Understanding Spotify Track Popularity

Akhil Ajith

Research Questions

1   Is track popularity evenly distributed, or concentrated among a small number of songs?

2   How are artist popularity and follower count related to track popularity?

3   Do explicit tracks differ in popularity compared to non-explicit tracks?

4   Are certain genres, album types, or release periods associated with higher popularity?

5   Can simple rule-based logic be used to identify potential hit songs?

Data

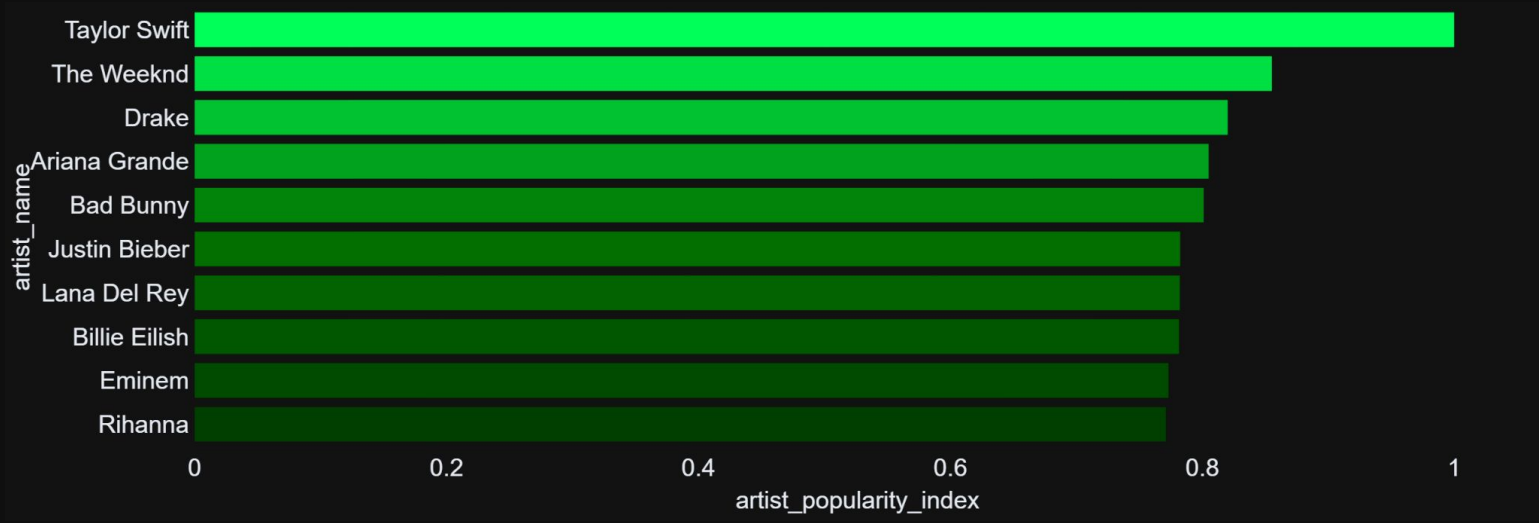| #	Feature Type	Columns
1	Track-level	Popularity, duration, explicit flag
2	Artist-level	Artist popularity, followers
3	Categorical	Genre, album type
4	Temporal	Release year

Data Cleaning and Wrangling Strategy

Steps:

- Cleaned and standardized text fields, and handled missing values

- Fixed data types and extracted the release year from album dates

- Processed genre information and assigned one main genre per track

- Removed unrealistic or invalid entries (e.g., impossible popularity or duration values)

