those features into clusters using K-means clustering methodology.

We explored several clustering technologies before making a selection for our final clustering methodology. The objective of K-means is to group similar data points together and discover underlying patterns. K-means clustering was chosen for a variety of reasons in addition to its performance on our test data. K-means clustering is one of the simplest, popular, and most studied unsupervised machine learning algorithms available. Clustering has a long history and was actually introduced in 1932 by H.E. Driver and A.L.Kroeber in their paper on "Quantitative expression of cultural relationship". K-means is also comparatively fast when compared with other clustering algorithms, which are known to be relatively slow. Since K-means is an iterative algorithm that involves calculating the distance between each point in your data and the center of each cluster instead of calculating pairwise distance between points, this means that K-means scales linearly with the number of data points being assessed.

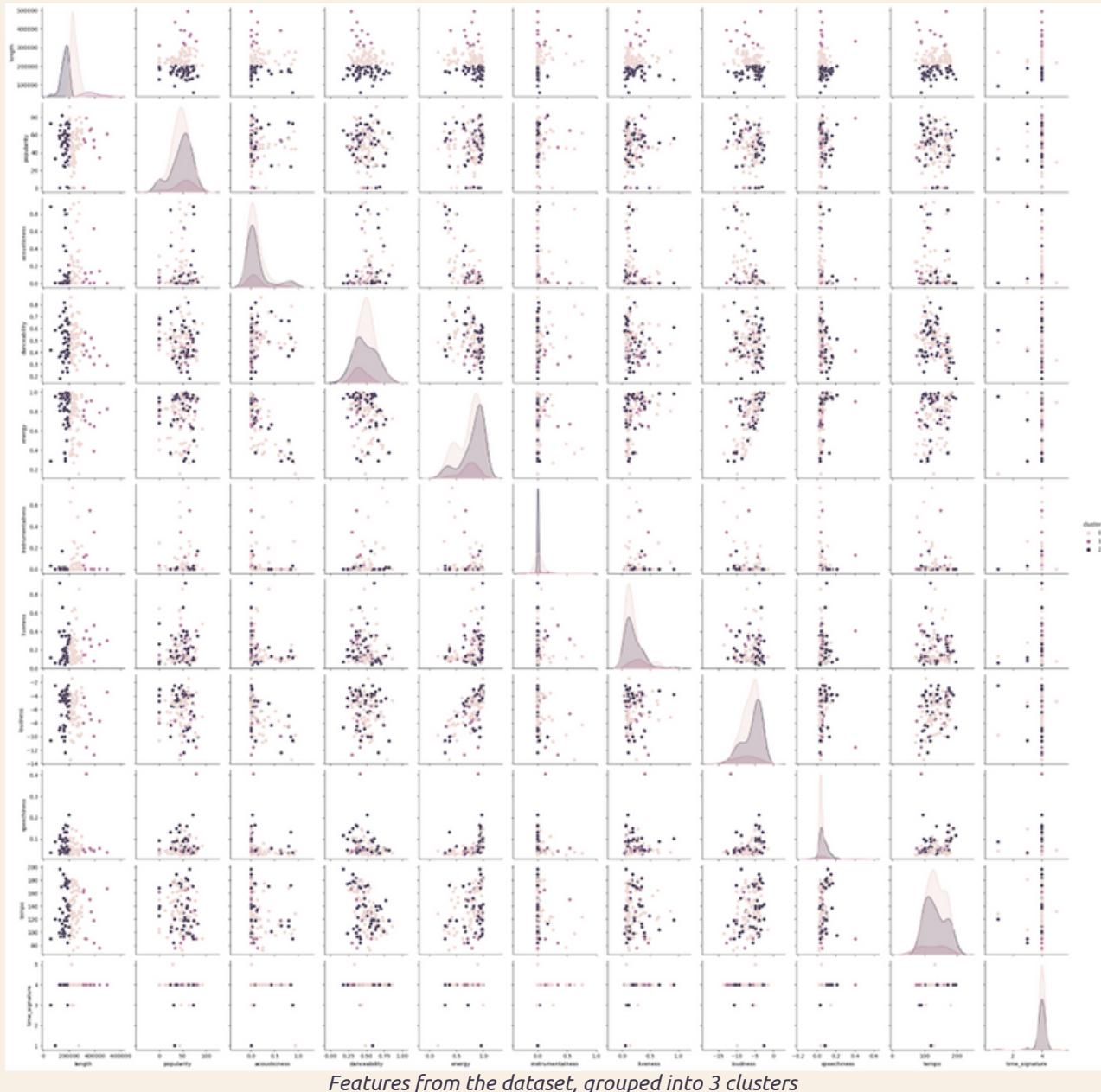


We used a combination of Elbow and Silhouette methods to determine the proper number of clusters (the value of K) for our implementation. In the Elbow Method, we plot the mean distance and look for the elbow point where the rate of decrease shifts. This elbow point can be used to determine K. In our dataset, Sum of Square Error (SSE) flattens out to a gradual improvement after a K value of 3. The Silhouette Method is used to measure the quality of the clusters by checking how similar a data point is within a cluster compared to the other clusters. In our dataset, the Silhouette score is highest with 8 clusters, but due to the complexity of using 8 clusters, 3 was chosen based on only involving a slight performance hit and the data gained from the Elbow Method.



After selection of the K value of 3, we ran multiple iterations using different playlists for comparison. We wanted to focus on the features that are most perceptible to listeners and chose three high level features of Danceability, Energy, and Valence. 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. Energy represents a perceptual measure of intensity and activity. Valence describes the musical positiveness of the track. We also added the feature of popularity to our filter to round out the track selection process based on whether listeners prefer very mainstream or more obscure titles.

After clustering the songs together, we select the middle cluster, with the most overlap between the two listeners. This provides the most similarity between songs and is a middle ground which should appeal to both listeners.



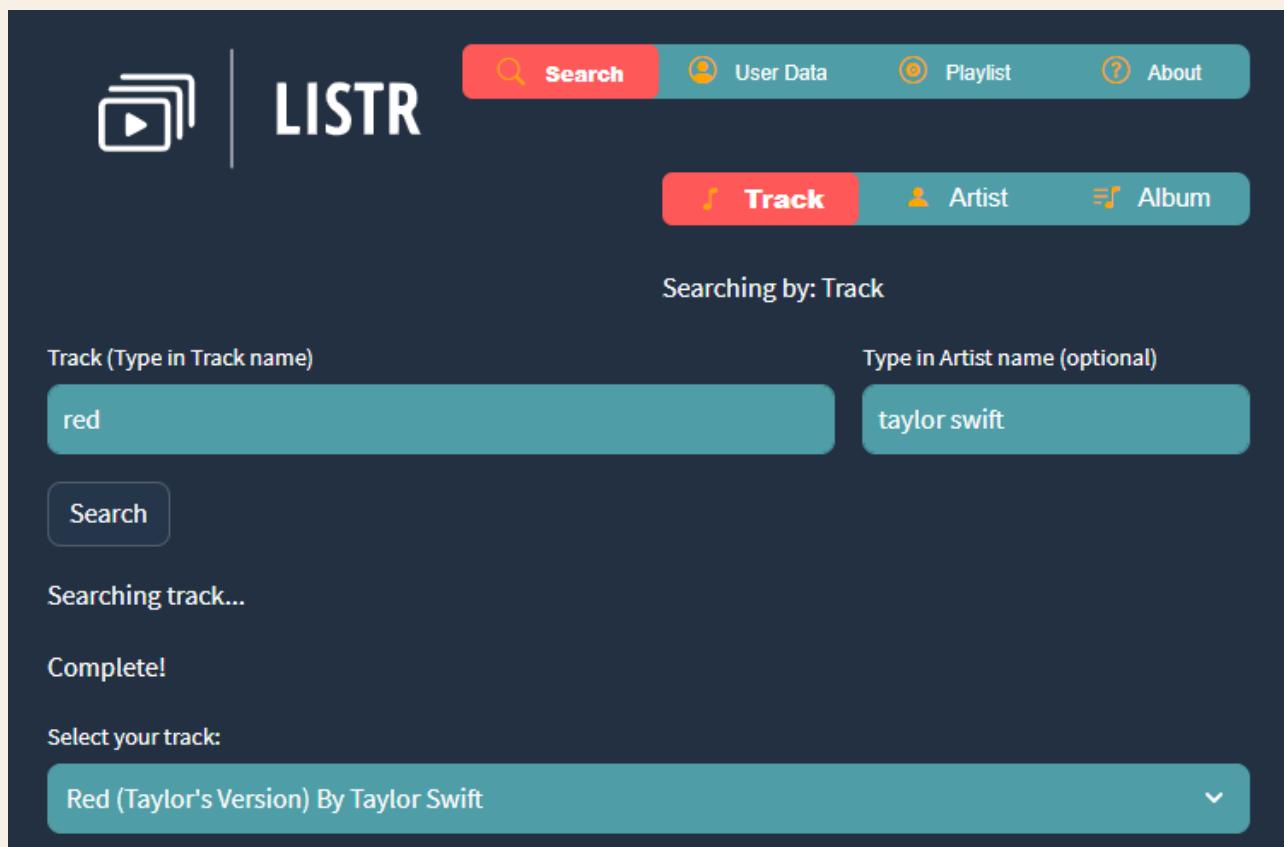
Features from the dataset, grouped into 3 clusters

