



Music Recommendation System

Presented by:

Parul (044)

Palak Mangla (043)

Palak Jethwani (042)

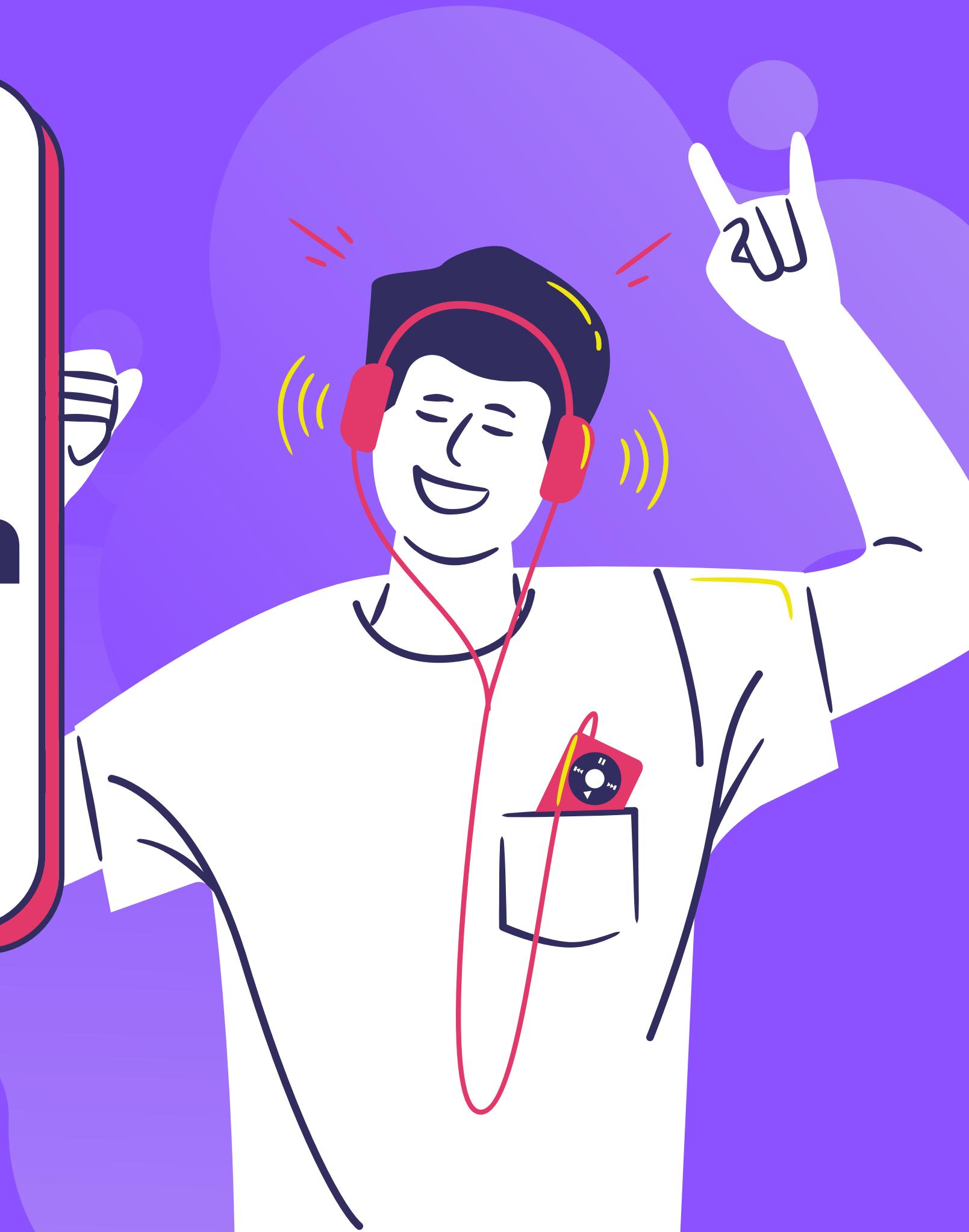


Table Of Content



Objective

01

Dataset

02

Methodology

03

Model

04

Visualisation

05

Inferences

06

OBJECTIVE



The purpose of this project is to develop a music recommendation system based on user preferences and song features. The system aims to provide personalized song recommendations to users based on their favourite genres and artists.



DATASET

The music dataset used in this project is obtained from the "musicdataset.csv" file. It contains information about various songs, including their titles, artists, genres, and popularity ratings. The dataset was loaded into R using the read.csv function.

| Index | Title | Artist | Top.Genre | Year | Beats.Per.Minute..BPM. | Energy | Danceability | Loudness..dB. | Liveness | Length..Duration. | Popularity |
|-------|---------------------------------------|-------------------|----------------------|------|------------------------|--------|--------------|---------------|----------|-------------------|------------|
| 1 | Sunrise | Norah Jones | adult standards | 2004 | 157 | 30 | 53 | -14 | 11 | 201 | 71 |
| 2 | Black Night | Deep Purple | album rock | 2000 | 135 | 79 | 50 | -11 | 17 | 207 | 39 |
| 3 | Clint Eastwood | Gorillaz | alternative hip hop | 2001 | 168 | 69 | 66 | -9 | 7 | 341 | 69 |
| 4 | The Pretender | Foo Fighters | alternative metal | 2007 | 173 | 96 | 43 | -4 | 3 | 269 | 76 |
| 5 | Waitin' On A Sunny Day | Bruce Springsteen | classic rock | 2002 | 106 | 82 | 58 | -5 | 10 | 256 | 59 |
| 6 | The Road Ahead (Miles Of The Unknown) | City To City | alternative pop rock | 2004 | 99 | 46 | 54 | -9 | 14 | 247 | 45 |
| 7 | She Will Be Loved | Maroon 5 | pop | 2002 | 102 | 71 | 71 | -6 | 13 | 257 | 74 |
| 8 | Knights of Cydonia | Muse | modern rock | 2006 | 137 | 96 | 37 | -5 | 12 | 366 | 69 |
| 9 | Mr. Brightside | The Killers | modern rock | 2004 | 148 | 92 | 36 | -4 | 10 | 223 | 77 |
| 10 | Without Me | Eminem | detroit hip hop | 2002 | 112 | 67 | 91 | -3 | 24 | 290 | 82 |
| 11 | Love Me Tender | Elvis Presley | adult standards | 2002 | 109 | 5 | 44 | -16 | 11 | 162 | 49 |
| 12 | Seven Nation Army | The White Stripes | alternative rock | 2003 | 124 | 46 | 74 | -8 | 26 | 232 | 74 |
| 13 | Als Het Golft | De Dijk | dutch indie | 2000 | 102 | 88 | 54 | -6 | 53 | 214 | 34 |
| 14 | I'm going home | Ten Years After | album rock | 2005 | 117 | 93 | 38 | -2 | 81 | 639 | 26 |
| 15 | Fluorescent Adolescent | Arctic Monkeys | garage rock | 2007 | 112 | 81 | 65 | -5 | 14 | 173 | 66 |
| 16 | Zonder Jou | Paul de Leeuw | dutch cabaret | 2006 | 133 | 42 | 42 | -10 | 16 | 236 | 48 |
| 17 | Speed of Sound | Coldplay | permanent wave | 2005 | 123 | 90 | 52 | -7 | 7 | 288 | 69 |
| 18 | Uninvited | Alanis Morissette | alternative rock | 2005 | 127 | 54 | 38 | -5 | 9 | 276 | 57 |

Showing 1 to 18 of 799 entries, 12 total columns

CONTENTS OF DATASET

- **Index:** ID
- **Title:** Name of the Track
- **Artist:** Name of the Artist
- **Top Genre:** Genre of the track
- **Year:** Release Year of the track
- **Beats per Minute (BPM):** The tempo of the song
- **Energy:** The energy of a song - the higher the value, the more energetic. song
- **Danceability:** The higher the value, the easier it is to dance to this song.
- **Loudness:** The higher the value, the louder the song.
- **Length:** The duration of the song.
- **Speechiness:** The higher the value the more spoken words the song contains
- **Popularity:** The higher the value the more popular the song is.

METHODOLOGY

Data Acquisition and Preprocessing

The music dataset used in this project was sourced from the "musicdataset.csv" file. The data was loaded into r using read.csv function. To ensure data quality, missing values in the dataset were examined using the sum(is.na(music_data)) command. The preprocessing phase involved several steps, including data cleaning and feature engineering, to prepare the dataset for the recommendation system.

Feature Extraction and Representation

Three main features were considered for song representation: keywords extracted from song titles, genre and artist information. The song titles were processed to extract keywords, which were then combined to form a consolidated keyword representation for each song.

User Profile Creation

The user profile was created based on the user's favourite genres. In this project, the user's favourite genres were set as "dutch pop" and "pop". A user profile matrix was created to match the dimensions of the feature matrix, and similarity scores were computed using the cosine similarity measure.

Similarity Calculation

To determine the similarity between the user profile and the song features, cosine similarity was employed. The cosine similarity measure computes the cosine of the angle between two vectors and provides a similarity score ranging from 0 to 1. The similarity scores were calculated using the simil function from the proxy package.