- Removed duplicate tracks and combined repeated versions of the same song

- Created additional features such as track length in minutes

- Performed final checks to ensure the dataset was complete and consistent
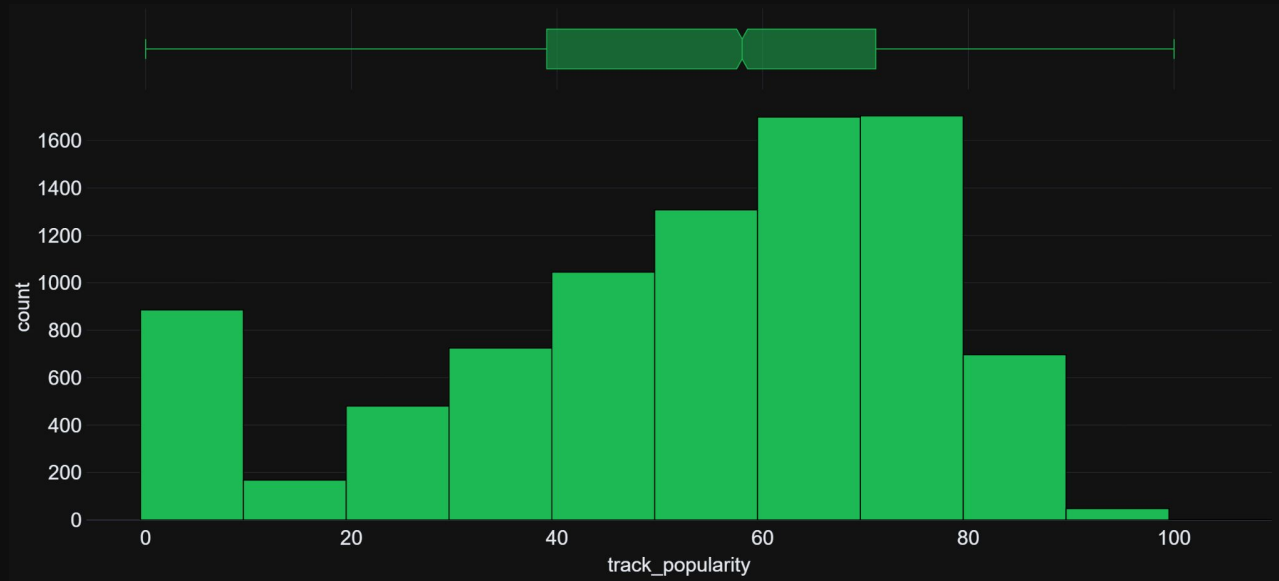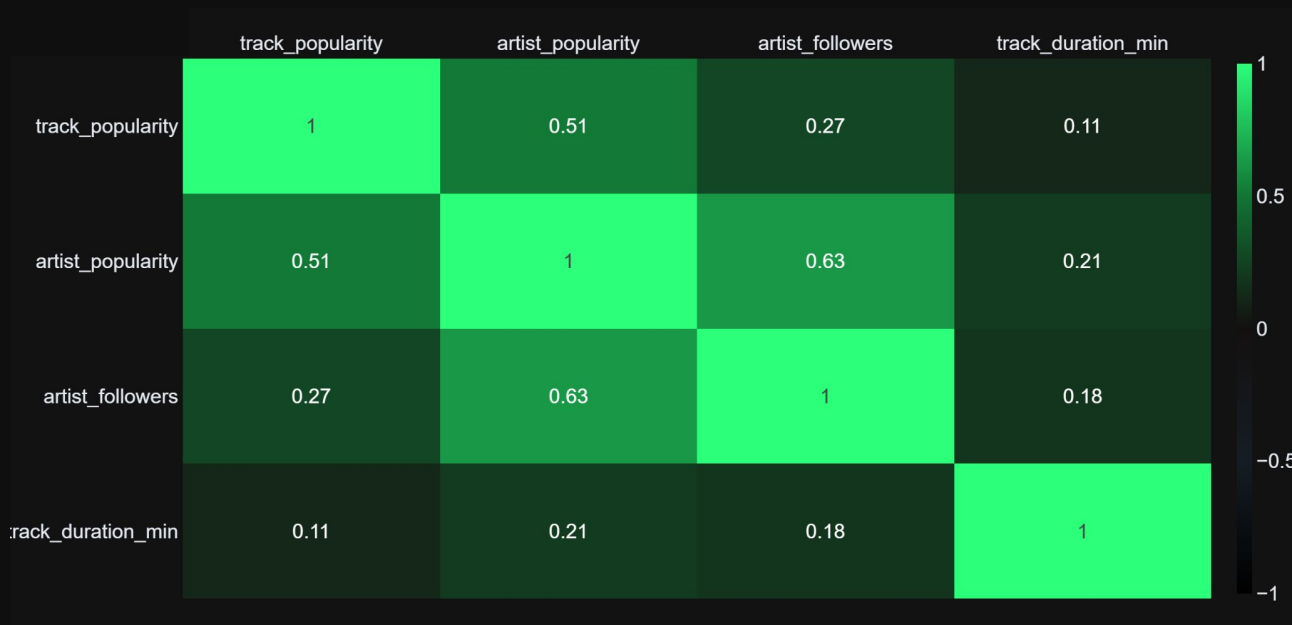
Top 10 artists