Application

The app provides 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, they are able to search by track and artist (optional). The app then shows the details of the selected track and lyrics along with visualizations characterizing the song.



Select your track:

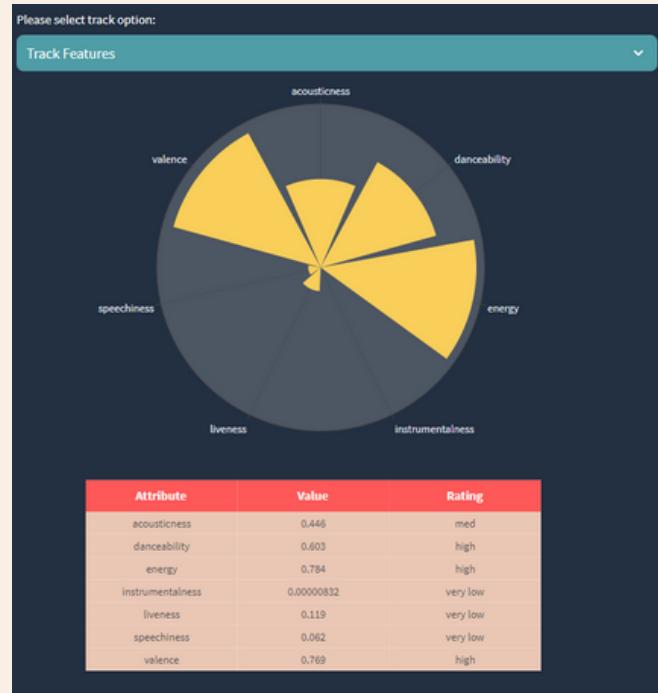
Red (Taylor's Version) By Taylor Swift

Please select track option:

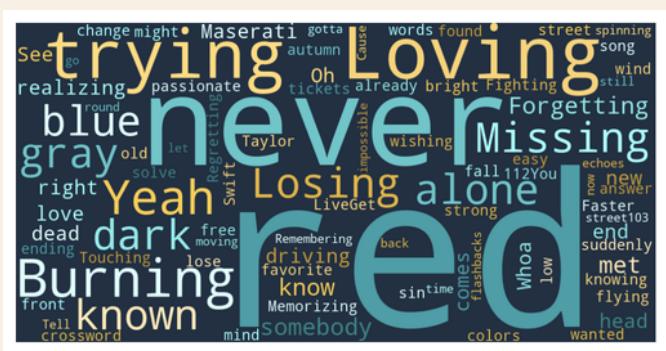
Track Features

| | acousticness | danceability | energy | instrumentalness | liveness | speechiness | valence |
|---|--------------|--------------|--------|------------------|----------|-------------|---------|
| 0 | 0.001 | 0.519 | 0.783 | 0 | 0.0776 | 0.0398 | 0.421 |

The initial results from track search



Visual of the song features



Word Cloud of the lyrics

Lyrics for 'Red' by Taylor Swift

Lyrics

Loving him is like driving a new Maserati down a dead-end street
Faster than the wind, passionate as sin, ending so suddenly
Loving him is like trying to change your mind once you're already flying through the free fall
Like the colors in autumn, so bright just before they lose it all

Losing him was blue like I'd never known
Missing him was dark gray, all alone
Forgetting him was like trying to know somebody you've never met
But loving him was red

(Red, red)
(Red, red)
Loving him was red
(Red, red)
(Red, red)

Touching him was like realizing all you ever wanted was right there in front of you

Full lyrics from Genius

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 Tracks Recommendation

| | name | artist | explicit | track duration | popularity |
|---|------------------------------|----------------------|-------------------------------------|----------------|------------|
| 0 | Same Boat | Lizzy McAlpine | <input type="checkbox"/> | 00:03:22 | 58 |
| 1 | I'm In Love With You | The 1975 | <input checked="" type="checkbox"/> | 00:04:22 | 74 |
| 2 | Best Friend Breakup | Lauren Spencer Smith | <input type="checkbox"/> | 00:02:21 | 66 |
| 3 | Summertime Sadness | Lana Del Rey | <input type="checkbox"/> | 00:04:24 | 65 |
| 4 | Liability | Lorde | <input type="checkbox"/> | 00:02:51 | 75 |
| 5 | don't come back | Tate McRae | <input type="checkbox"/> | 00:02:32 | 67 |
| 6 | this is me trying | Taylor Swift | <input type="checkbox"/> | 00:03:15 | 79 |
| 7 | Take A Chance With Me | NIKI | <input type="checkbox"/> | 00:05:03 | 59 |
| 8 | Music For a Sushi Restaurant | Harry Styles | <input type="checkbox"/> | 00:03:13 | 82 |
| 9 | Kids Are Born Stars | Laув | <input type="checkbox"/> | 00:03:07 | 64 |

Song Recommendations for "Red" by Taylor Swift



Artist Search

If the user chooses the "Artist" option, the app displays all the albums created by that artist along with the release date and song count. There is also an option to see that artist's top tracks.

Select your artist:

Taylor Swift

Start artist search...

Complete!

Select artist choice

Albums

| album type | album name | album release | track count |
|------------|-------------------------------------|---------------|-------------|
| album | Speak Now (Taylor's Version) | 2023-07-07 | 22 |
| album | Midnights (The Til Dawn Edition) | 2023-05-26 | 23 |
| album | Midnights (3am Edition) | 2022-10-22 | 20 |
| album | Midnights | 2022-10-21 | 13 |
| album | Red (Taylor's Version) | 2021-11-12 | 30 |
| album | Fearless (Taylor's Version) | 2021-04-09 | 26 |
| album | evermore (deluxe version) | 2021-01-07 | 17 |
| album | evermore | 2020-12-11 | 15 |
| album | folklore: the long pond studio sess | 2020-11-25 | 34 |
| album | folklore (deluxe version) | 2020-08-18 | 17 |

Select artist choice

Top Tracks

| | track name | track duration | danceability | energy | key | loudness | mode | speechiness | acou |
|---|-------------------|----------------|--------------|--------|-----|----------|------|-------------|------|
| 0 | Cruel Summer | 00:02:58 | 0.552 | 0.702 | 9 | -5.707 | 1 | 0.157 | |
| 1 | Anti-Hero | 00:03:20 | 0.637 | 0.643 | 4 | -6.571 | 1 | 0.0519 | |
| 2 | Blank Space | 00:03:51 | 0.753 | 0.678 | 5 | -5.421 | 1 | 0.0644 | |
| 3 | Style | 00:03:51 | 0.598 | 0.786 | 2 | -5.572 | 1 | 0.0383 | |
| 4 | august | 00:04:21 | 0.532 | 0.623 | 5 | -9.208 | 1 | 0.0331 | |
| 5 | cardigan | 00:03:59 | 0.613 | 0.581 | 0 | -8.588 | 0 | 0.0424 | |
| 6 | I Can See You (Ta | 00:04:33 | 0.694 | 0.764 | 6 | -4.893 | 1 | 0.0337 | |
| 7 | Karma | 00:03:24 | 0.642 | 0.62 | 8 | -7.091 | 1 | 0.0666 | |
| 8 | Don't Blame Me | 00:03:56 | 0.615 | 0.534 | 9 | -6.719 | 0 | 0.0386 | |
| 9 | Lover | 00:03:41 | 0.359 | 0.543 | 7 | -7.582 | 1 | 0.0919 | |

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.