Recommendation Generation

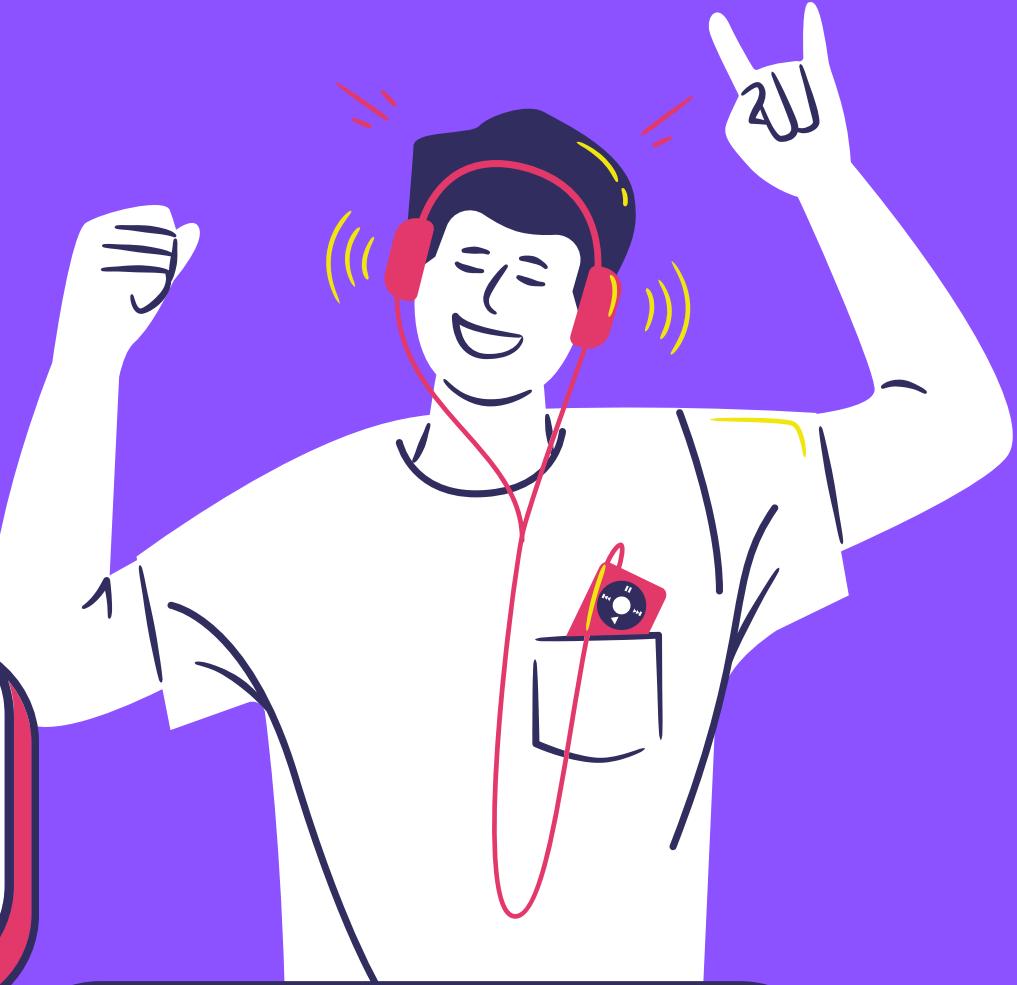
Based on the similarity scores, song recommendations were generated for the user. The similarity scores were merged with the original dataset, and the recommendations were sorted in descending order of similarity. The top N recommendations were selected to provide a personalized list of songs for the user.



MODEL

The word "MODEL" is written in large, bold, dark blue capital letters. The letters are partially obscured by a vibrant, multi-colored trail of musical notes that sweeps from the bottom left towards the top right, creating a sense of motion and rhythm.

Recommendation on the basis of GENRES



Rock

Pop

EDM

Country

Hiphop

Jazz

K-Pop

Blues

Reggae



GENRE MATRIX

FEATURE MATRIX

Songs x song_profiles1 x Untitled1 x .Rhistory x Untitled2* x user_profile_matrix x feature_matrix x genre_m x

Filter | Cols: << 1 - 50 >>

| | adult standards | album rock | alternative hip hop | alternative metal | classic rock | alternative pop rock | pop | modern rock | detroit hip hop | alternative rock | dutch indie | garage rock | country | punk | rock | metal | indie | electronic | rap | hip hop | pop rock | classic rock | alternative metal | alternative hip hop | adult standards | genre matrix |
|----|-----------------|------------|---------------------|-------------------|--------------|----------------------|-----|-------------|-----------------|------------------|-------------|-------------|---------|------|------|-------|-------|------------|-----|---------|----------|--------------|-------------------|---------------------|-----------------|--------------|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

Showing 1 to 17 of 799 entries, 116 total columns

USER PROFILE

USER PROFILE MATRIX

| genre_matrix | |
|--------------|---|
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 0 |
| 6 | 0 |
| 7 | 1 |
| 8 | 0 |
| 9 | 0 |
| 10 | 0 |
| 11 | 0 |
| 12 | 0 |
| 13 | 0 |
| 14 | 0 |
| 15 | 0 |
| 16 | 0 |
| 17 | 0 |
| 18 | 1 |

SIMILARITY DATAFRAME CONTAINING SIMILARITY SCORES OF FEATURE MATRIX AND USER PROFILE MATRIX

The screenshot shows a data frame titled "similarity_df" in an RStudio environment. The data frame has two columns: "Index" and "similarity_scores". The "Index" column contains integer values from 1 to 18. The "similarity_scores" column contains floating-point numbers. Most values are 0.0000000, except for the entry where both indices are 6, which is 0.4082483.

| Index | similarity_scores |
|-------|-------------------|
| 1 | 0.0000000 |
| 2 | 0.0000000 |
| 3 | 0.0000000 |
| 4 | 0.0000000 |
| 5 | 0.0000000 |
| 6 | 0.4082483 |
| 7 | 0.7071068 |
| 8 | 0.0000000 |
| 9 | 0.0000000 |
| 10 | 0.0000000 |
| 11 | 0.0000000 |
| 12 | 0.0000000 |
| 13 | 0.0000000 |
| 14 | 0.0000000 |
| 15 | 0.0000000 |
| 16 | 0.0000000 |
| 17 | 0.0000000 |
| 18 | 0.0000000 |

Showing 1 to 19 of 799 entries, 2 total columns

SORTED RECOMMENDATIONS

Screenshot of an RStudio session showing a sorted recommendations table.

The session bar at the top includes tabs for r_profile, sorted_recommendations, top_songs, song_profiles1, Untitled1, .Rhistory, Untitled2*, and use.

The main area displays a data frame with the following columns:

| Index | Title | Artist | Top.Genre | Year | Beats.Per.Minute.. |
|-------|-------------------------------------------|-------------------------------------------------|-----------|------|--------------------|
| 23 | 23 Als De Morgen Is Gekomen | Jan Smit | dutch pop | 2006 | |
| 25 | 25 Dichterbij Dan Ooit | BLØF | dutch pop | 2002 | |
| 40 | 40 De Weg | Guus Meeuwis | dutch pop | 2005 | |
| 52 | 52 Dansen Aan Zee | BLØF | dutch pop | 2000 | |
| 82 | 82 Wêr Bisto | Twarres | dutch pop | 2001 | |
| 86 | 86 Aanzoek Zonder Ringen | BLØF | dutch pop | 2006 | |
| 94 | 94 Blauwe Ruis | BLØF | dutch pop | 2002 | |
| 114 | 114 Bloed, Zweet En Tranen | Andre Hazes | dutch pop | 2002 | |
| 123 | 123 Voltooid Verleden Tijd | IOS | dutch pop | 2009 | |
| 148 | 148 Zij Maakt Het Verschil | De Poema's | dutch pop | 2003 | |
| 177 | 177 Hier | BLØF | dutch pop | 2000 | |
| 181 | 181 Father & Friend | Alain Clark | dutch pop | 2007 | |
| 205 | 205 Pak Maar M'n Hand | Nick & Simon | dutch pop | 2007 | |
| 228 | 228 Mijn Houten Hart | De Poema's | dutch pop | 2003 | |
| 256 | 256 Rain Down on Me | Kane | dutch pop | 2003 | |
| 267 | 267 Omarm | BLØF | dutch pop | 2003 | |
| 282 | 282 Laat Me / Vivre – Lange Versie | Alderliefste met Ramses Shaffy en Liesbeth List | dutch pop | 2005 | |
| 298 | 298 Une Belle Histoire (Een Mooi Verhaal) | Alderliefste | dutch pop | 2008 | |