Artists are ranked using a combined popularity index based on artist popularity, followers, and track count. Features are normalized and weighted to reflect overall visibility, with emphasis on artist popularity.
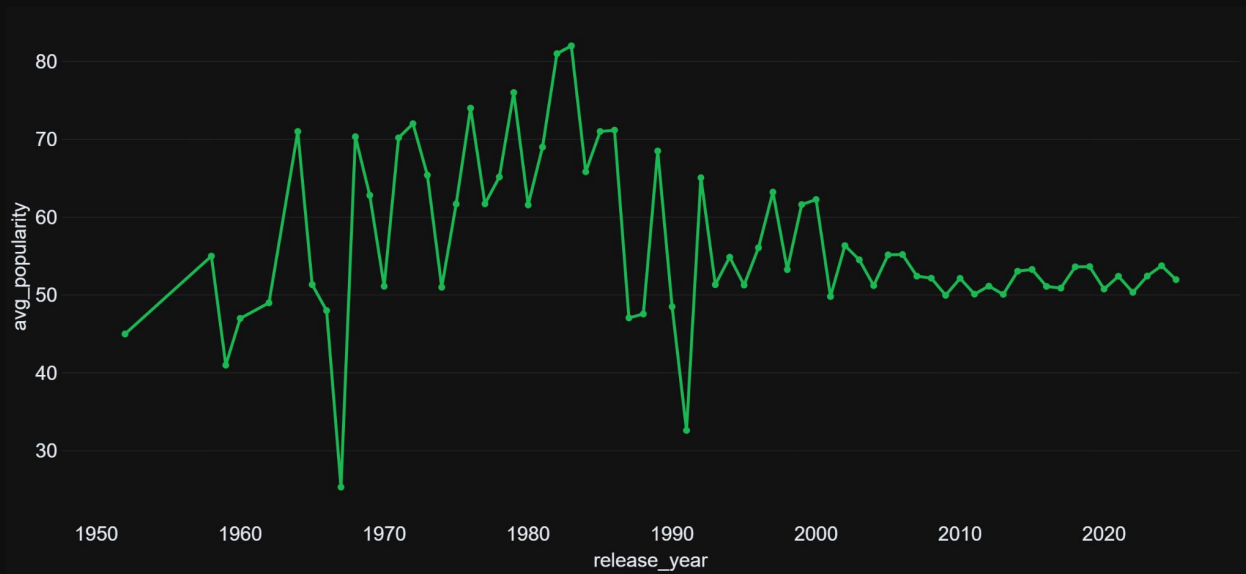
Track Popularity Is Highly Uneven

Track popularity is highly uneven, with many songs receiving little attention and only a small fraction becoming very popular. This skewed pattern reflects strong competition on the platform.

Artist Visibility Matters More Than Track Features

	track_popularity	artist_popularity	artist_followers	track_duration_min
track_popularity	1	0.51	0.27	0.11
artist_popularity	0.51	1	0.63	0.21
artist_followers	0.27	0.63	1	0.18
track_duration_min	0.11	0.21	0.18	1

