

Froeny Trocts



Abrochicurn



Spotify Tracks Analysis

CONTENTS

Project Overview

Problem Statement

Data Description

Data Preparation

Quantitative
Snapshot

Understanding
Data Spread

What's in the
Catalog?

Distribution of Key
Audio Features

Distribution of Key
Audio Features - II

What Drives
Popularity?

Audio Feature
Patterns in
Winning Tracks

Correlation
Between Every
Feature

Temporal Analysis

Top Artists by
Popularity, Energy,
and Danceability

Cluster
Suggestions

Energetic Tracks
Cluster

Chill Tracks Cluster

Podcasts Track
Cluster

Thank You!

1. Project Overview

Contents of the Project

1 Problem Statement

Identifying effective strategies for enhancing song recommendations and promotions on Spotify.

3 Data Cleaning and Preprocessing

Steps taken to ensure data quality for analysis.

5 Visual Analysis

- Univariate: Analysis focused on individual variables.
- Bivariate: Insights drawn from the relationships between two variables.

7 Conclusion

Summary of the study and its implications.

2 Dataset Overview and Description

An introduction to the datasets utilized in the analysis.

4 Quantitative Analysis

Summary of quantitative insights derived from the data.

6 Key Insights

Major findings from the analysis.

2. Problem Statement

Spotify aims to gain deeper insights into what makes songs resonate with listeners to enhance recommendations, guide music production, and optimize artist promotion strategies. The primary audience includes music directors and mixing engineers who seek to optimize song composition and production. Producers, record labels, and streaming strategists benefit from these insights to make informed creative and business decisions. The challenge involved performing a comprehensive Exploratory Data Analysis (EDA) on a provided tracks dataset to uncover actionable insights into music popularity and feature patterns that drive success in the streaming industry.



3. Data Description

Column Name	Descripti...
0 track_id	A unique identifier for the track on Spotify
1 track_name	The title of the song
2 artist_name	The name of the artist(s) who performed the song
3 year	The release year of the song
4 popularity	A measure of how popular a track is, ranging f...
5 artwork_url	A URL pointing to the album artwork for the tr...
6 album_name	The name of the album the track belongs to
7 acousticness	A confidence measure indicating whether the tr...
8 danceability	A measure of how suitable a track is for danci...
9 duration_ms	The duration of the track in milliseconds
10 energy	A perceptual measure of intensity and activity
11 instrumentalness	Predicts whether a track contains no vocal con...
12 key	The key the track is in, represented as an int...
13 liveness	Detects the presence of an audience in the rec...
14 loudness	The overall loudness of a track in decibels (dB)
15 mode	Indicates the modality (major or minor) of a t...
16 speechiness	A measure detecting the presence of spoken wor...
17 tempo	The overall estimated tempo of a track in beat...
18 time_signature	An estimated overall time signature of a track
19 valence	A measure from -1.0 to 1.0 describing the musi...
20 track_url	A URL to the Spotify track
language	The detected language of the song's lyr...

```
.class 'pandas.core.frame.DataFrame'>
RangeIndex: 62317 entries, 0 to 62316
Data columns (total 22 columns):
#   Column              Non-Null Count  Dtype
---  -
0   track_id            62317 non-null  object
1   track_name          62317 non-null  object
2   artist_name         62317 non-null  object
3   year                62317 non-null  int64
4   popularity          62317 non-null  int64
5   artwork_url        62317 non-null  object
6   album_name         62317 non-null  object
7   acousticness        62317 non-null  float64
8   danceability        62317 non-null  float64
9   duration_ms         62317 non-null  float64
10  energy              62317 non-null  float64
11  instrumentalness     62317 non-null  float64
12  key                 62317 non-null  float64
13  liveness            62317 non-null  float64
14  loudness            62317 non-null  float64
15  mode                62317 non-null  float64
16  speechiness         62317 non-null  float64
17  tempo               62317 non-null  float64
18  time_signature      62317 non-null  float64
19  valence             62317 non-null  float64
20  track_url           62317 non-null  object
21  language            62317 non-null  object
dtypes: float64(13), int64(2), object(7)
memory usage: 10.5+ MB
```

The datasets used in the analysis include:

1

spotify_data_description.csv

2

spotify_tracks.csv

4. Data Preparation

Data Cleaning Steps

Initial Checks

No null values were found; 78 duplicate entries were removed.

Columns Dropped

Non-analytical columns (track_id, artwork_url, track_url) were removed to focus on core features.

Feature Engineering

- Created popularity level bins (Low, Moderate, Average, Massive) for categorical analysis.
- Converted duration_ms to duration_min for better interpretability.
- Applied a log transformation (duration_log) to handle extreme outliers in song duration.

Data Filtering

Removed physiologically implausible loudness values (e.g., -100000 dB) to ensure data quality.

Final Dataset Shape

62,191 tracks, 20 features.

5. QuantitativeSnapshot

Key Findings

1

Year Distribution

More songs were published in recent years, with fewer preceding 2000.