Select your album:

folklore By Taylor Swift

Collecting all the tracks for the album: folklore

| | track name | explicit | track duration | danceability | energy | key | loudness | mode | spee |
|---|-------------------------|--------------------------|----------------|--------------|--------|-----|----------|------|------|
| 0 | the 1 | <input type="checkbox"/> | 00:03:30 | 0.765 | 0.348 | 0 | -6.963 | 1 | |
| 1 | cardigan | <input type="checkbox"/> | 00:03:59 | 0.613 | 0.581 | 0 | -8.588 | 0 | |
| 2 | the last great american | <input type="checkbox"/> | 00:03:50 | 0.684 | 0.661 | 7 | -8.507 | 1 | |
| 3 | exile (feat. Bon Iver) | <input type="checkbox"/> | 00:04:45 | 0.298 | 0.38 | 6 | -8.426 | 1 | |
| 4 | my tears ricochet | <input type="checkbox"/> | 00:04:15 | 0.456 | 0.263 | 0 | -10.63 | 1 | |
| 5 | mirrorball | <input type="checkbox"/> | 00:03:28 | 0.553 | 0.413 | 2 | -10.037 | 1 | |
| 6 | seven | <input type="checkbox"/> | 00:03:28 | 0.596 | 0.488 | 4 | -10.769 | 1 | |
| 7 | august | <input type="checkbox"/> | 00:04:21 | 0.532 | 0.623 | 5 | -9.208 | 1 | |
| 8 | this is me trying | <input type="checkbox"/> | 00:03:15 | 0.515 | 0.476 | 9 | -9.843 | 1 | |
| 9 | illicit affairs | <input type="checkbox"/> | 00:03:10 | 0.559 | 0.31 | 10 | -10.445 | 1 | |



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.

Although this feature is not live in the deployed app, there is a demo sample to showcase the functionality.

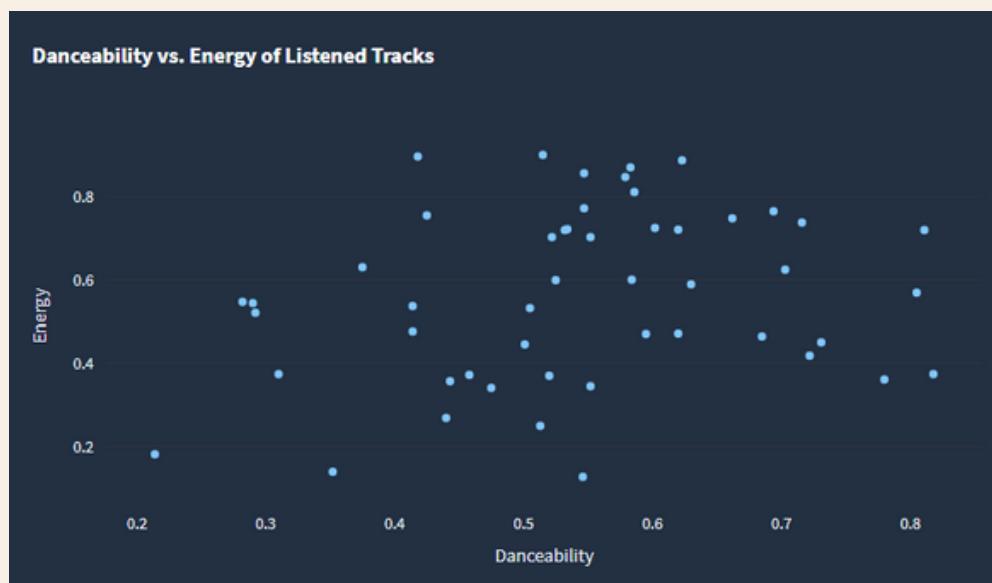
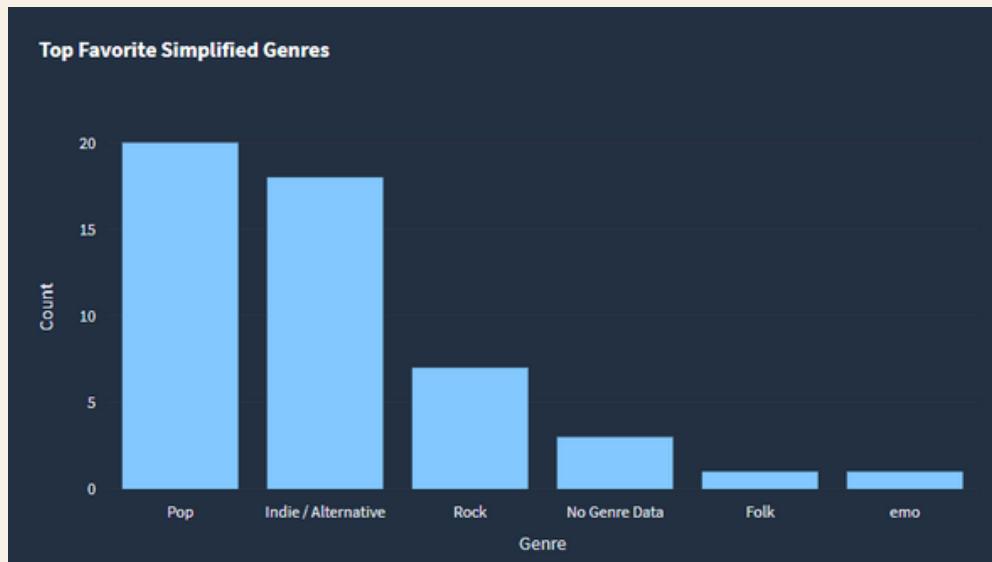
Your Spotify Listening Data

This feature is on its way and will be available shortly! In the meantime, feel free to explore the dashboard using the sample data provided by our team. Stay tuned for updates!