Artist-level variables show the strongest relationship with track popularity, while track-specific features contribute much less. This suggests that artist visibility matters more than technical song attributes.
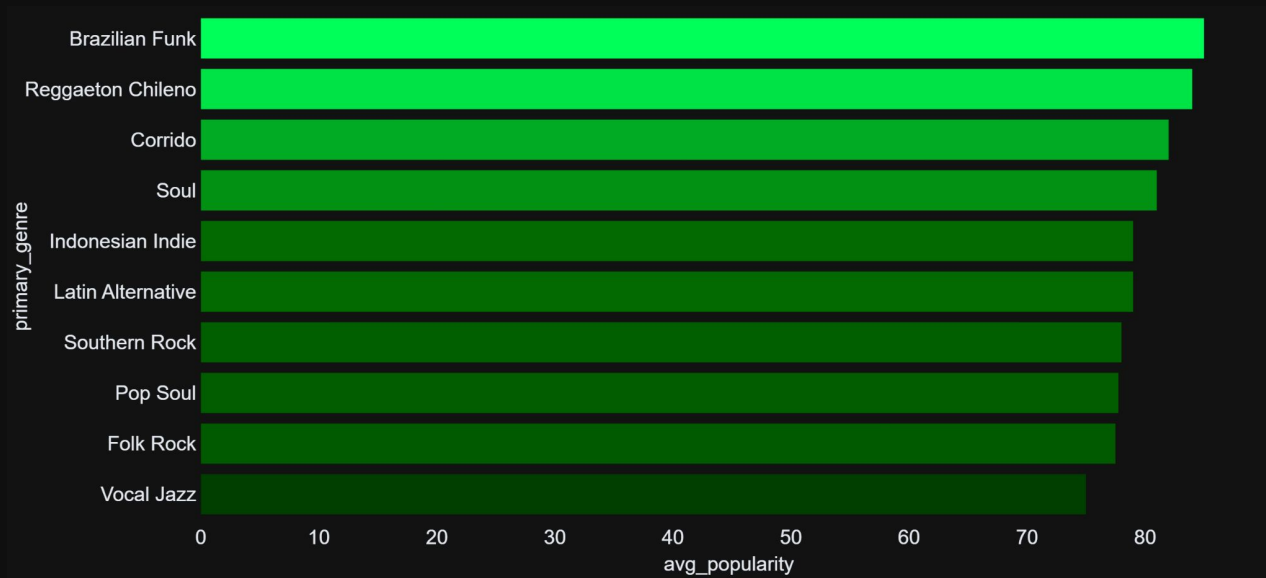
Popularity Patterns Stabilize Over Time

Early years show volatile averages due to limited data, while recent years are more stable. This indicates that current popularity patterns are reliable and structurally consistent.

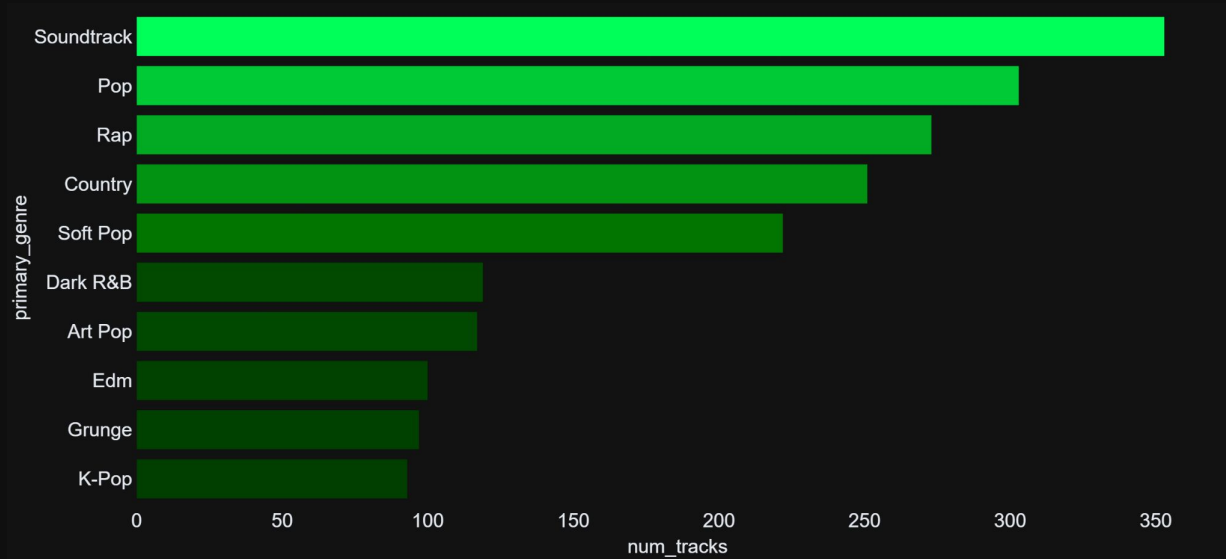# Explicit Content Does Not Guarantee Success