2

Popularity Score

The average song popularity score is low (~ 15), indicating a "long-tail" distribution.

3

Danceability and Energy

These scores are moderately high, suggesting a catalog leaning towards upbeat music.

4

Audio Feature Distributions

- Acousticness, Tempo, and Loudness show uniform spread except duration_min.

	count	mean	std	min	25%	50%	75%	max	range	c
year	62191.0	2014.415221	9.648624	1971.000000	2011.0000	2017.000000	2022.00000	2024.00000	53.000000	0.00
popularity	62191.0	15.364538	18.632701	0.000000	0.0000	7.000000	26.00000	93.00000	93.000000	1.21
acousticness	62191.0	0.363120	0.313107	0.000002	0.0673	0.286000	0.63300	0.99600	0.995998	0.86
danceability	62191.0	0.597654	0.182408	0.000000	0.4980	0.631000	0.73000	0.98600	0.986000	0.31
energy	62191.0	0.603274	0.243300	0.000020	0.4400	0.639000	0.80300	0.99900	0.998980	0.40
instrumentalness	62191.0	0.146700	0.306530	0.000000	0.0000	0.000025	0.01515	0.99900	0.999000	2.09
key	62191.0	5.105385	3.551422	0.000000	2.0000	5.000000	8.00000	11.00000	11.000000	0.70
liveness	62191.0	0.194836	0.169783	0.011900	0.0932	0.125000	0.24300	0.99800	0.986100	0.87
loudness	62191.0	-8.946058	5.327861	-45.920000	-10.7215	-7.505000	-5.45500	-0.00500	45.915000	-0.60
mode	62191.0	0.586934	0.492388	0.000000	0.0000	1.000000	1.00000	1.00000	1.000000	0.84
speechiness	62191.0	0.088343	0.112315	0.000000	0.0367	0.048900	0.08910	0.95900	0.959000	1.27
tempo	62191.0	117.992223	28.369054	0.000000	95.9505	117.995000	135.07650	239.97000	239.970000	0.24
time_signature	62191.0	3.859787	0.489434	0.000000	4.0000	4.000000	4.00000	5.00000	5.000000	0.13
valence	62191.0	0.496042	0.262453	0.000000	0.2930	0.508000	0.71000	0.99500	0.995000	0.53
duration_min	62191.0	4.045008	1.882724	0.167117	3.2060	3.939650	4.77200	76.35805	76.190933	0.47

6. Understanding Data Spread

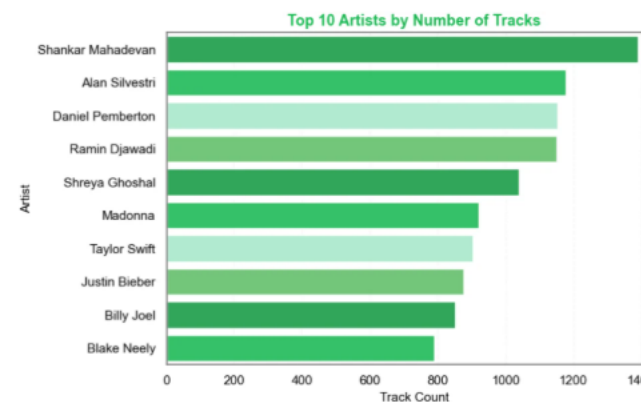
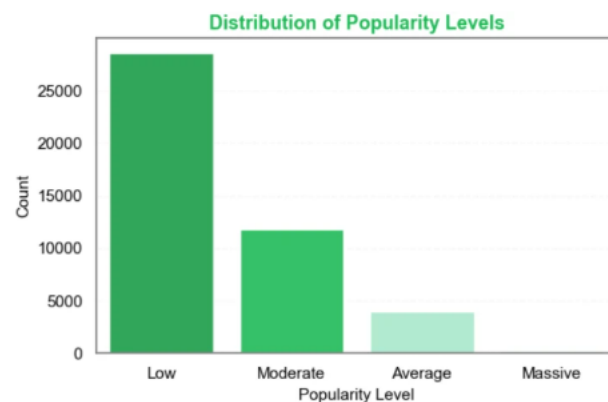
year	-1.519647	year	2.090412
popularity	1.230852	popularity	0.660541
acousticness	0.486151	acousticness	-1.137858
danceability	-0.763755	danceability	0.131169
energy	-0.537597	energy	-0.534778
instrumentalness	1.824831	instrumentalness	1.559414
key	0.057805	key	-1.265500
liveness	2.296321	liveness	5.829478
loudness	-1.824510	loudness	4.315808
mode	-0.353122	mode	-1.875365
speechiness	4.488898	speechiness	25.988183
tempo	0.366447	tempo	-0.168984
time_signature	-3.311762	time_signature	15.975065
valence	-0.121822	valence	-1.023577
duration_min	9.381201	duration_min	235.760268
dtype: float64		dtype: float64	



Skewness and Kurtosis

High positive skewness in instrumentality and duration_min confirms a right-skewed distribution, indicating most songs have vocals and standard lengths, with a few long/instrumental outliers.