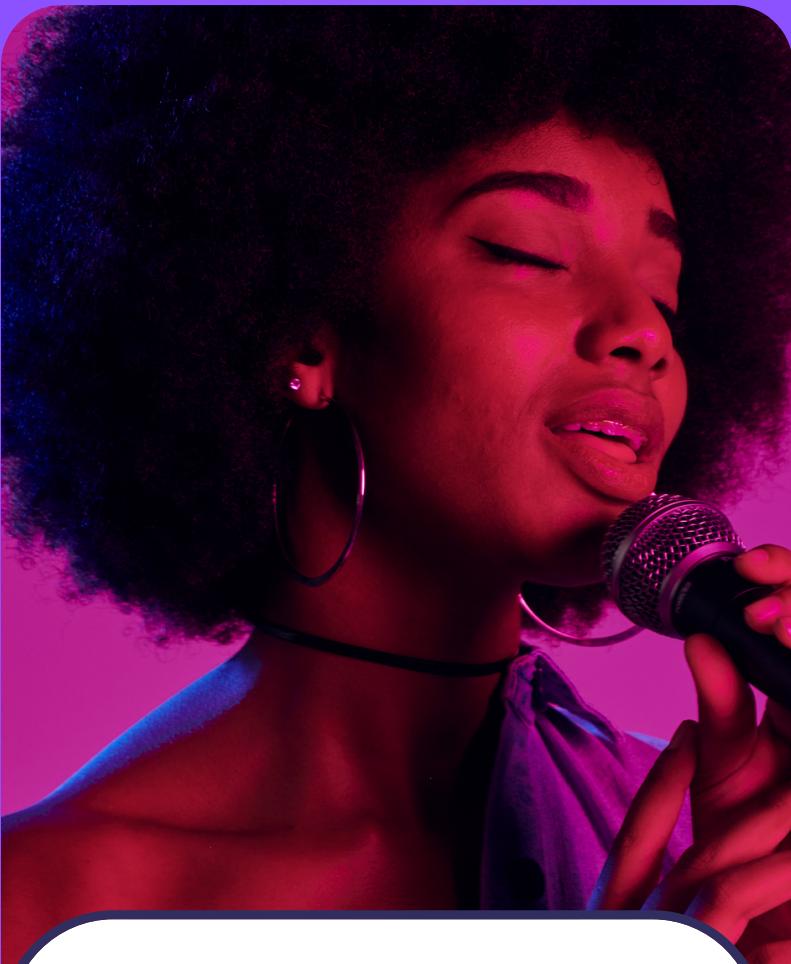
Text at the bottom of the table area: Showing 1 to 18 of 799 entries, 13 total columns

TOP SONGS

Showing 1 to 15 of 15 entries, 5 total columns

| | Index | Title | Artist | Top.Genre | Popularity |
|-----|-------|--------------------------|--------------|-----------|------------|
| 23 | 23 | Als De Morgen Is Gekomen | Jan Smit | dutch pop | 55 |
| 25 | 25 | Dichterbij Dan Ooit | BLØF | dutch pop | 16 |
| 40 | 40 | De Weg | Guus Meeuwis | dutch pop | 42 |
| 52 | 52 | Dansen Aan Zee | BLØF | dutch pop | 52 |
| 82 | 82 | Wêr Bisto | Twarres | dutch pop | 48 |
| 86 | 86 | Aanzoek Zonder Ringen | BLØF | dutch pop | 42 |
| 94 | 94 | Blauwe Ruis | BLØF | dutch pop | 41 |
| 114 | 114 | Bloed, Zweet En Tranen | Andre Hazes | dutch pop | 59 |
| 123 | 123 | Voltooid Verleden Tijd | IOS | dutch pop | 39 |
| 148 | 148 | Zij Maakt Het Verschil | De Poema's | dutch pop | 54 |
| 177 | 177 | Hier | BLØF | dutch pop | 40 |
| 181 | 181 | Father & Friend | Alain Clark | dutch pop | 35 |
| 205 | 205 | Pak Maar M'n Hand | Nick & Simon | dutch pop | 59 |
| 228 | 228 | Mijn Houten Hart | De Poema's | dutch pop | 45 |
| 256 | 256 | Rain Down on Me | Kane | dutch pop | 47 |

Recommendation on the basis of ARTIST



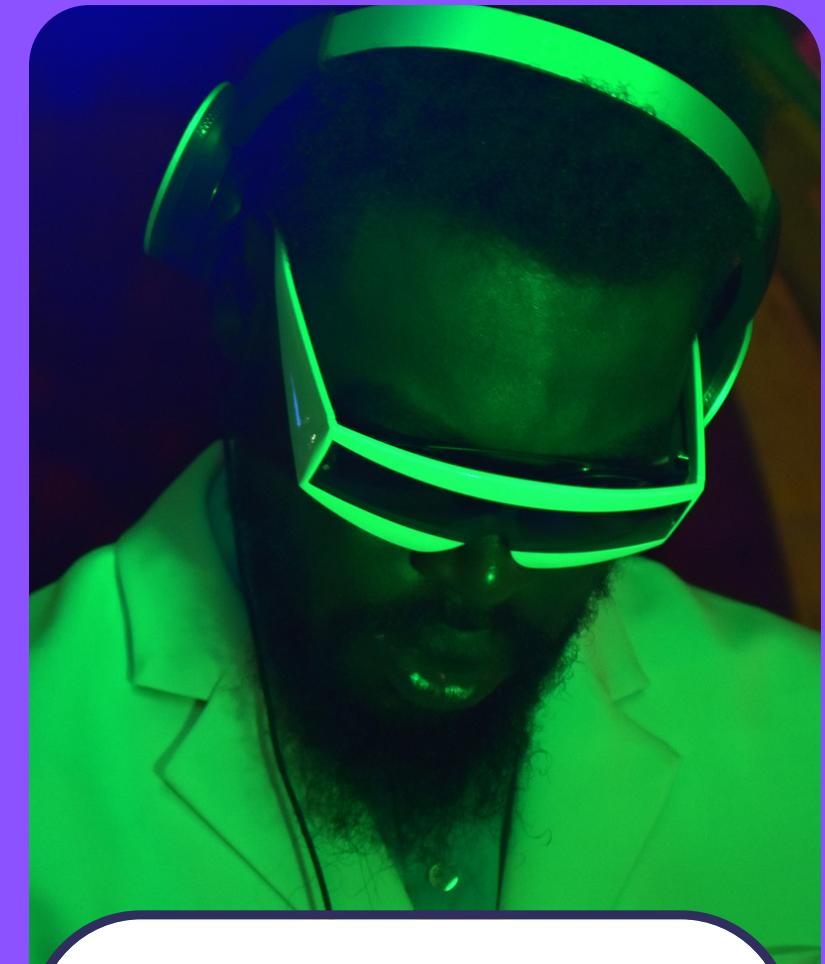
Reese Miller
Jazz Artist



Lars Peeters
Pop Artist



Benjamin Shah
Hiphop Artist



Avery Davis
EDM Artist

ARTIST MATRIX

R Untitled2* artist_matrix top_songs feature_matrix

Filter Cols: << 1 - 50 >>

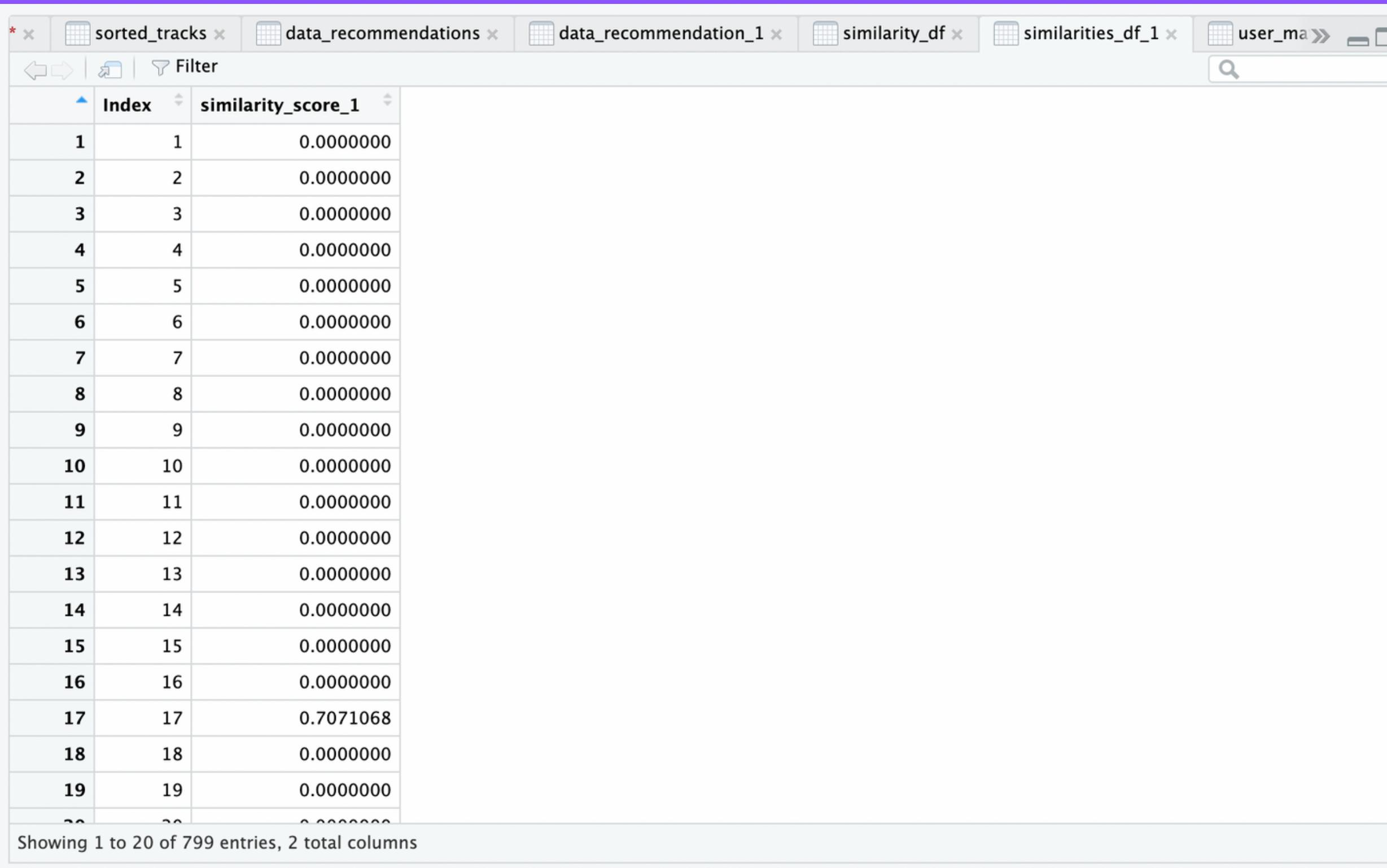
| | Norah Jones | Deep Purple | Gorillaz | Foo Fighters | Bruce Springsteen | City To City | Maroon 5 | Muse | The Killers | Eminem | Elvis Presley | The White Stripes |
|----|-------------|-------------|----------|--------------|-------------------|--------------|----------|-------|-------------|--------|---------------|-------------------|
| 1 | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 2 | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 3 | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 4 | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 5 | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 6 | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 7 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 8 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE | FALSE |
| 9 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE | FALSE |
| 10 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE | FALSE |
| 11 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE |
| 12 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE |
| 13 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 14 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 15 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 16 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 17 | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |