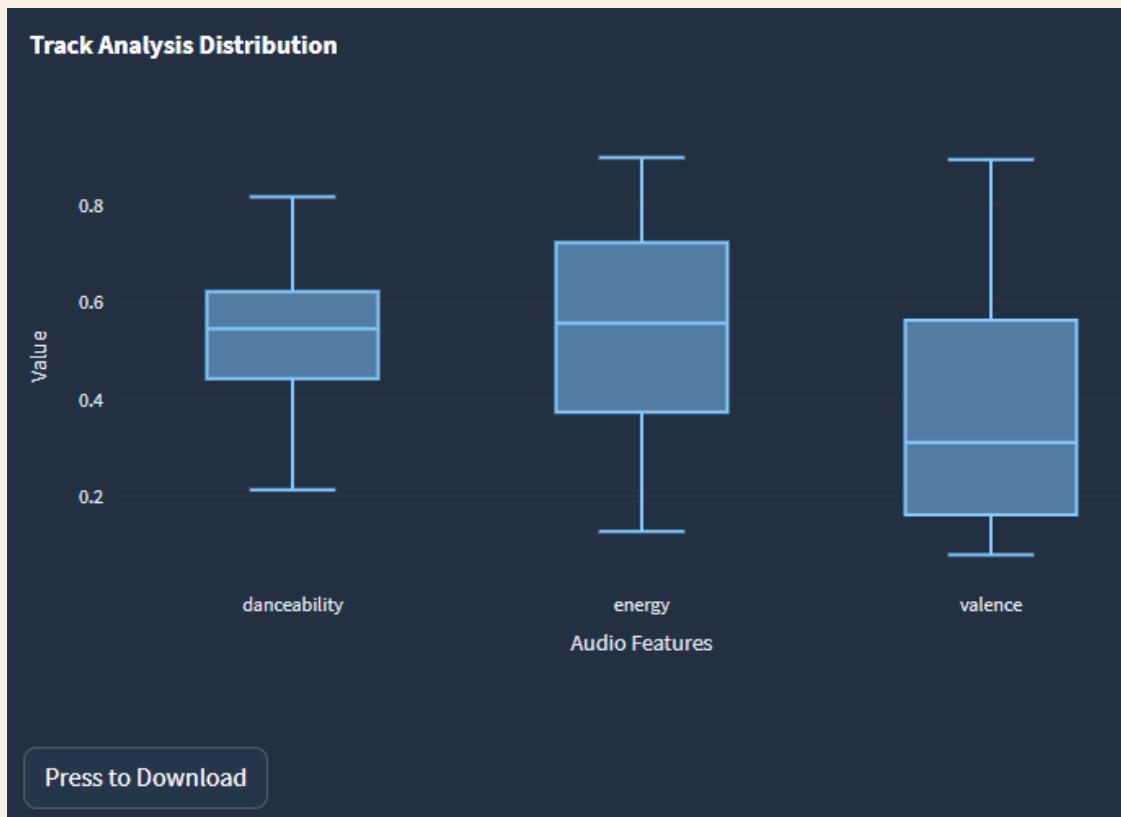
[Login to Spotify](#)

| | track name | track duration | track release | artist name | genre |
|---|--|----------------|---------------|------------------------|---------|
| 0 | Golden | 00:02:50 | 2023-04-19 | Barns Courtney | Rock |
| 1 | Shut Up Kiss Me | 00:03:22 | 2016-09-02 | Angel Olsen | Indie / |
| 2 | Mermaids | 00:04:35 | 2023-04-21 | Florence + The Machine | Pop |
| 3 | Sad Disco (Pine Studios) | 00:03:47 | 2023-02-14 | flipturn | Indie / |
| 4 | Gold | 00:03:57 | 2023-02-10 | Ivan & Alyosha | Pop |
| 5 | Anti-Hero - Roosevelt Remix | 00:04:59 | 2022-11-11 | Taylor Swift | Pop |
| 6 | Mermaids | 00:04:35 | 2023-04-21 | Florence + The Machine | Pop |
| 7 | I Can See You (Taylor's Version) (From The | 00:04:33 | 2023-07-07 | Taylor Swift | Pop |
| 8 | Miss Americana & The Heartbreak Prince | 00:03:54 | 2019-08-23 | Taylor Swift | Pop |
| 9 | Some things Cosmic | 00:02:54 | 2011-04-11 | Angel Olsen | Indie / |

Users are able to see visualizations depicting their favorite genres and the danceability and energy spread of their songs.



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.

Playlist Comparison

This functionality is designed for friends to create playlists that would be enjoyed by both. Just enter the link of two **public** playlists to get a new playlist built from the originals.

Playlist Comparison

Let's compare two Spotify playlist of your choosing. Please make sure that they are public playlist.

In Spotify, go to the playlist and click the three '...' >> 'Share' >> 'Copy Link to Playlist'

Copy and paste the URL link to the playlist in Listr.

Sample URL: <https://open.spotify.com/playlist/5Qnr7ct4jZ4ad1yDRuM1to?si=0bdc634e4b6c43f6>