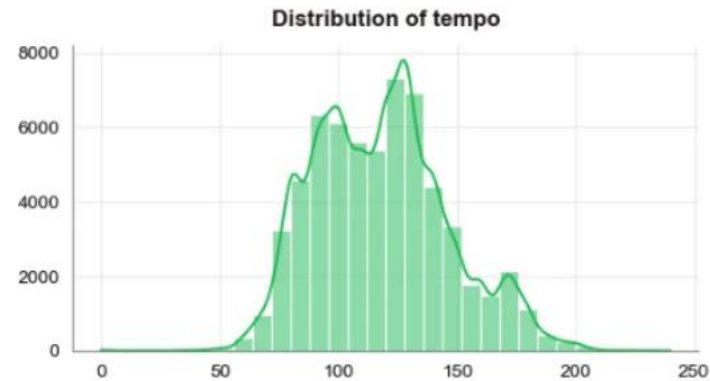
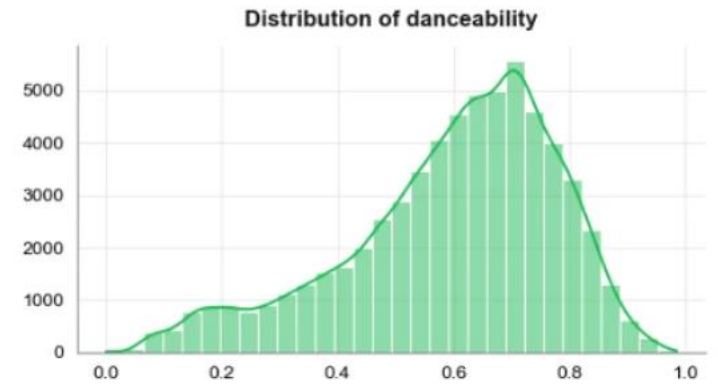
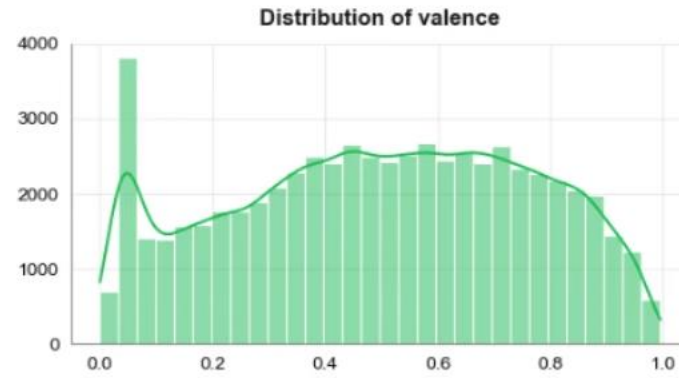
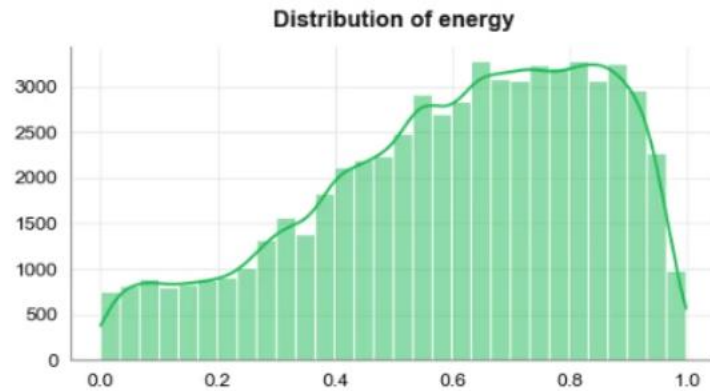
7. What's in the Catalog?



The majority of tracks fall into the "Low" popularity category, driven by social media trends.

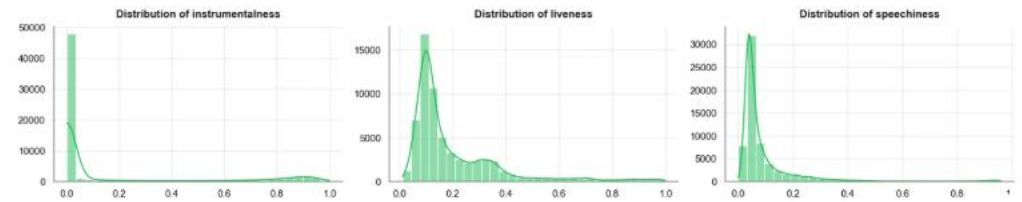
Certain artists, like Shankar Mahadevan and Alan Silvestri, appear frequently, indicating a focus on specific genres or film scores in the dataset.