Showing 1 to 17 of 799 entries, 396 total columns

REAL MATRIX

| | Norah Jones | Deep Purple | Gorillaz | Foo Fighters | Bruce Springsteen | City To City | Maroon 5 | Muse | The Killers | Eminem | Elvis Presley | The White Stripes | De Dijk | Ten Years After | Arctic Monkeys |
|----|-------------|-------------|----------|--------------|-------------------|--------------|----------|------|-------------|--------|---------------|-------------------|---------|-----------------|----------------|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

SIMILARITY DATAFRAME CONTAINING SIMILARITY SCORES OF REAL MATRIX AND USER MATRIX

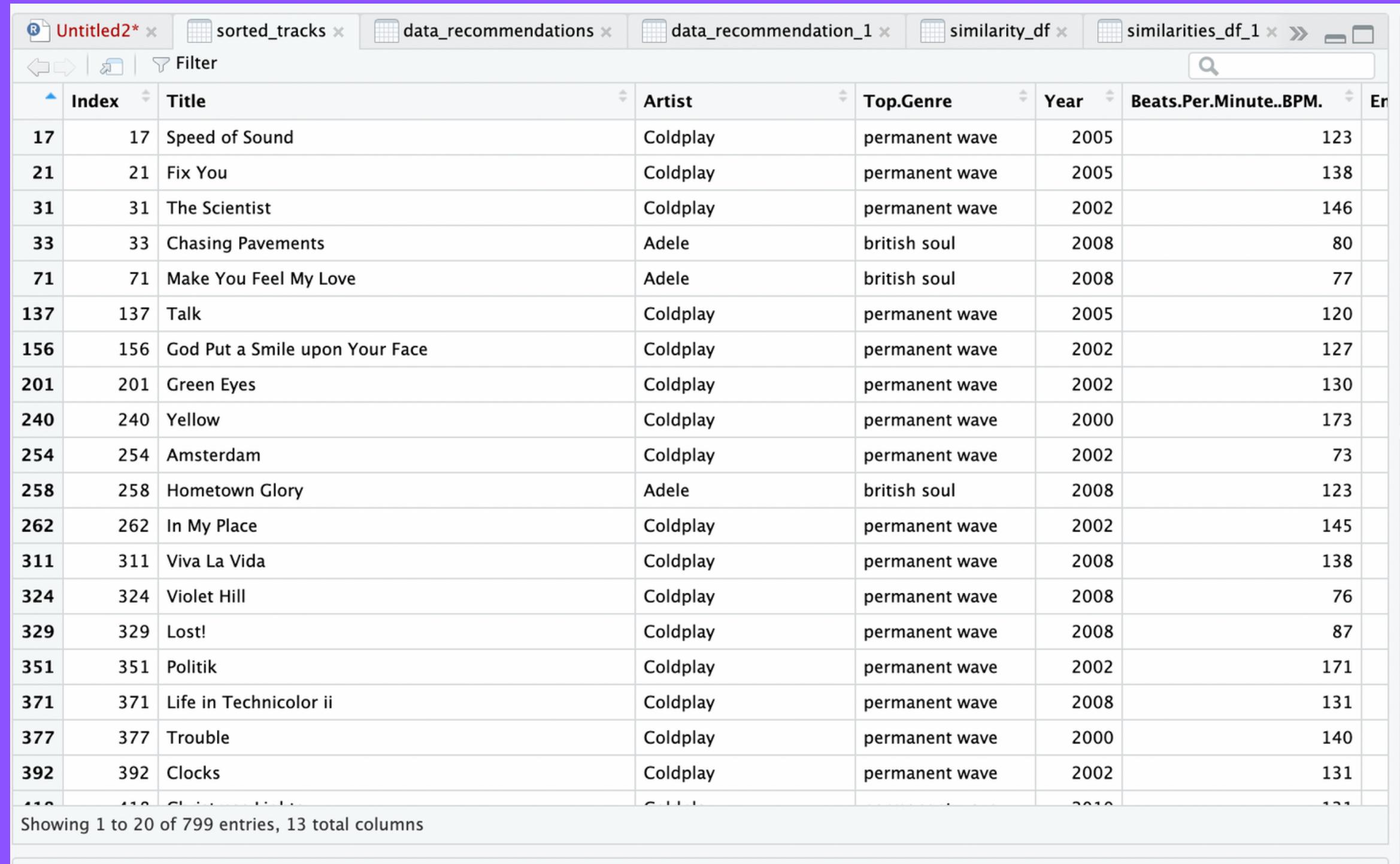


A screenshot of a Jupyter Notebook interface showing a DataFrame titled "similarity_score_1". The DataFrame has two columns: "Index" and "similarity_score_1". The "Index" column contains integers from 1 to 20. The "similarity_score_1" column contains floating-point numbers, with the value 0.7071068 appearing at index 17. The notebook also shows tabs for other data frames like "sorted_tracks", "data_recommendations", "data_recommendation_1", "similarities_df", "similarities_df_1", and "user_ma".

| Index | similarity_score_1 |
|-------|--------------------|
| 1 | 0.0000000 |
| 2 | 0.0000000 |
| 3 | 0.0000000 |
| 4 | 0.0000000 |
| 5 | 0.0000000 |
| 6 | 0.0000000 |
| 7 | 0.0000000 |
| 8 | 0.0000000 |
| 9 | 0.0000000 |
| 10 | 0.0000000 |
| 11 | 0.0000000 |
| 12 | 0.0000000 |
| 13 | 0.0000000 |
| 14 | 0.0000000 |
| 15 | 0.0000000 |
| 16 | 0.0000000 |
| 17 | 0.7071068 |
| 18 | 0.0000000 |
| 19 | 0.0000000 |
| 20 | 0.0000000 |