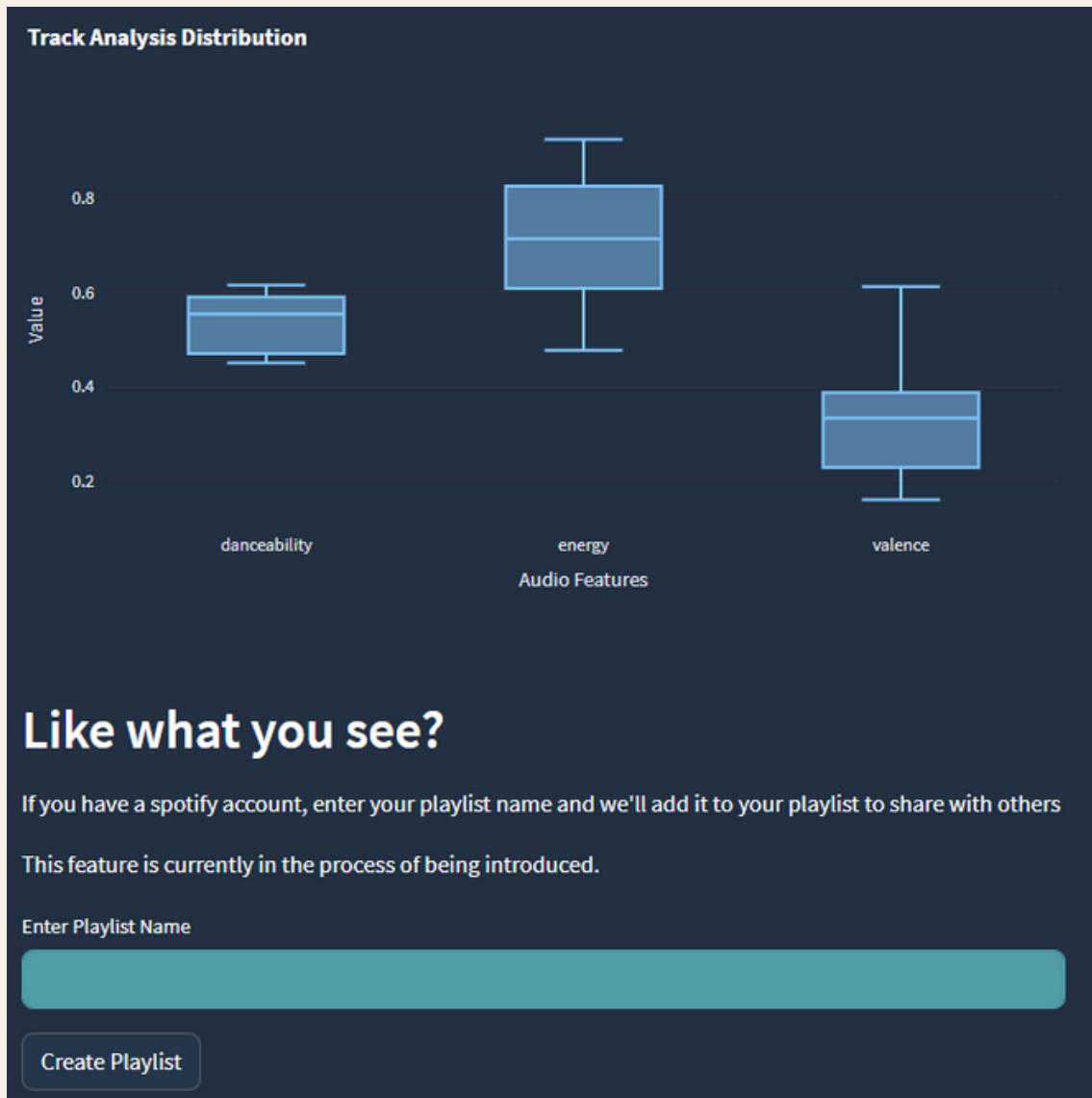
Enter Playlist 1 URL or Playlist ID

Enter Playlist 2 URL or Playlist ID

Analyze Playlist

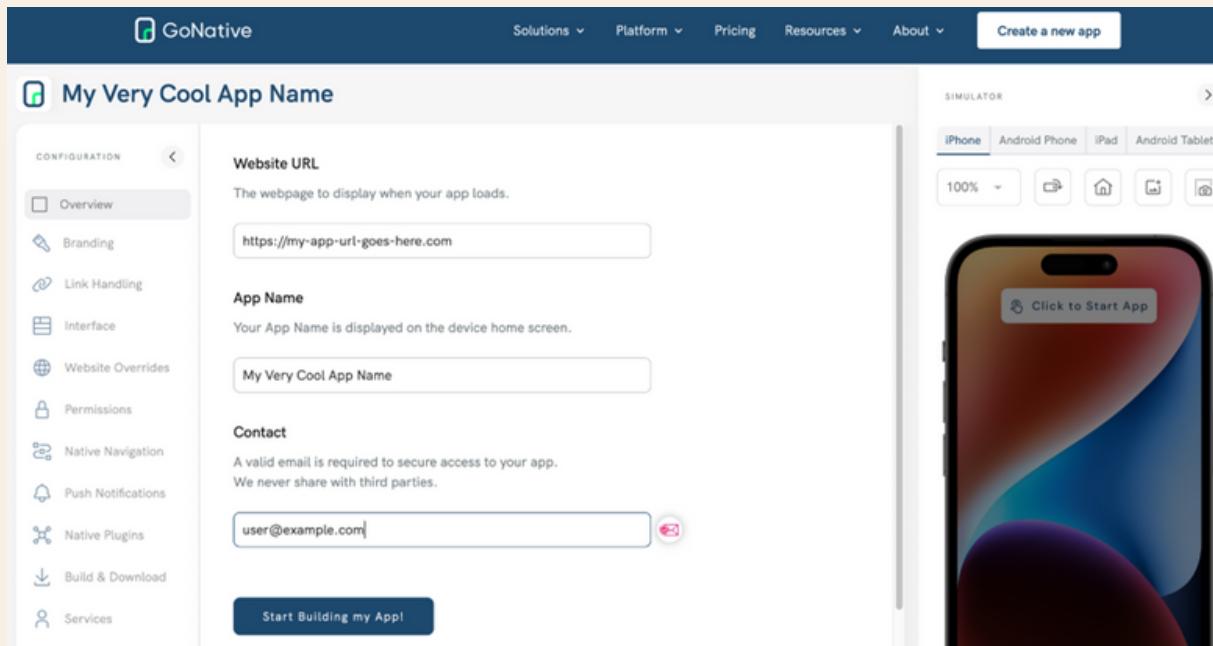
| track_id | track_name | artist | album |
|------------------------|-----------------------------------|----------------|---------------------------------|
| 3x7NGlfDKQYN2wnwEnL7r9 | The Kids Aren't Alright | Fall Out Boy | American Beauty/American Psycho |
| 2BedHMJWiW5CuiTd2wza1u | Sad Disco | flipturn | Shadowglow |
| 3x7NGlfDKQYN2wnwEnL7r9 | The Kids Aren't Alright | Fall Out Boy | American Beauty/American Psycho |
| 3HFvMlvCBNGkypFCQUHjyV | Supernatural | Barns Courtney | Supernatural |
| 6YArGkJUr82ehzRFa2YaRK | Wake Me | Bleachers | Strange Desire |
| 6lzc0Al0zfZOIFsFvBS1ki | State Of Grace (Taylor's Version) | Taylor Swift | Red (Taylor's Version) |
| 3HFvMlvCBNGkypFCQUHjyV | Supernatural | Barns Courtney | Supernatural |
| 6lzc0Al0zfZOIFsFvBS1ki | State Of Grace (Taylor's Version) | Taylor Swift | Red (Taylor's Version) |

Visualizations describing the new playlist are displayed. In future versions, users will also be able to automatically add this playlist onto Spotify to share with others.



Mobile Application

After creation and deployment of our web application, we wanted to attempt fully integrating the web application into a mobile device application that could be downloaded and installed on a phone or tablet by users. Research led us down multiple paths that ended selection of GoNative, who provides an end-to-end solution to deploy websites and web-based applications as native iOS and Android apps on mobile devices (<https://gonative.io/>). GoNative offers several benefits over traditional application development by converting your existing website into a native app, requiring us to maintain a single web codebase rather than separate iOS and Android apps. To create a cross platform application, users can navigate to (<https://gonative.io/app>) and walk through the process to create an application in a matter of minutes. To get started, users only needs the website URL, Application Name, and a Contact Email address.

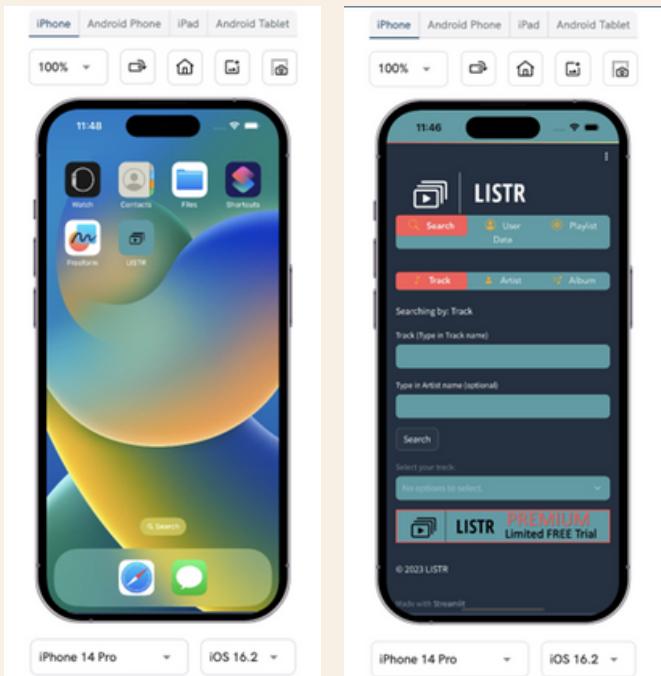


Once this information entered, the application builder launches and allows users to customize a host of options including:

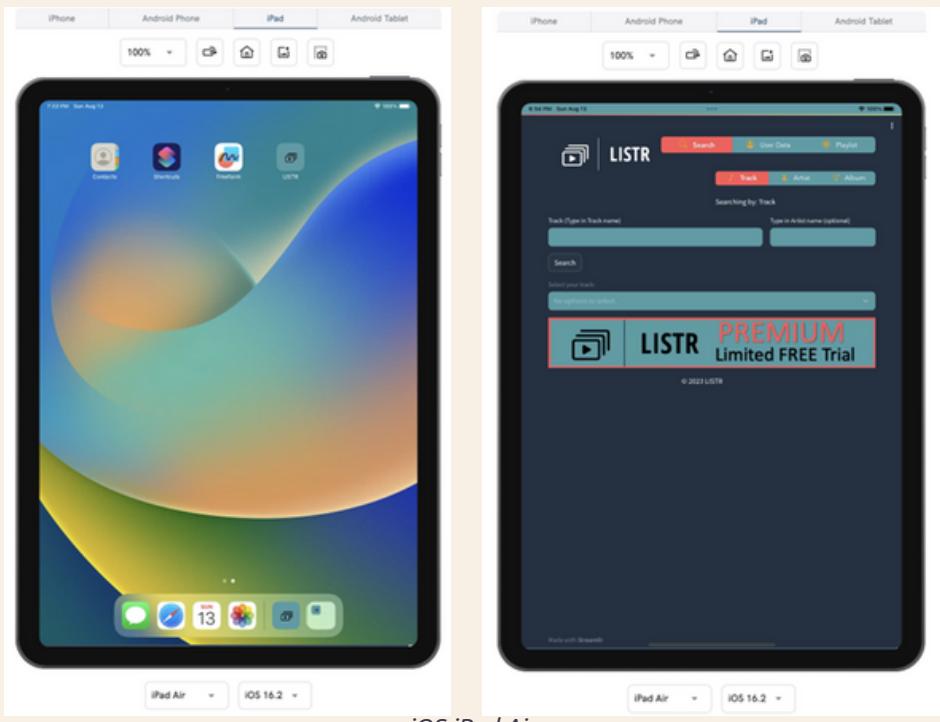
- Branding - Application Icon, Splash Screen, Light/Dark Mode Options, etc
- Link Handling - Select items to open within app or with default browser
- Interface - Keep Screen On, Full Screen, Screen Orientation, etc
- Website Overrides - User Agent, Custom Headers, Custom CSS, Custom JavaScript, etc
- Permissions - App Tracking Transparency, Location Services, Downloads, Audio, etc
- Native Navigation - Top and Sidebar Navigation options, Toolbars, etc
- Push Notifications - Push Notification, requires a license to upgrade
- Native Plugins - Multiple paid upgrades to application, ranging from \$200 to \$5500+
- Build & Download - Rebuild and download Application Source, APK, AAB files
- Services - GoNative paid services for app development and publishing pricing
- Support - Free and paid support/update options and pricing
- License - License purchase options and pricing