8. Distribution of Key Audio Features



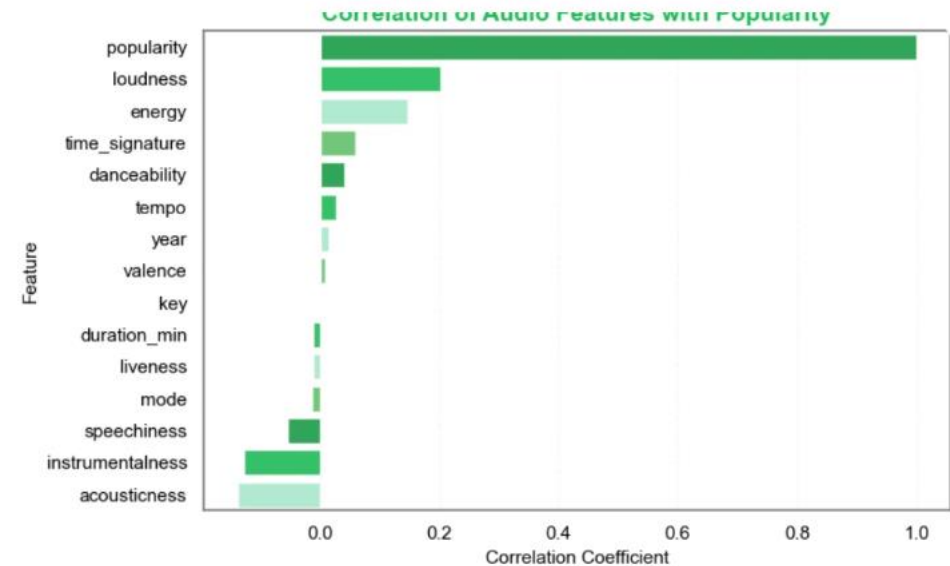
9. Distribution of Key Audio Features - II

The analysis reveals a strong preference for vocal-driven, studio-produced music, suggesting a shift in content acquisition strategies.



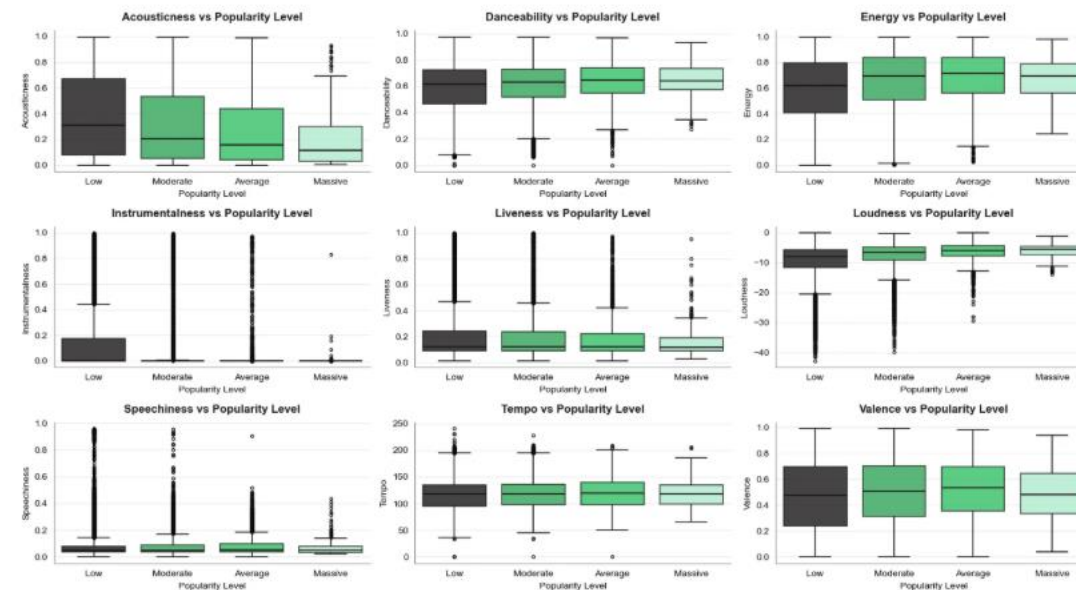
10. What Drives Popularity?

Loudness and Energy show the strongest positive correlation with popularity, while Acousticness and Instrumentalness exhibit negative correlations.