Showing 1 to 20 of 799 entries, 2 total columns

SORTED TRACKS

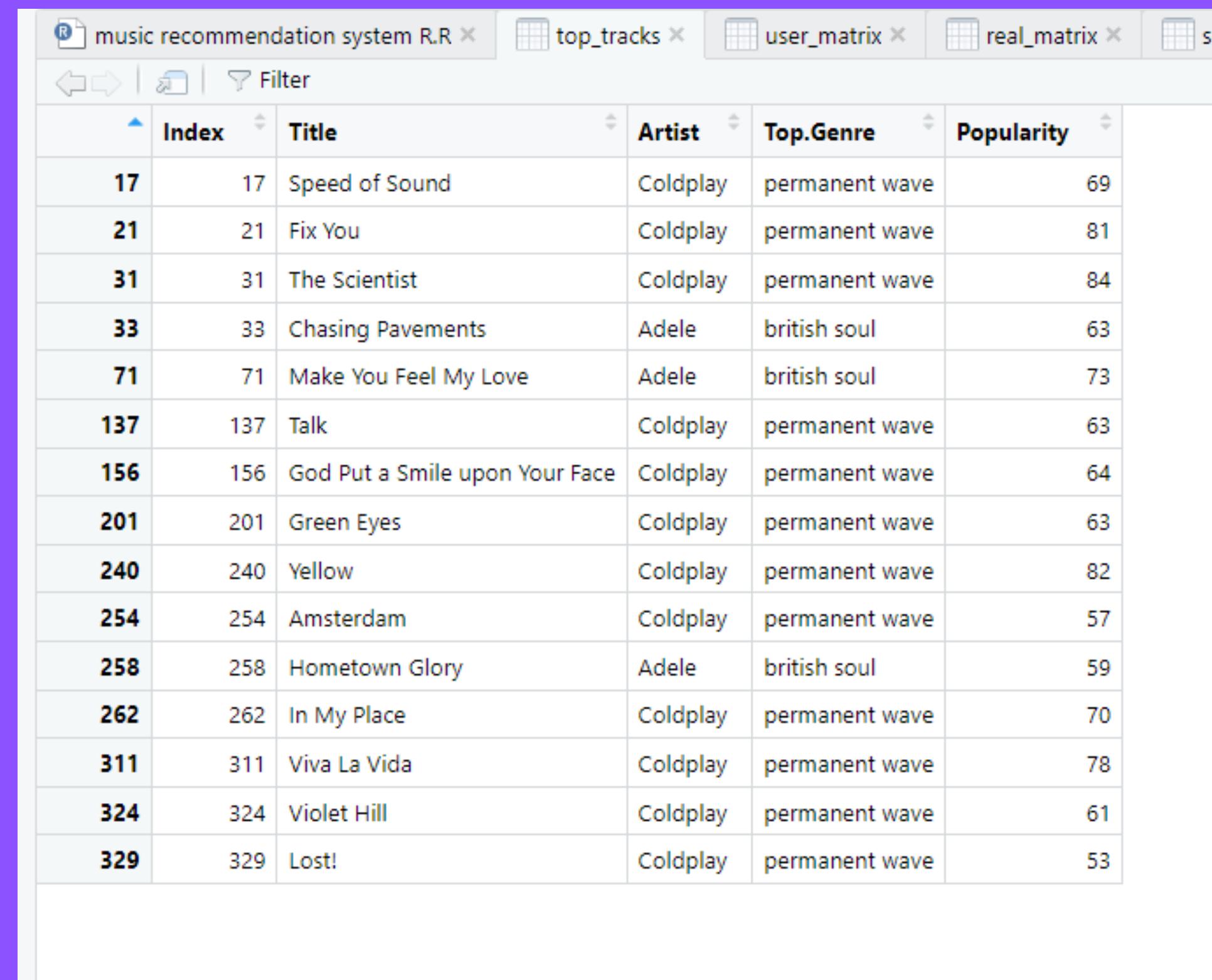


The screenshot shows a data frame titled "sorted_tracks" in an RStudio session. The data frame contains 799 rows and 13 columns. The columns are: Index, Title, Artist, Top.Genre, Year, Beats.Per.Minute..BPM., Energy, Danceability, Key, Acousticness, Duration..ms, and Explicit. The data is sorted by the "Index" column, which lists various song titles. The most frequent artist is Coldplay, appearing in nearly all entries. The genres are predominantly "permanent wave" and "british soul". The years span from 2000 to 2008. The BPM values range from 73 to 173. The energy levels are generally high, with most songs having an energy level above 0.6. The danceability scores are mostly between 0.3 and 0.6. The acousticness scores are mostly low, around 0.1. The duration of the songs varies significantly, with some being very short and others quite long.

| | Index | Title | Artist | Top.Genre | Year | Beats.Per.Minute..BPM. | Energy | Danceability | Key | Acousticness | Duration..ms | Explicit |
|-----|-------|--------------------------------|----------|----------------|------|------------------------|--------|--------------|-----|--------------|--------------|----------|
| 17 | 17 | Speed of Sound | Coldplay | permanent wave | 2005 | 123 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 21 | 21 | Fix You | Coldplay | permanent wave | 2005 | 138 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 31 | 31 | The Scientist | Coldplay | permanent wave | 2002 | 146 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 33 | 33 | Chasing Pavements | Adele | british soul | 2008 | 80 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 71 | 71 | Make You Feel My Love | Adele | british soul | 2008 | 77 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 137 | 137 | Talk | Coldplay | permanent wave | 2005 | 120 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 156 | 156 | God Put a Smile upon Your Face | Coldplay | permanent wave | 2002 | 127 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 201 | 201 | Green Eyes | Coldplay | permanent wave | 2002 | 130 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 240 | 240 | Yellow | Coldplay | permanent wave | 2000 | 173 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 254 | 254 | Amsterdam | Coldplay | permanent wave | 2002 | 73 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 258 | 258 | Hometown Glory | Adele | british soul | 2008 | 123 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 262 | 262 | In My Place | Coldplay | permanent wave | 2002 | 145 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 311 | 311 | Viva La Vida | Coldplay | permanent wave | 2008 | 138 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 324 | 324 | Violet Hill | Coldplay | permanent wave | 2008 | 76 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 329 | 329 | Lost! | Coldplay | permanent wave | 2008 | 87 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 351 | 351 | Politik | Coldplay | permanent wave | 2002 | 171 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 371 | 371 | Life in Technicolor ii | Coldplay | permanent wave | 2008 | 131 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 377 | 377 | Trouble | Coldplay | permanent wave | 2000 | 140 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 392 | 392 | Clocks | Coldplay | permanent wave | 2002 | 131 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |
| 410 | 410 | Clocks | Coldplay | permanent wave | 2002 | 121 | 0.85 | 0.65 | 1 | 0.1 | 150000 | 0 |

Showing 1 to 20 of 799 entries, 13 total columns

TOP TRACKS



The screenshot shows a data frame titled "top_tracks" in an RStudio session. The session also includes tabs for "music recommendation system.R.R", "user_matrix", and "real_matrix". The "top_tracks" data frame has columns: Index, Title, Artist, Top.Genre, and Popularity. The data consists of 15 rows, all of which have "Coldplay" as the artist and "permanent wave" as the genre. The popularity scores range from 53 to 84.

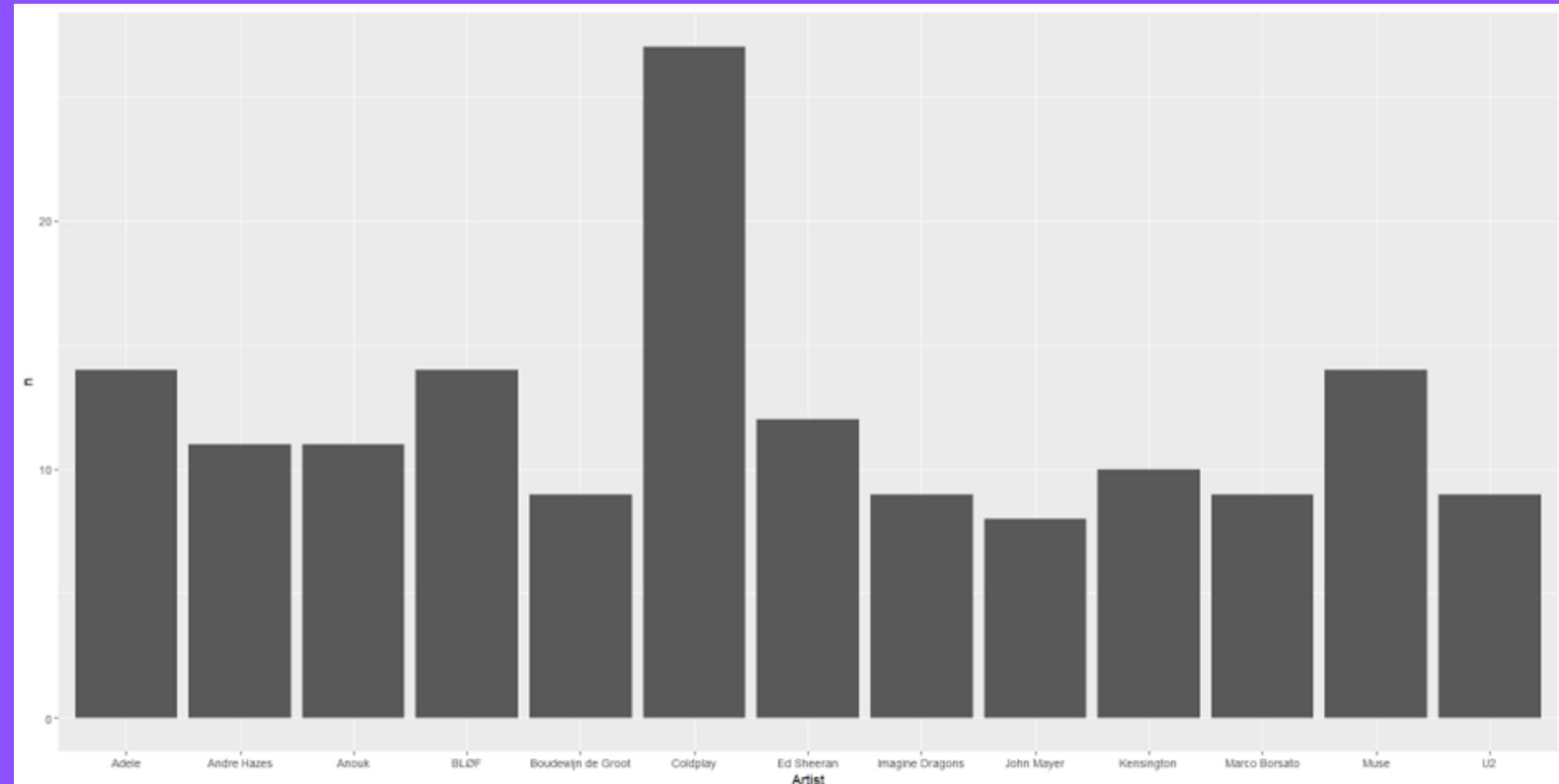
| | Index | Title | Artist | Top.Genre | Popularity |
|--|-------|------------------------------------|----------|----------------|------------|
| | 17 | 17 Speed of Sound | Coldplay | permanent wave | 69 |
| | 21 | 21 Fix You | Coldplay | permanent wave | 81 |
| | 31 | 31 The Scientist | Coldplay | permanent wave | 84 |
| | 33 | 33 Chasing Pavements | Adele | british soul | 63 |
| | 71 | 71 Make You Feel My Love | Adele | british soul | 73 |
| | 137 | 137 Talk | Coldplay | permanent wave | 63 |
| | 156 | 156 God Put a Smile upon Your Face | Coldplay | permanent wave | 64 |
| | 201 | 201 Green Eyes | Coldplay | permanent wave | 63 |
| | 240 | 240 Yellow | Coldplay | permanent wave | 82 |
| | 254 | 254 Amsterdam | Coldplay | permanent wave | 57 |
| | 258 | 258 Hometown Glory | Adele | british soul | 59 |
| | 262 | 262 In My Place | Coldplay | permanent wave | 70 |
| | 311 | 311 Viva La Vida | Coldplay | permanent wave | 78 |
| | 324 | 324 Violet Hill | Coldplay | permanent wave | 61 |
| | 329 | 329 Lost! | Coldplay | permanent wave | 53 |