Our application is now launched and available for both iOS and Android platforms, usable across a variety of phones and tablets. To access a mobile device emulator and view the application, testers can navigate to <https://gonative.io/share/ejqdpk>. Users can select the system, device, and operating system to test the application in an emulator environment for multiple models of device.

iOS

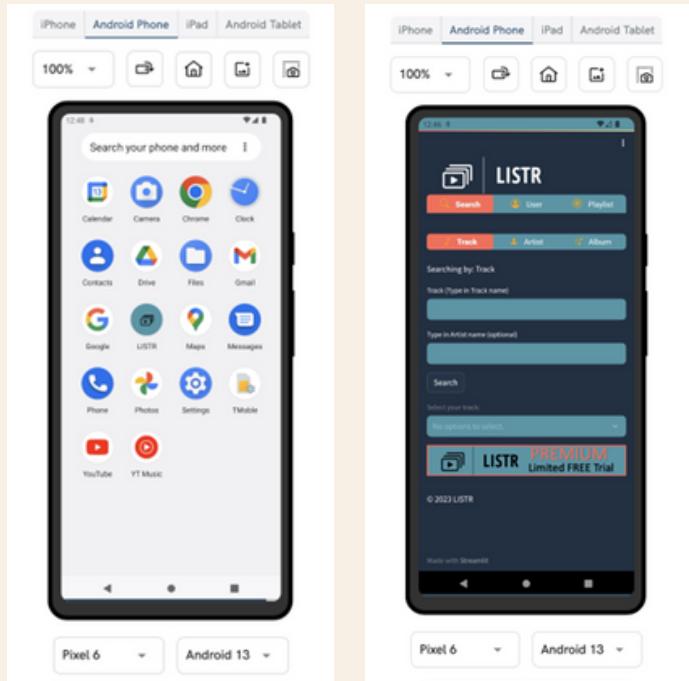


iOS iPhone 14 Pro

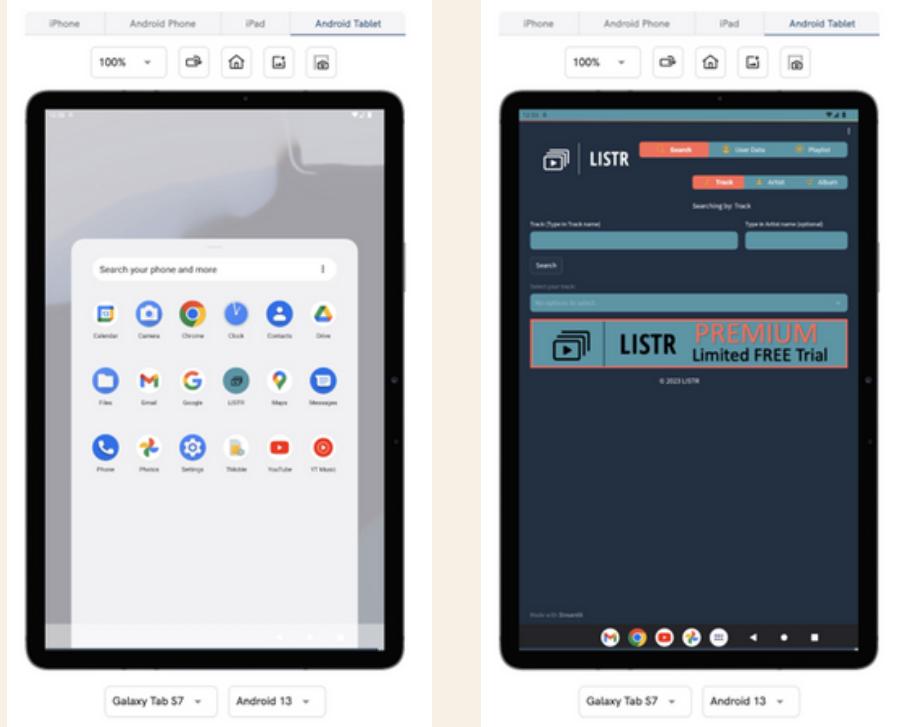


iOS iPad Air

Android



Android Google Pixel 6

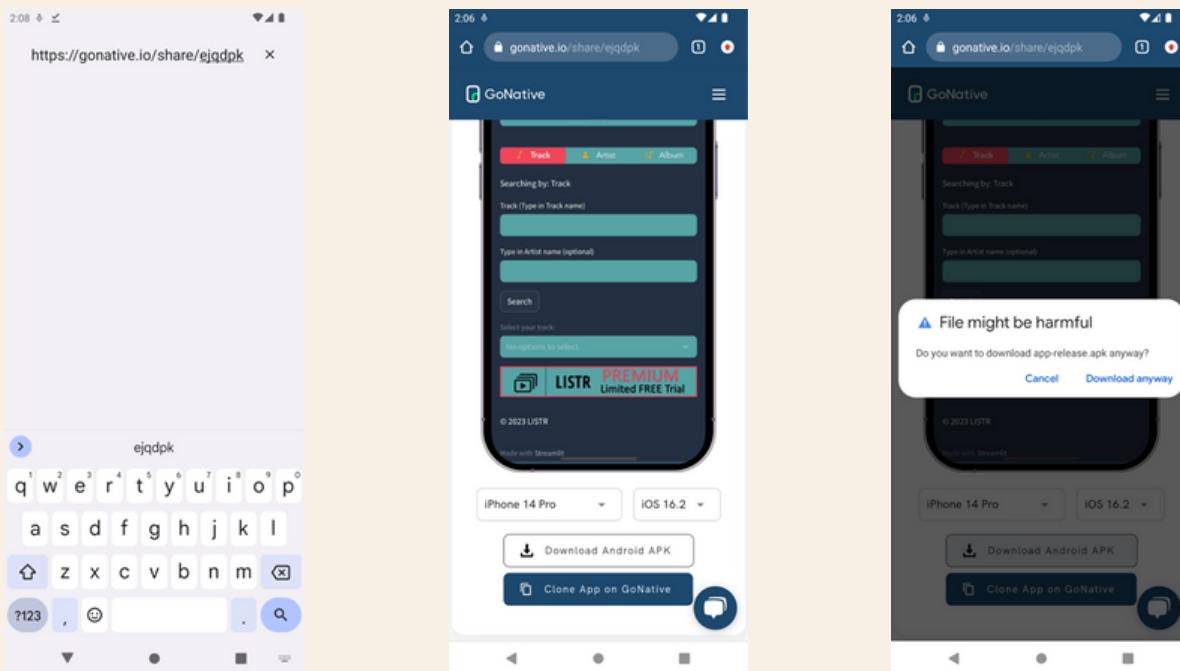


Android Samsung Galaxy Tab S7

Submissions of applications to the official application store for Android and iOS come with different pricing options, but there are no free options. The Google Play Store carries a one-time, \$25 fee for publishing an application and publishing to the Apple App Store by enrolling in the Apple Developer Program costs \$99 per year. To keep application development costs to a minimum, our team developed and self published our application using source code and Android Package Kit (APK) file options. In order to utilize the iOS application on a physical device via source code, a user must have an Apple Developer account. Android is more flexible, and allows installation of an APK file for any user.