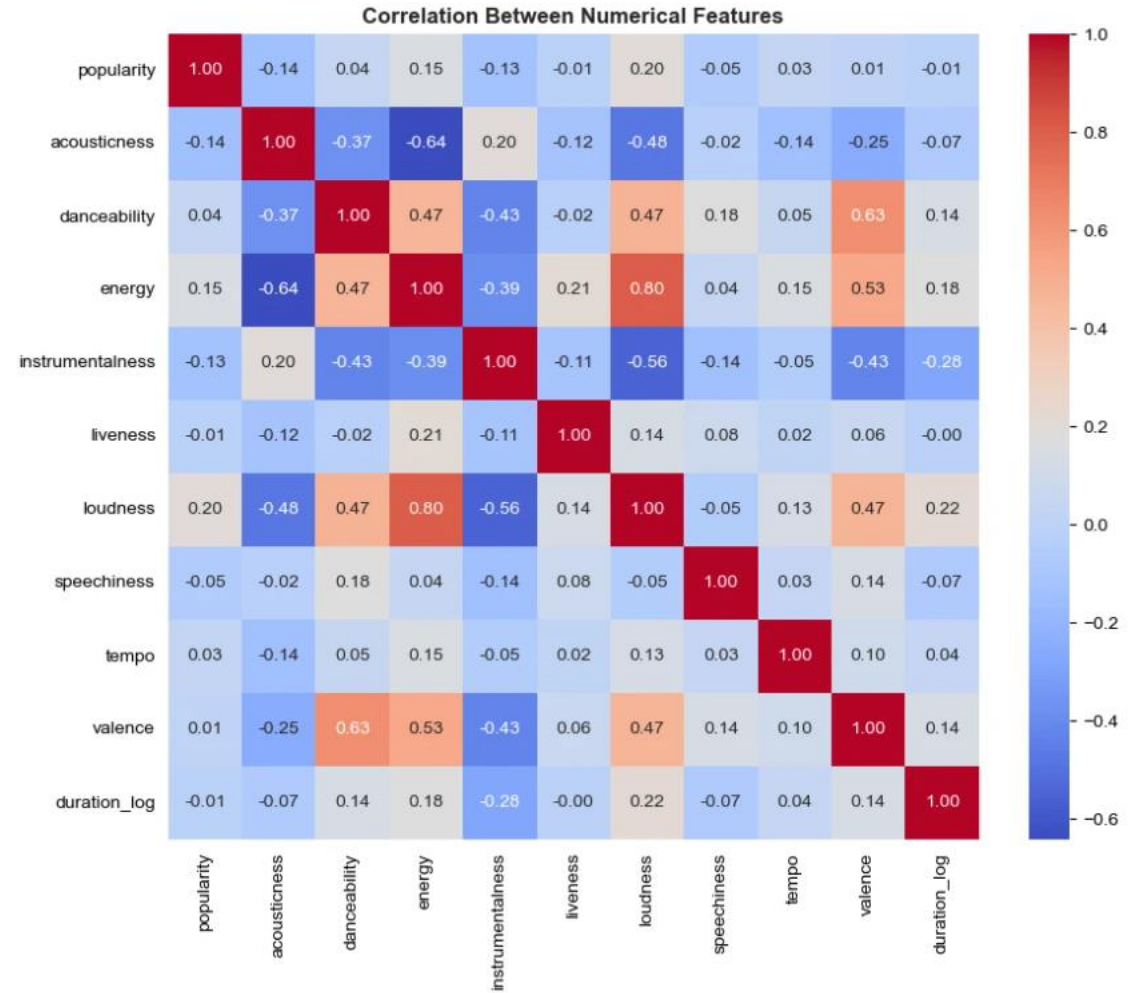
11. Audio Feature Patterns in Winning Tracks

Popular tracks generally show higher energy, danceability, and loudness, emphasizing trends toward upbeat sounds.