Explicit tracks have slightly higher median popularity, but the heavy overlap between groups shows that explicit content alone does not drive success.
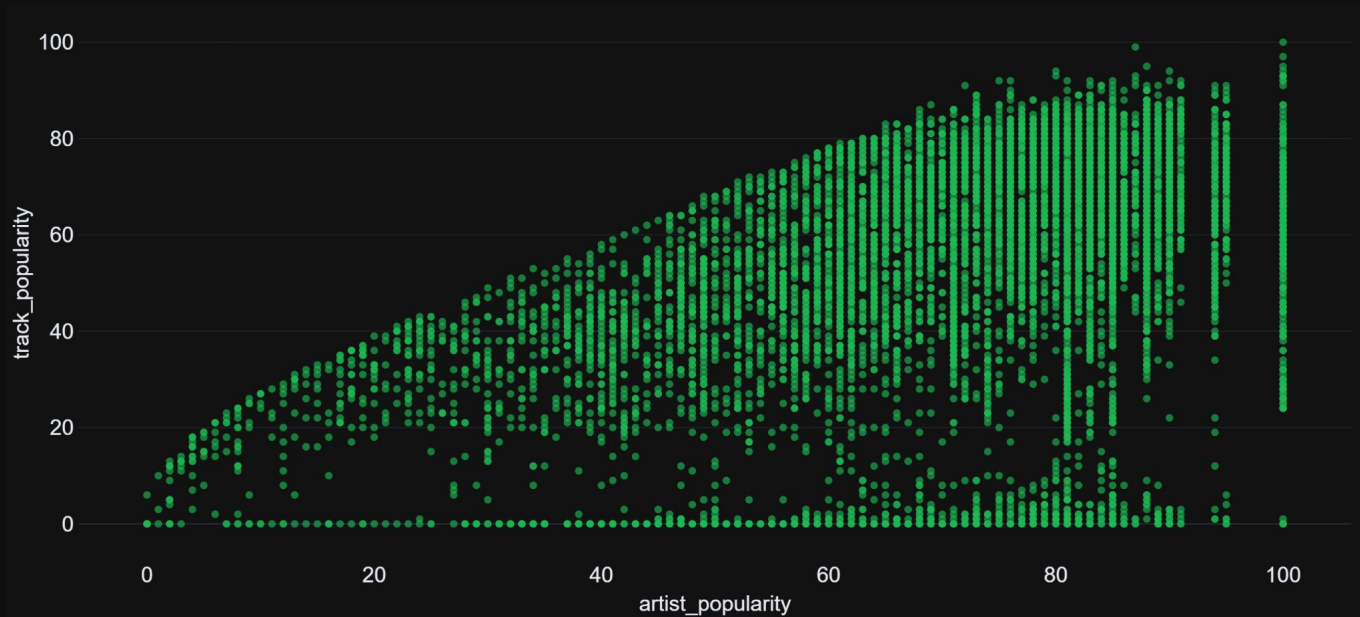
High Average Popularity Can Be Misleading

Some niche genres rank highly by average popularity, often due to small sample sizes. Average popularity should therefore be interpreted alongside genre frequency.
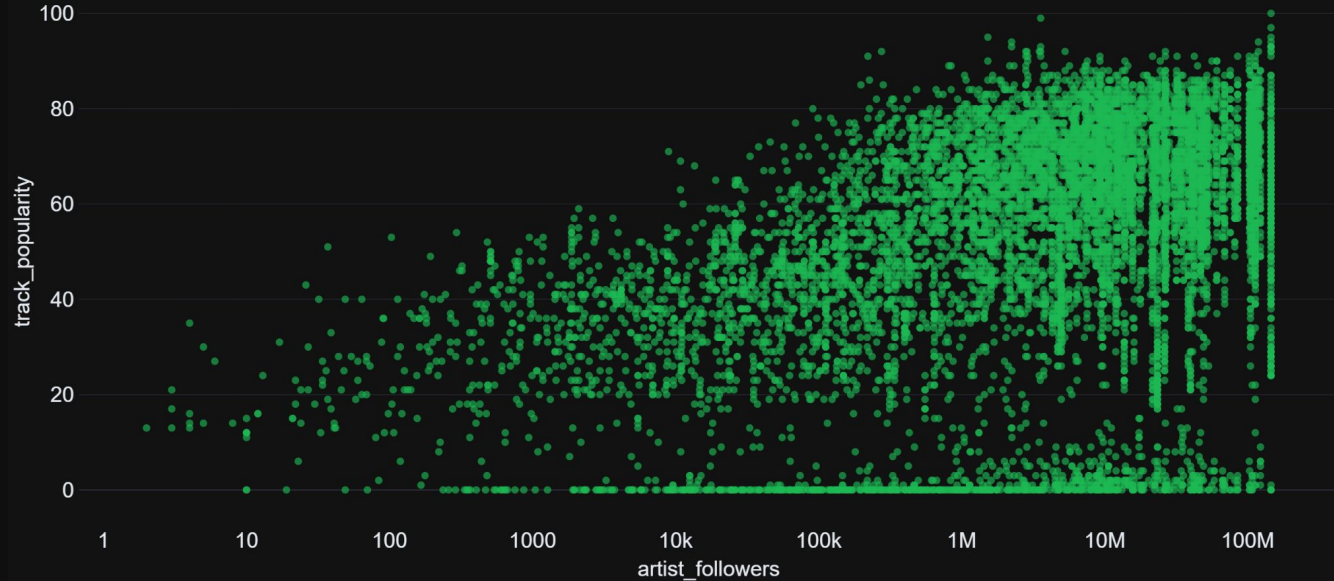
Popularity and Frequency Are Not the Same