The installation process for the application via APK file is as follows:

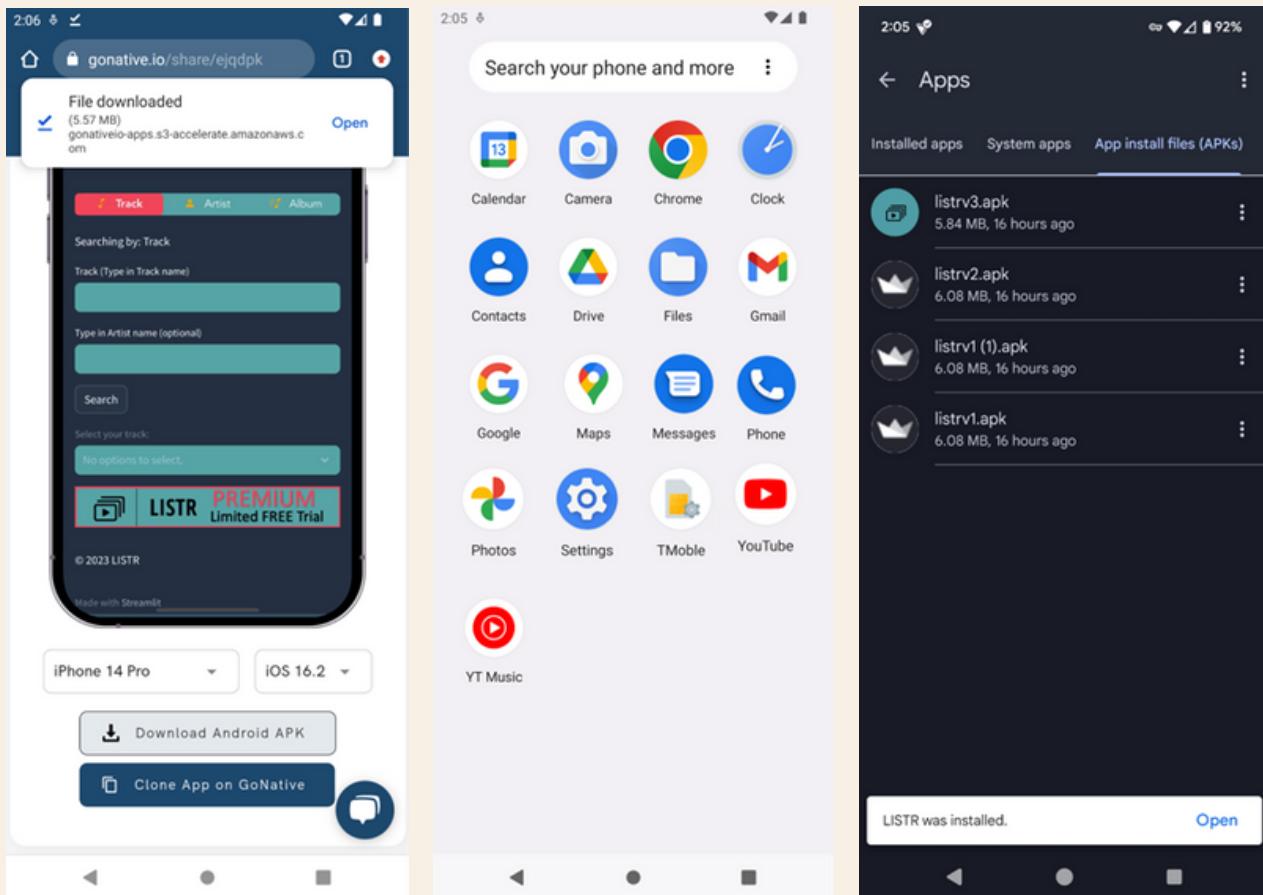
1. Navigate to our Application Public Sharing Page at <https://gonative.io/share/ejqdpk> on your Android mobile device.
2. Click "Download Android APK" near the bottom center of the page.



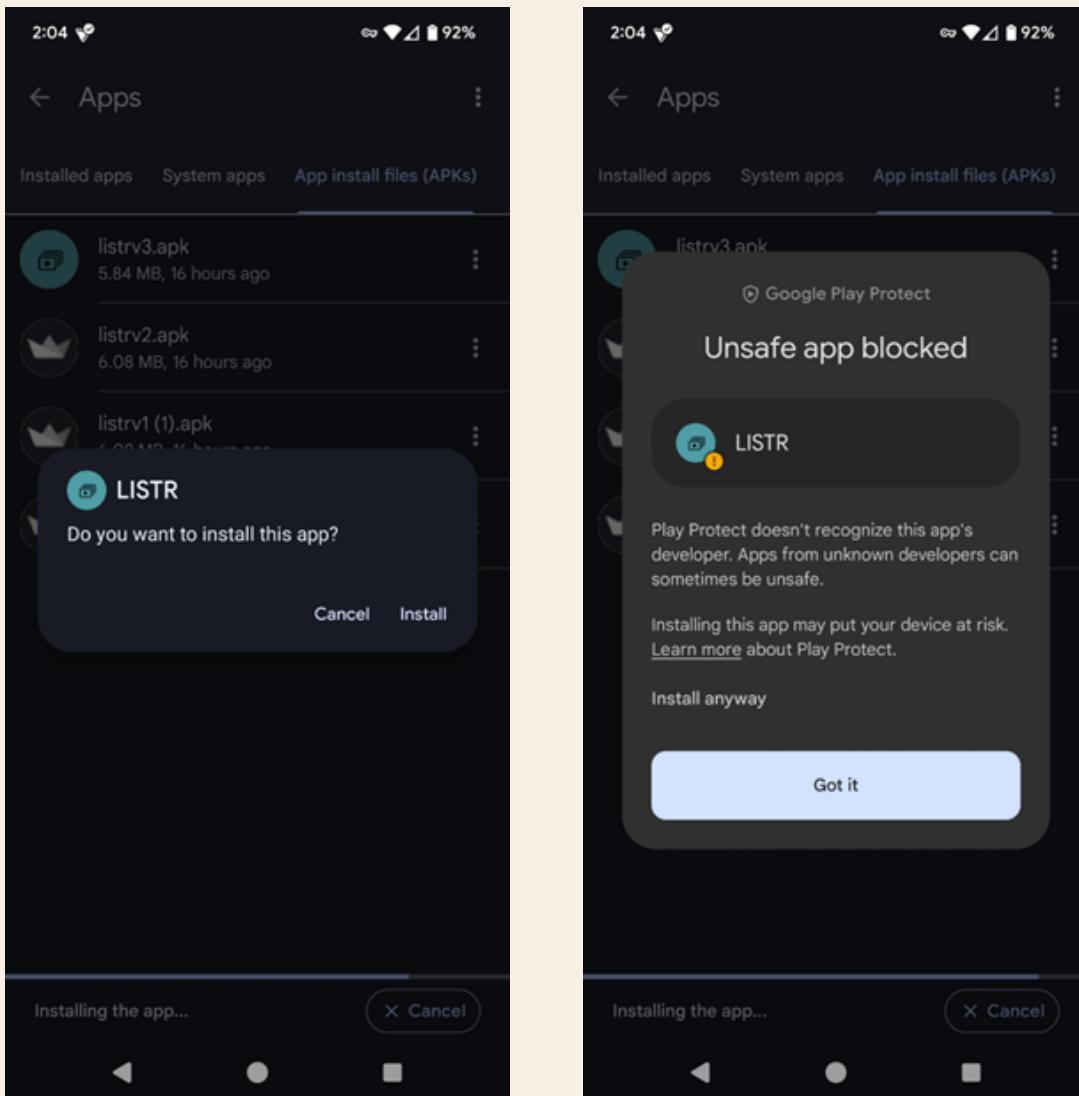
You may receive a warning that the file might be harmful. You can ignore this warning since it is because our application file is not downloaded from the Google Play store as Google prefers. Select "Download Anyway" if this message appears.