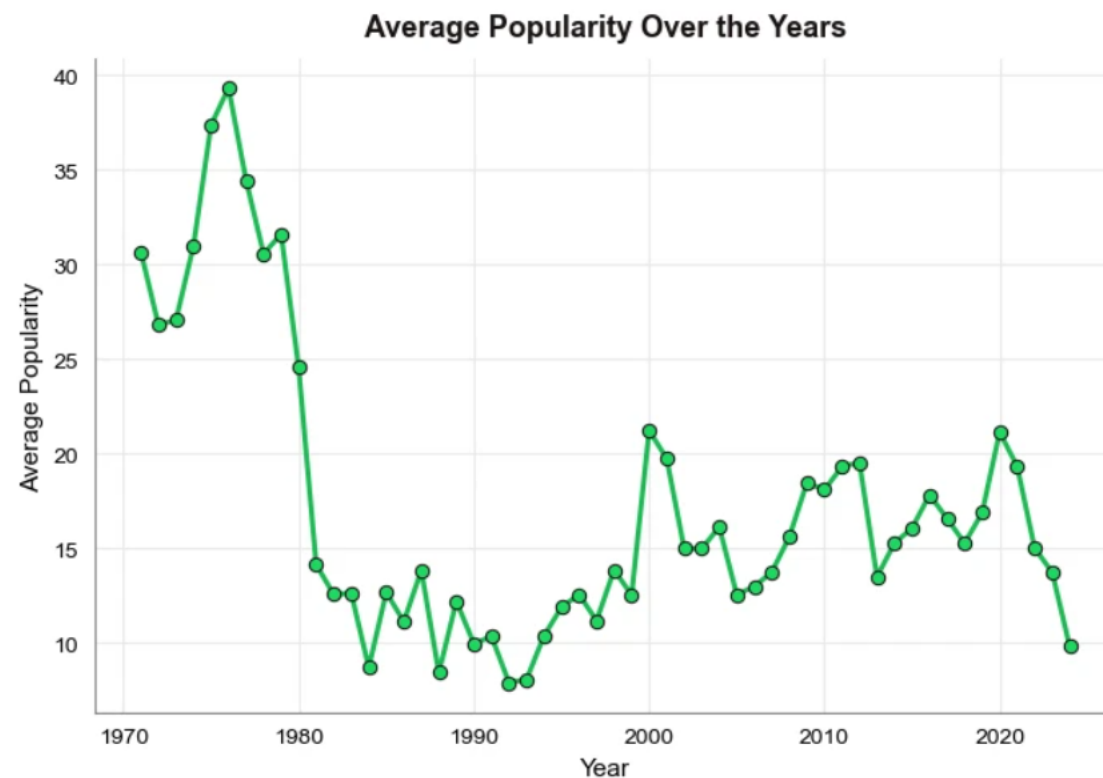
12. Correlation Between Every Feature

The correlation matrix provides insights into the relationships among all features.

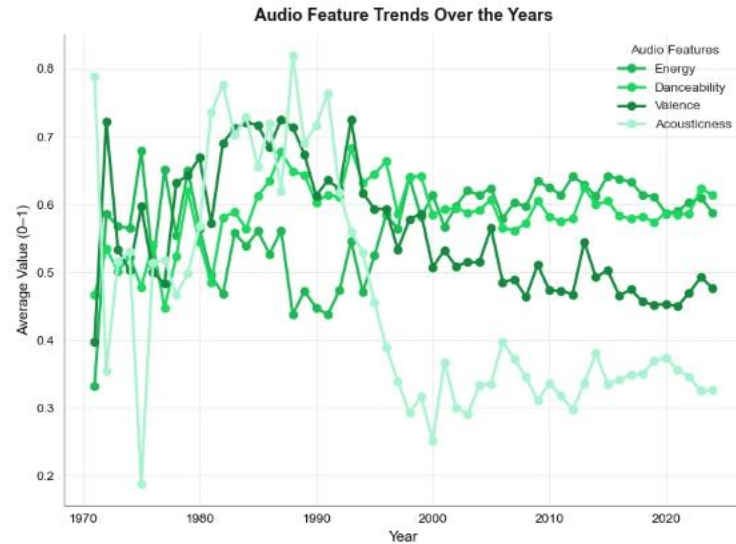


13. Temporal Analysis

Popularity Over the Years



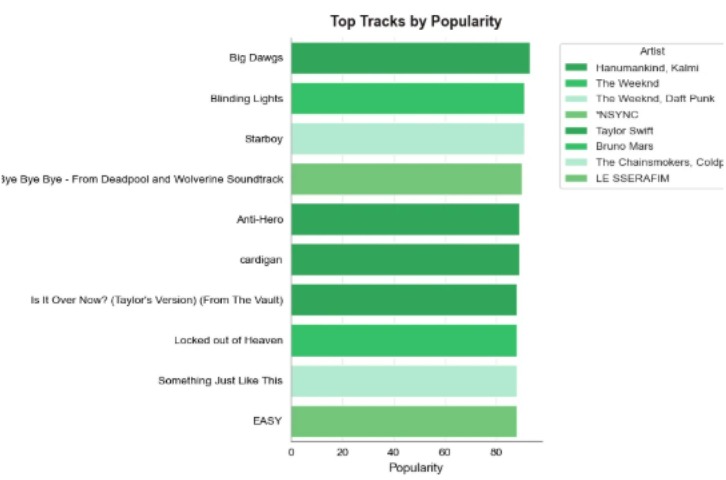
Trends in Audio Features Over Time



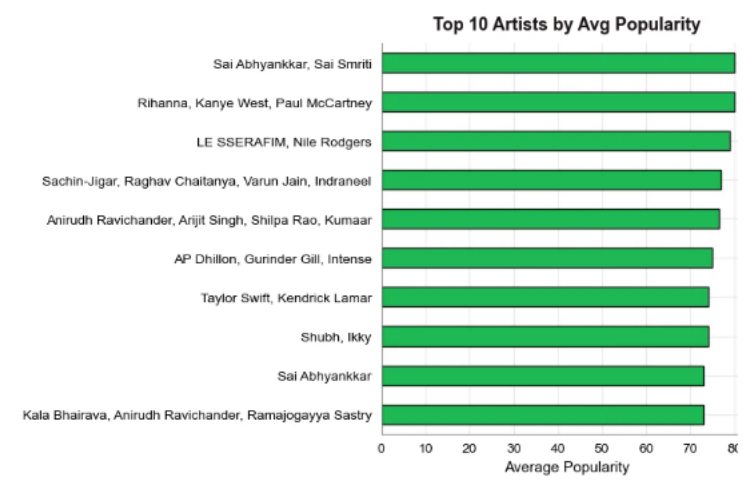
The analysis indicates that music popularity has generally decreased over the years.