VISUALISATION

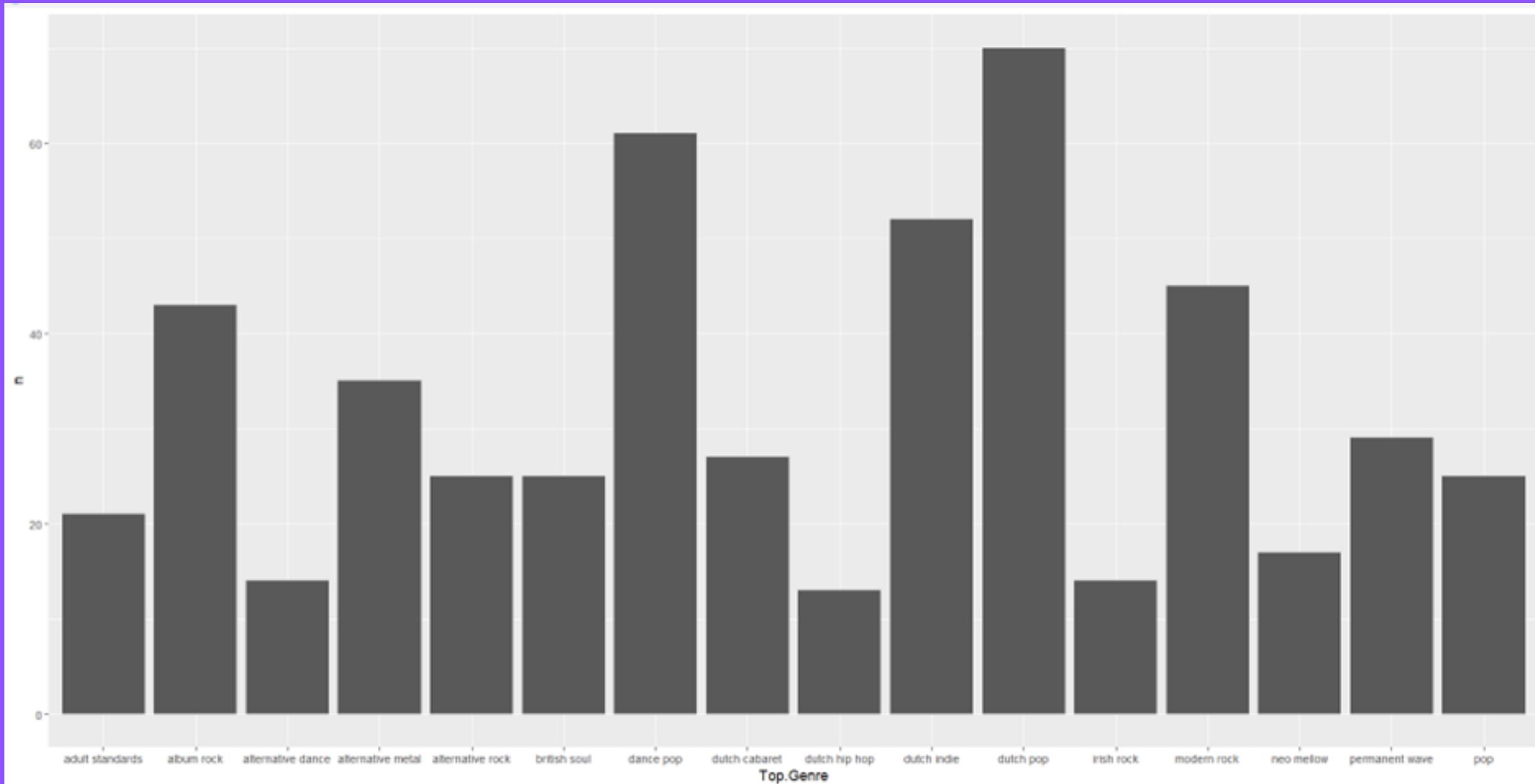
Data visualization was performed to gain a visual understanding of the music dataset.

- The count of songs by genre was visualized using a bar plot. The `count` function from the `dplyr` package was used to calculate the frequency of each genre, and the `ggplot2` package was used to create the bar plot.
- The count of songs by year was visualized using a bar plot. Again, the `count` function and `ggplot2` package were used for this analysis.
- The density plot of artist counts was created to show the distribution of song counts per artist. The `ggplot2` package was used to generate the plot.
- The relationship between the length of songs and their popularity was visualized using a scatter plot. The `ggplot2` package was used to create the scatter plot.

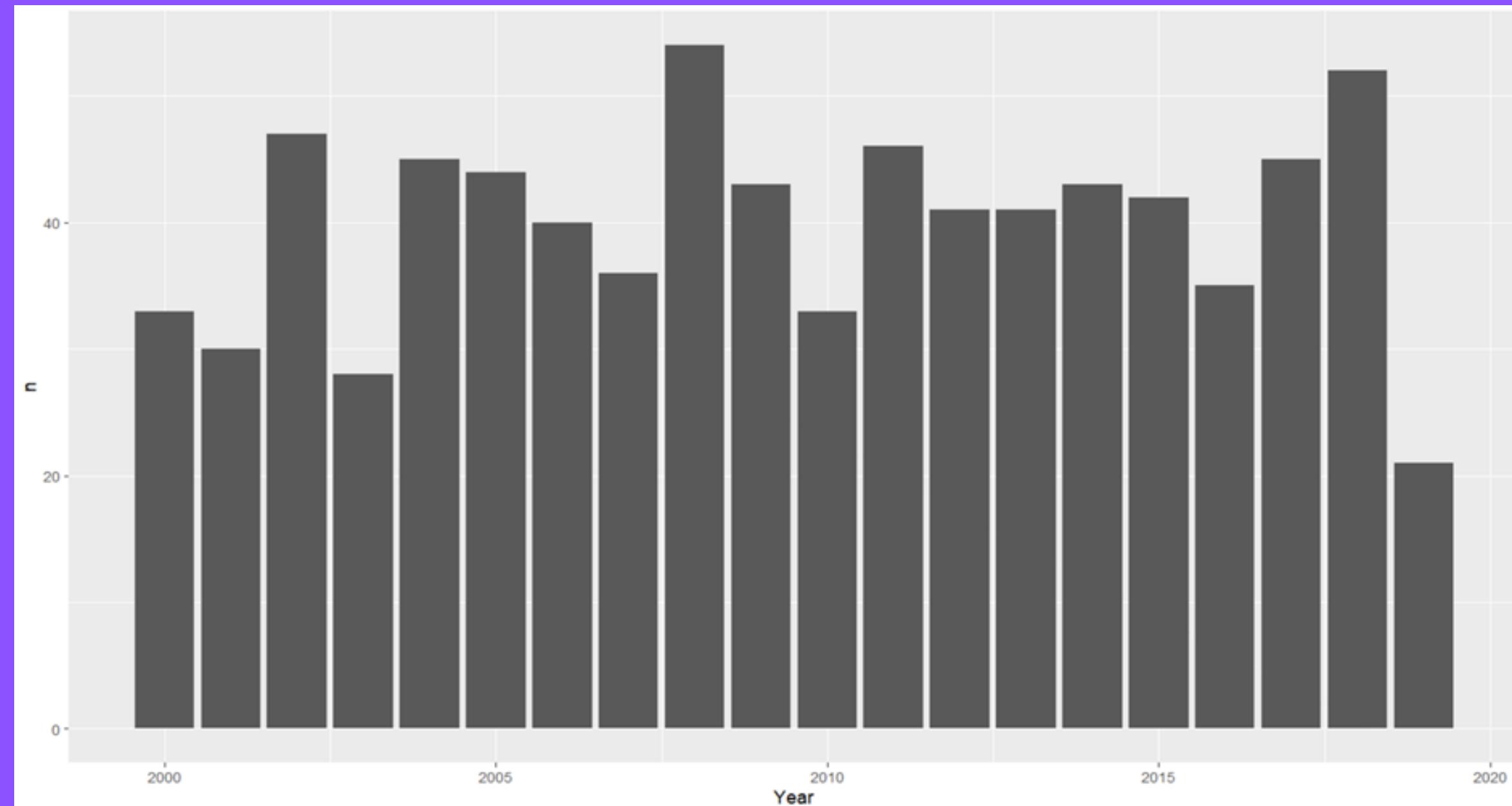
Artist Count Plot: The plot represents the count of songs for each artist. Artists with more than 6 songs are included. It gives an insight into the number of songs contributed by different artists.