High-frequency genres dominate the dataset but are not always the most successful on average. This highlights the difference between genre saturation and performance.
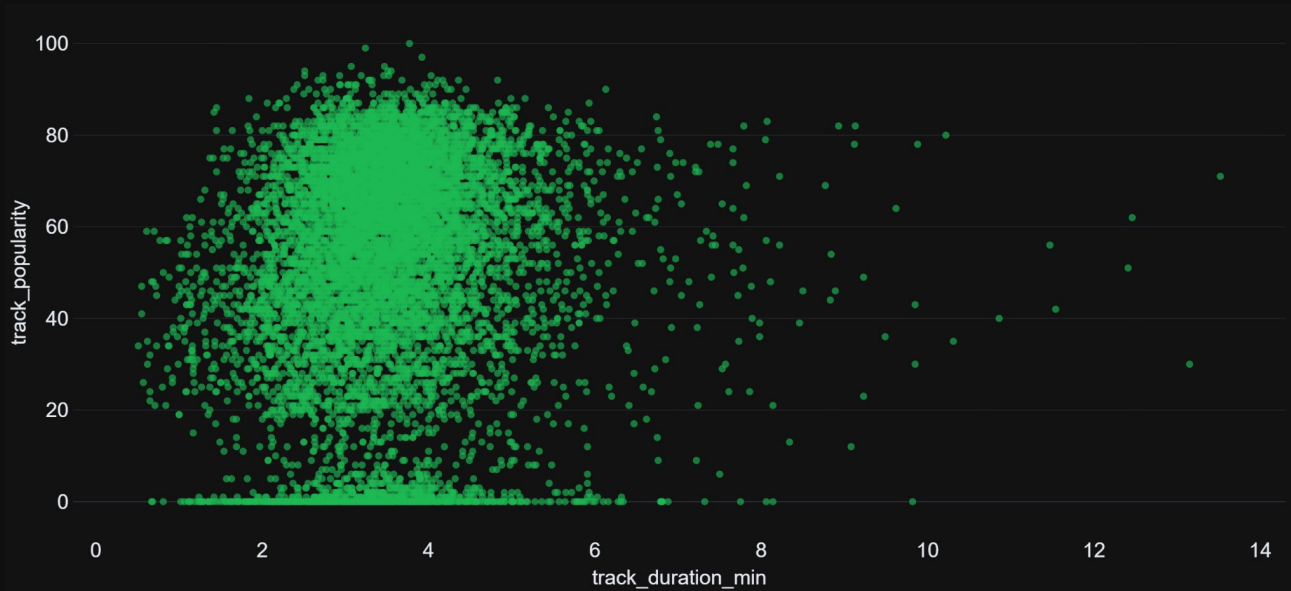
Artist Popularity Increases Upside Potential