Recent years have seen a slight increase in energy and danceability whereas valence and acousticness had decreased considerably over the time spread.

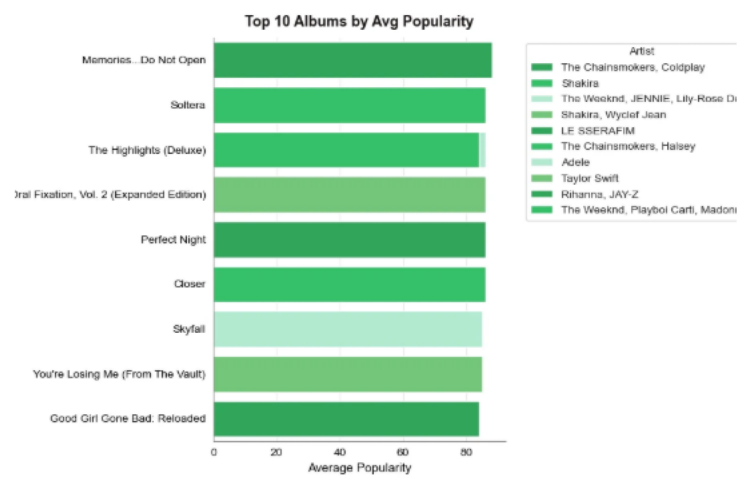
Trends in Categories.



Most popular tracks.



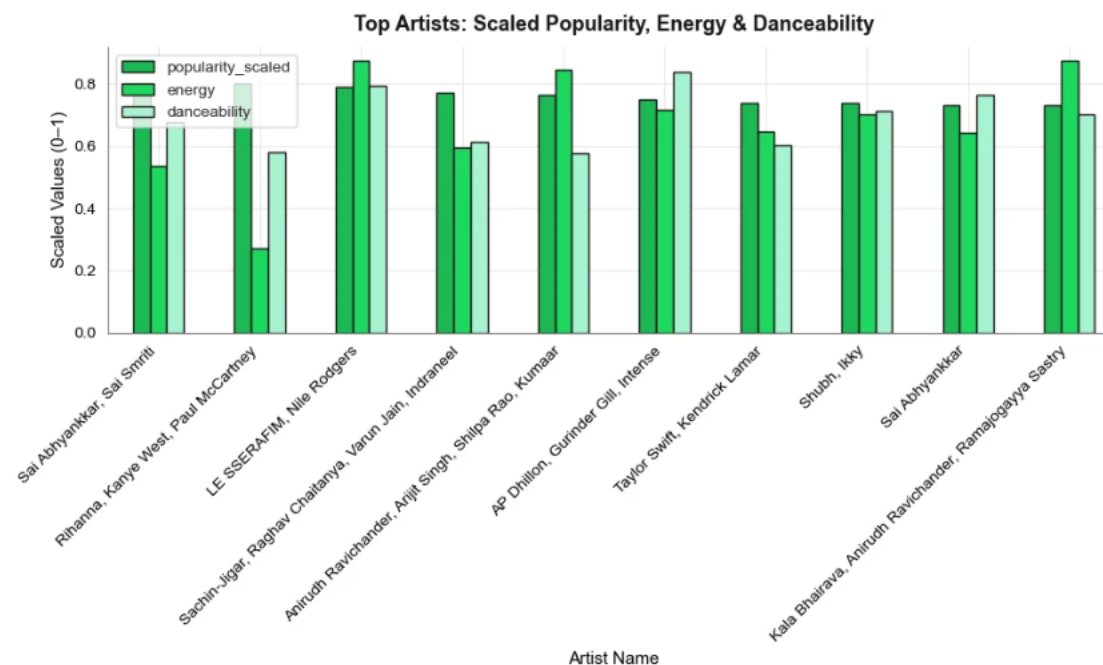
Most popular artists.



Most popular

14. Top Artists by Popularity, Energy, and Danceability

This chart identifies artists who consistently produce popular, energetic, and danceable tracks.



15. Cluster Suggestions

Cluster suggestions can be made by categorizing top songs from different genres to optimize user profiles and increase app reach.