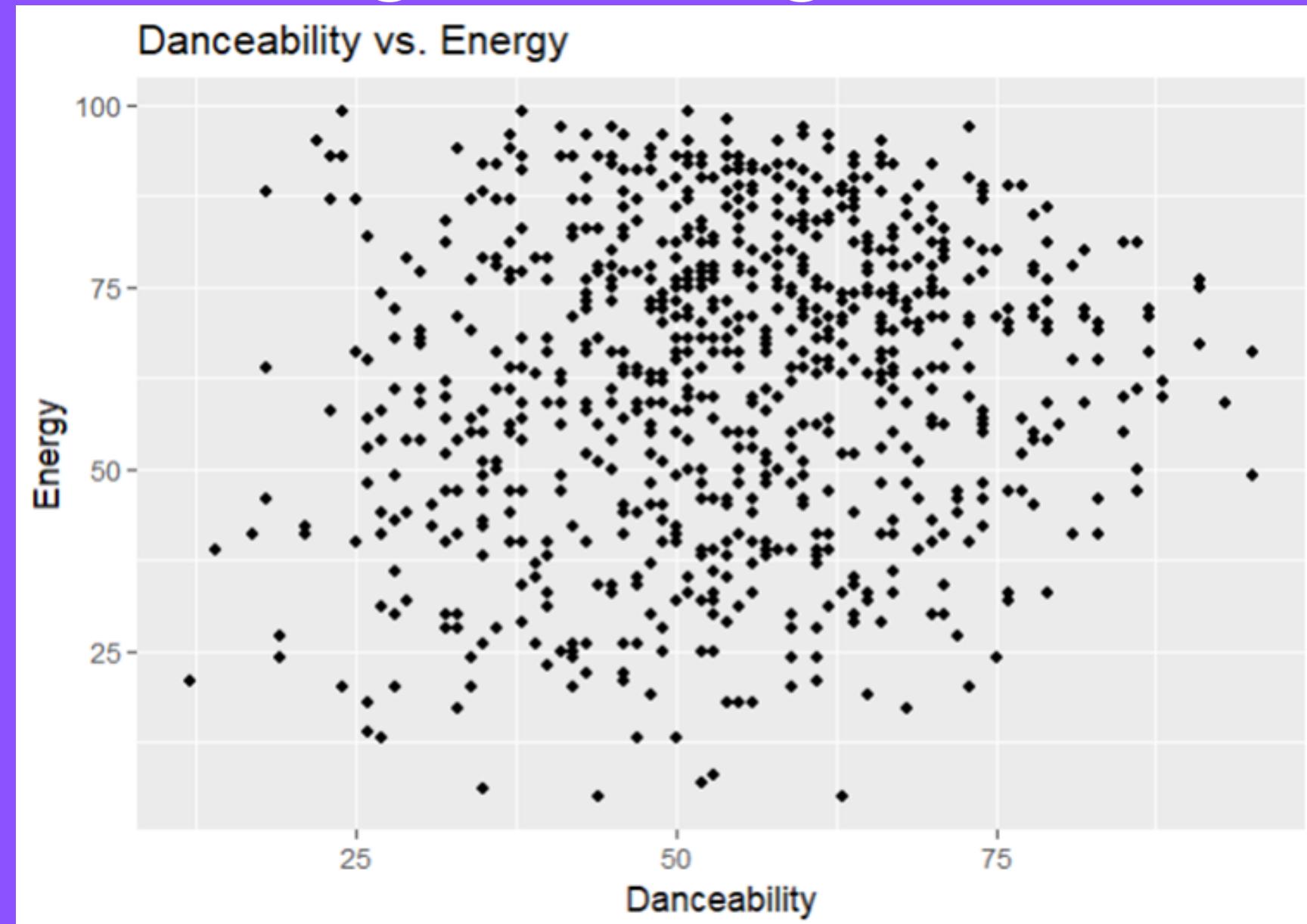
Genre Count Plot: The plot shows the count of songs for each top genre. Genres with more than 10 songs are included. It provides an overview of the distribution of songs across different genres.



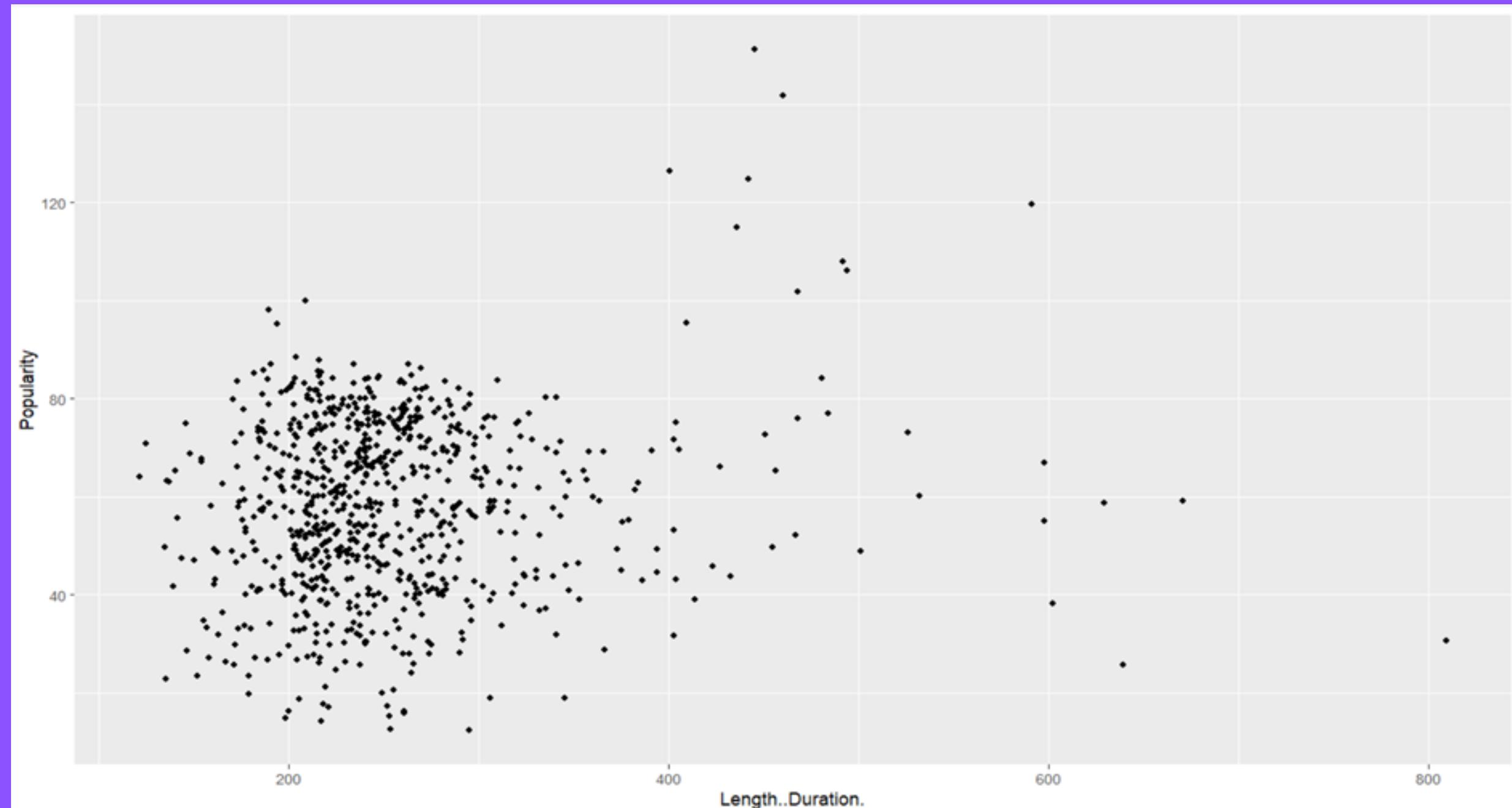
Year Count Plot: This plot displays the count of songs for each year. It helps visualize the distribution of songs over time, indicating which years have a higher or lower number of releases.