4. Depending on how your device is set up, once the download is complete, it may present a popup to open the file.

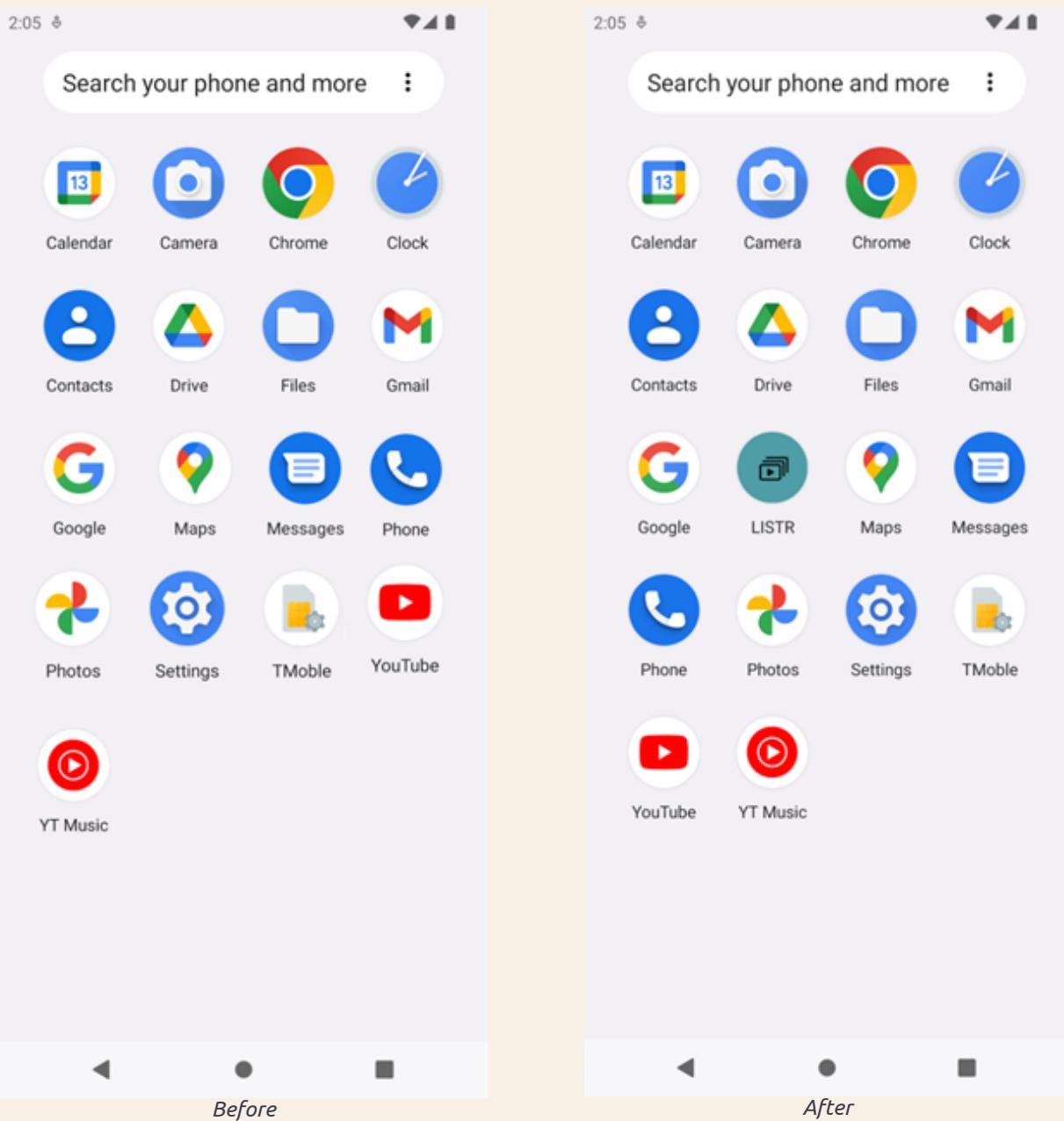
- a. If your device presents a popup that says "File Downloaded, click on "Open"
- b. If your device does not present a popup, navigate to "Files" or "My Files" and select the LISTR file. NOTE: There should be a category for Apps or APKs (if not, you can just choose downloads or most recent and it should show up).



5. The screen will ask "Do you want to install this app?", Click "Install"
6. You will likely get a warning from Google Play Protect stating that it has blocked an unsafe application. This is totally OK and occurs because our application has not been scanned by Google Play Protect since it is not from the Google Play Store. Click on "more details" to reveal an option for "Install anyway" and click this option.



7. The application will install and now be visible in your applications area. This is usually accessible by swiping up on your device from the bottom of the screen. The application will be visible similar to the following.



Future Enhancements

Chatbot

Allows app users to ask a chatbot for recommendations based on similar songs, artists, moods, or activities

Song Recs

The current recommendation engine is solely based on Spotify's API. There are additional internal models in the works that will enhance those recommendations

Clusters

There will be clustering functionalities that will allow users to compare albums, artists, and more

User Data

Users will be able to seamlessly log into their Spotify account and extract their user data, generate user reports, and get recommendations based on their activity



Financial Projections

Listr is projected to build a 500K user base in the first year based on market research conducted. The cost per subscription will be \$1.99 monthly, so this will yield \$12M in revenue, but still -\$13M in profits due to the initial start up costs.

There was an initial investment of \$10M and an additional \$80M of funding was secured which will help over the first year's worth of overhead costs.

At 1M users, the application will break even and begin turning a profit.

Year 1

\$12M

500K Users

Initial User Base

Ad Revenue

Year 2

\$48M

2M Users

Marketing & Product

Enhancement Efforts

Year 3

\$120M

5M Users

International Expansion

Sensitivity Analysis

| User Acquisition Rate | Annual Revenue | Annual Profit |
|-------------------------|----------------|---------------|
| Base Case (50%) | \$298M | \$273M |
| Best Case (70%) | \$419M | \$394M |
| Worst Case (30%) | \$183M | \$158M |

| Subscription Fee | Annual Revenue | Annual Profit |
|----------------------------|----------------|---------------|
| Base Case (\$1.99) | \$298M | \$273M |
| Best Case (\$24.9) | \$372M | \$347M |
| Worst Case (\$1.79) | \$226M | \$201M |

| Operational Costs | Annual Revenue | Annual Profit |
|---------------------------|----------------|---------------|
| Base Case (\$25M) | \$298M | \$273M |
| Best Case (\$20M) | \$345M | \$300M |
| Worst Case (\$30M) | \$271M | \$246M |

Closing Remarks

Although the scope shifted from the beginning of the project, Listr still launched with many interesting features for music lovers, filling a gap in the current market. The web based application is compatible on both desktop and mobile browsers with the eventual plan to launch on iOS and Android app stores.

Future version releases will introduce new and improved features such as a chatbot for recommendations and seamless integration with Spotify.

Listr is currently projected to bring in \$12M in revenue in Year 1 and breaking even halfway through Year 2. After that, Listr will begin turning a profit.



References

"Build with Lyrics." Musixmatch Developer. Accessed July 8, 2023.
[https://developer.musixmatch.com/.](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%20subscribers%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](https://huggingface.co/mrm8488/t5-base-finetuned-emotion). 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>.