16. Energetic Tracks Cluster

	track_name	artist_name	album_name	energy
33569	I Was Crazy About You	2PM	1:59PM	0.999
2232	Action With Aggressive	Yuvan Shankar Raja	Sandakozhi (Original Soundtrack)	0.999
33357	Tetris - Buzz & Flipswitch Remix	2PM, Buzz & Flipswitch	Tetris	0.999
5957	Skanda Title Theme	Thaman S	Skanda (Original Background Score)	0.998
29790	Rockstar - Rearranged Version, Live	SUPER JUNIOR	SUPER SHOW 6 - SUPER JUNIOR The 6th WORLD TOUR	0.998
34839	Oxytocin	Infinite	Oxytocin	0.998
34874	The Edge	Infinite	The Edge	0.998
9538	Dhan Bada Ye Janu	Munna Mohit, Sadhna Chauhan	Dhan Bada Ye Janu	0.998
43855	Pts.Of.Athrty (Crystal Method Remix) - LPU Rar...	Linkin Park, The Crystal Method	Hybrid Theory (20th Anniversary Edition)	0.998
16903	Dom Kudithe	Shankar Mahadevan, Chitra	Khatarnak	0.998
15317	Razia (Remix)	Pritam, Master Saleem, Ritu Pathak, Abhijit Va...	Thank You	0.997
16519	Gaon Ke Bhauji Jindabad	Ankit Tiwari	Gaon Ke Bhauji Jindabad	0.997
12781	Mado Kabaar Deb	Pradeep Kumar	Kusumi Chunariya	0.997
43500	Papercut - Live	Linkin Park	Live in Texas	0.997
55729	Tere Vaaste - Slowed n Reverbed	Bhau Mast Vajtay, Sachin-Jigar, Altamash Farid...	Open Stage Melodies, Vol. 23	0.997

17. Chill Tracks Cluster

	track_name	artist_name	album_name	acousticness
35964	Hawk Meditations	Blue	Indigo Soulstice	0.996
46851	Lullaby (Goodnight, My Angel)	Billy Joel, Lara Downes	Lullaby (Goodnight, My Angel)	0.996
58250	Chapters Of History	Daniel Alexis Pemberton	Creative Review	0.996
35451	Discrepency	Treasure	Discrepency	0.996
54557	Boon Bestowed: III. The Madonna - Life	Katie Madonna Lee, Cecily Terhune	Boon Bestowed	0.996
58044	Christmas Lullaby	Daniel Pemberton	Christmas Cheer	0.996
59570	Ramin Djawadi: Game of Thrones Theme in the St...	Ramin Djawadi, Claudio Ferrarini	Ramin Djawadi: Game of Thrones Theme in the St...	0.996
45927	Alan Silvestri: Believe (Arr. for flute by Cla...	Alan Silvestri, Claudio Ferrarini	Alan Silvestri: Believe (Arr. for flute by Cla...	0.996
57869	Alone In Exile	Daniel Alexis Pemberton	Revolution And Empire	0.996
54517	With Love I Survive	Madonna Z	With Love I Survive	0.996
57911	Toby And Julia Drink	Daniel Pemberton	The Haunted Airman (Starring Robert Pattinson,...	0.996
35963	Meditation's in Sea	Blue	Indigo Soulstice	0.996
52240	Zdrowaś Maryjo	Black Madonna Choir	Polish religious and patriotic songs and the m...	0.996
58135	Toby And Julia Drink	Daniel Pemberton	The Haunted Airman (Starring Robert Pattinson,...	0.996
10264	Dj Jangan Lupakan	Abdur Rahman	Dj Jangan Lupakan	0.995

18. Podcasts Track Cluster

	track_name	artist_name	album_name	speechiness
61653	Chapter 51 - Nowhere to Hide	Keri Beevis, Shakira Shute	Nowhere to Hide (Unabridged)	0.959
61568	Chapter 109 - Cottage by the Sea - A BRAND NEW...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.958
61300	Chapter 81 - Nowhere to Hide	Keri Beevis, Shakira Shute	Nowhere to Hide (Unabridged)	0.957
29371	Dear Diary : 2016.07.29	ATEEZ	ZERO : FEVER Part.1	0.957
61434	Chapter 68 - Nowhere to Hide	Keri Beevis, Shakira Shute	Nowhere to Hide (Unabridged)	0.956
61556	Chapter 54 - Cottage by the Sea - A BRAND NEW ...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.956
61531	Chapter 156 - Cottage by the Sea - A BRAND NEW...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.956
60212	A Tragedy	Shakira Searle	Poetry of Sir Arthur Conan Doyle	0.955
61349	Chapter 32 - Nowhere to Hide	Keri Beevis, Shakira Shute	Nowhere to Hide (Unabridged)	0.955
61486	Chapter 112 - Cottage by the Sea - A BRAND NEW...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.955
61285	Chapter 3 - Nowhere to Hide	Keri Beevis, Shakira Shute	Nowhere to Hide (Unabridged)	0.955
61562	Chapter 131 - Cottage by the Sea - A BRAND NEW...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.955
61506	Chapter 139 - Cottage by the Sea - A BRAND NEW...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.955
46769	Performing in Cuba as Precursor to Russia	Billy Joel	A Matter Of Trust - The Bridge To Russia (Comm...	0.955
61569	Chapter 87 - Cottage by the Sea - A BRAND NEW ...	Keri Beevis, Shakira Shute	Cottage by the Sea [A BRAND NEW pulse-pounding...	0.955

Thank You!

by Debanjan Sarkar,

Madhurima Das