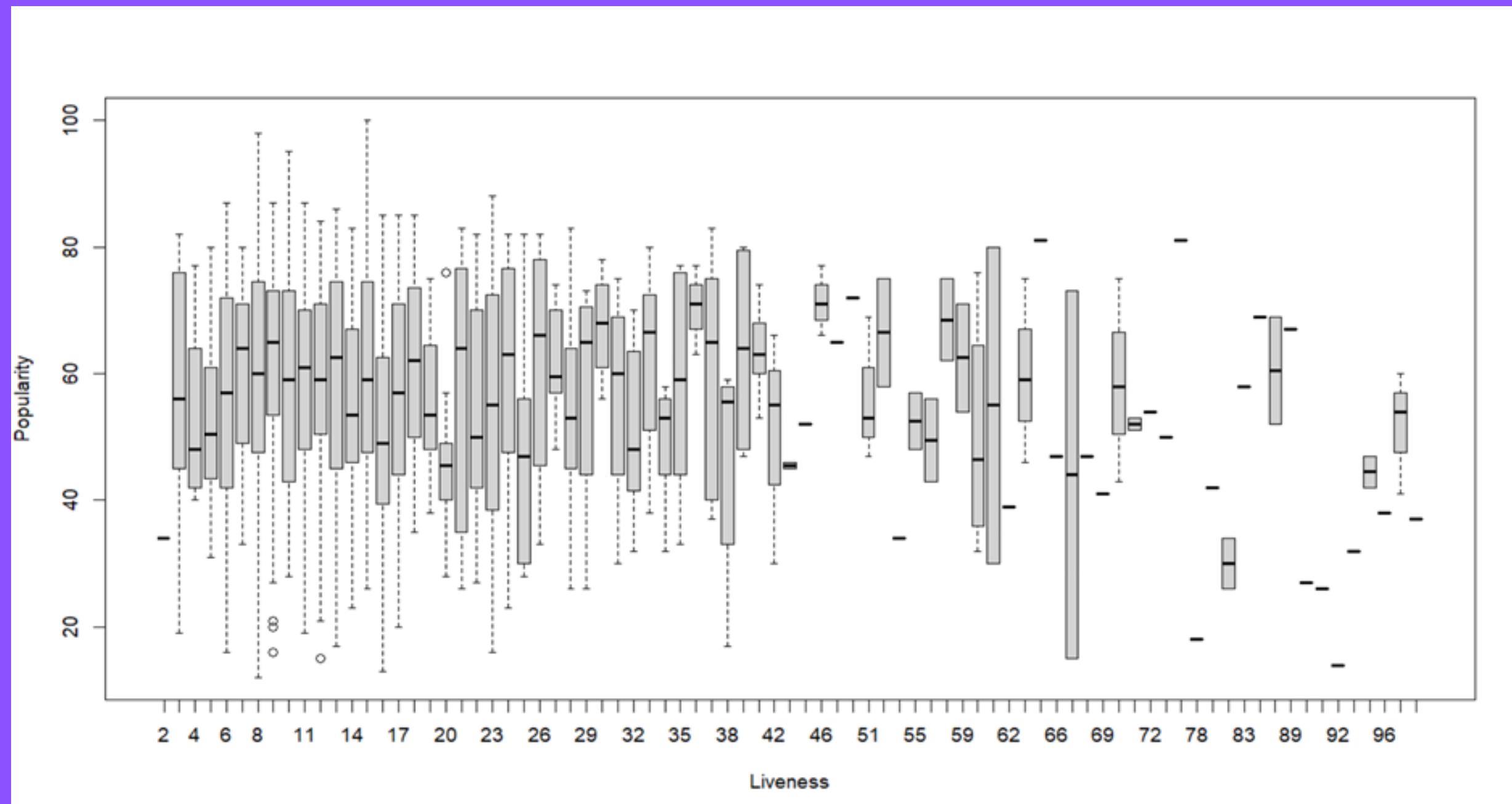
Danceability vs. Energy: This scatter plot displays the relationship between danceability and energy. It helps visualize if there is any correlation between these attributes and whether songs with higher danceability tend to have higher energy levels.



Length vs. Popularity: The jitter plot shows the relationship between song length and popularity. It gives a scattered representation of how the duration of a song relates to its popularity.



Popularity vs. Liveness (Boxplot): The boxplot shows the distribution of popularity across different levels of liveness. It helps compare the median, quartiles, and potential outliers in popularity for different liveness categories.

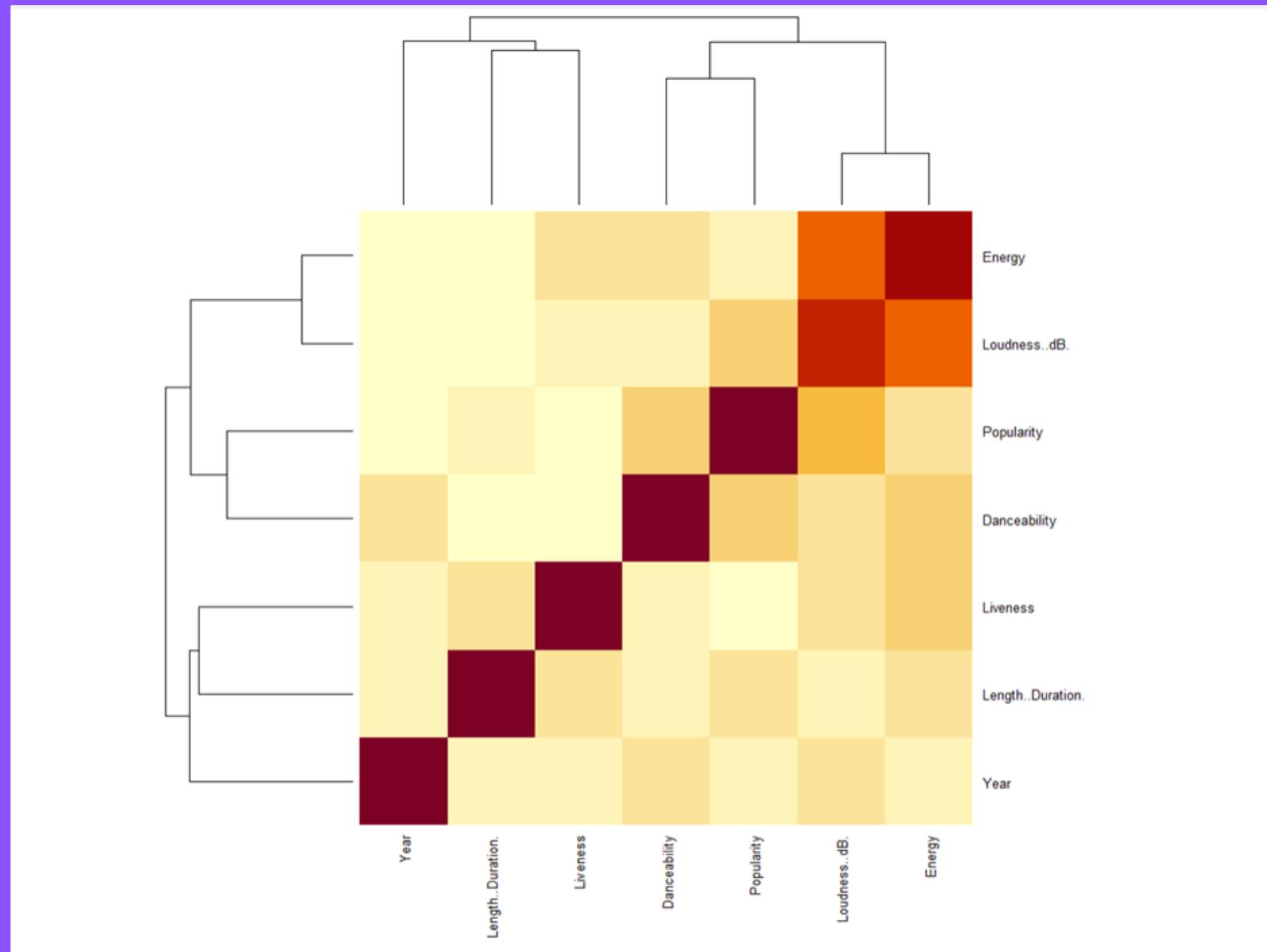


CORRELATION ANALYSIS

Correlation analysis was performed to explore the relationships between different numerical attributes of the songs.

- The `cor` function was used to calculate the correlation matrix between the attributes "Year," "Energy," "Danceability," "Loudness..dB.," "Liveness," "Length..Duration.," and "Popularity."
- The correlation matrix was printed to examine the strength and direction of the correlations between attributes.

Heatmap of Correlation Matrix: The heatmap visually represents the correlation matrix using colors. It allows for a quick assessment of the strength and patterns of correlations between variables.



```

> correlation_matrix<-cor(music_data[,c("Year","Energy","Danceability","Loudness..dB.","Liveness","Length..Duration.", "Popularity")]
> print(correlation_matrix)

          Year      Energy Danceability Loudness..dB.    Liveness Length..Duration. Popularity
Year  1.00000000 -0.065945349  0.05240345 -0.02153571 -0.02926872 -0.043848059 -0.039250432
Energy -0.06594535  1.000000000  0.14734541   0.72238675  0.16652483  0.005369493  0.119782470
Danceability  0.05240345  0.147345415  1.00000000  0.04883080 -0.09015558 -0.098747655  0.216114542
Loudness..dB. -0.02153571  0.722386750  0.04883080  1.00000000  0.05688718 -0.045013724  0.299709702
Liveness     -0.02926872  0.166524826 -0.09015558   0.05688718  1.00000000  0.026290429 -0.113455882
Length..Duration. -0.04384806  0.005369493 -0.09874766 -0.04501372   0.02629043  1.000000000 -0.002756057
Popularity   -0.03925043  0.119782470  0.21611454   0.29970970 -0.11345588 -0.002756057  1.000000000
>

```

| VARIABLE NAME | DEPENDENT / INDEPENDENT | DEPENDENT VARIABLES |
|-------------------|-------------------------|--------------------------------------------------|
| Year | INDEPENDENT | NIL |
| Energy | DEPENDENT | Loudness(0.72), Popularity(0.12) |
| Danceability | DEPENDENT | Popularity(0.22) |
| Loudness..dB. | DEPENDENT | Energy(0.72) |
| Liveness | INDEPENDENT | NIL |
| Length..Duration. | INDEPENDENT | NIL |
| Popularity | DEPENDENT | Energy(0.12), Danceability(0.22), Loudness(0.30) |

Conclusion



This project successfully analyzed a music dataset and provided personalized music recommendations based on user preferences. By utilizing various data manipulation and visualization techniques, the project uncovered insights about the dataset, such as genre distribution, popularity of artists, and correlations between variables. The user-based and artist-based recommendation models enhanced the music listening experience by suggesting songs that aligned with the user's preferences. This project demonstrates the application of data analysis and recommendation systems in the music industry, providing valuable insights for music enthusiasts and industry professionals alike.