Higher artist popularity increases the likelihood of success, but outcomes remain highly variable. Visibility raises potential without guaranteeing popularity.
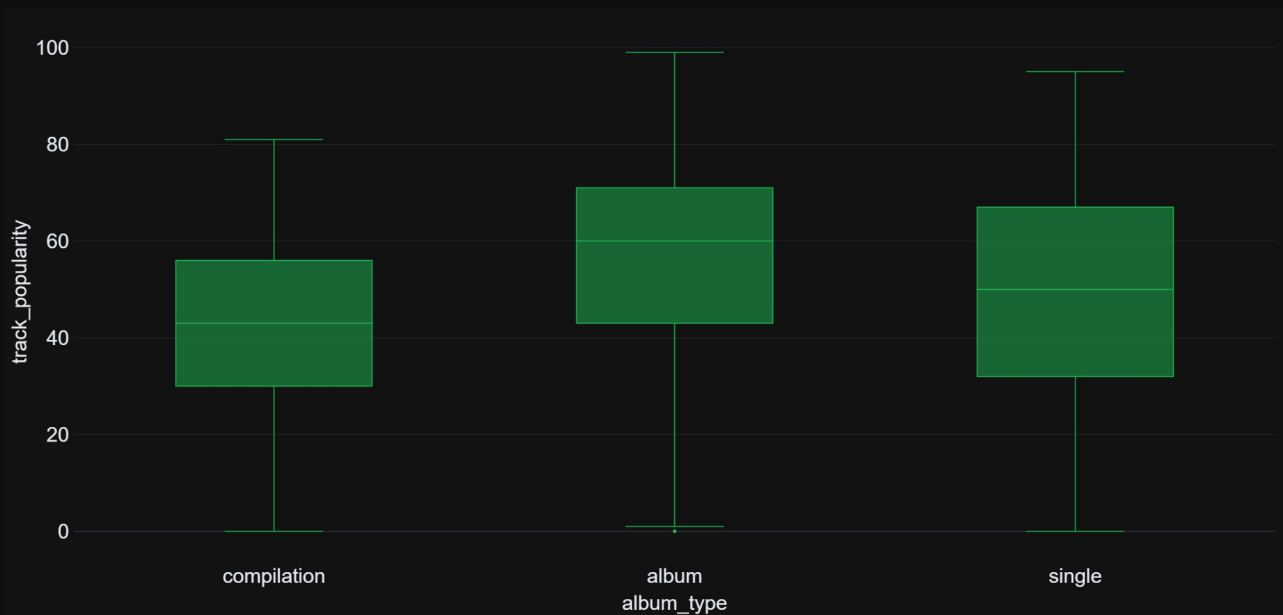
Audience Size Reflects Potential, Not Engagement

Follower count reflects potential reach but shows a noisy relationship with track popularity. Compared to artist popularity, it is a weaker indicator of success.

Track Length Has Little Impact on Popularity

Most tracks cluster around standard lengths regardless of popularity. This suggests track duration has little impact on performance.

Album Releases Perform Slightly Better on Average

Album tracks show slightly higher median popularity than singles or compilations, but strong overlap indicates album type alone does not determine success.

Simple Rule-Based Hit Identification

	Predicted Hit	Predicted Non-Hit
Actual Hit	1955	849
Actual Non-hit	2354	3603

How the rule works

- A track is a hit if it falls in the top 30% of track popularity
- A track is predicted as a hit if:
  - Artist popularity ≥ 75
  - Artist followers ≥ 1,000,000

What I learned

- Artist visibility strongly increases the chance of success
- Many tracks by popular artists still fail
- Useful for understanding and screening, but not precise enough for prediction

Similarity Score: Approach & Observations

How the rule works

- Chose a highly popular song as a reference point

- Compared other tracks using artist popularity, audience size, and track duration

- Normalized all features to ensure fair comparison

- Calculated a similarity score showing how close each track is to the reference

What I learned

- Tracks by the same or similarly popular artists appear closest to the reference

- Artist popularity decreases gradually as similarity scores increase

- Track popularity itself does not follow a clear pattern

- Individual song success remains difficult to predict

Final Research Findings

Popularity is highly concentrated

Most tracks receive little attention, while a small number capture most engagement.

Artist visibility matters most

Tracks by popular artists with large audiences tend to perform better, but success is not guaranteed.

Explicit content has little impact

Explicit and non-explicit tracks show very similar popularity patterns.

Genre, album type, and timing matter indirectly

Some genres and formats perform better on average, but competition and artist visibility play a bigger role.

Simple rules help, but are limited

Rule-based logic can identify some popular tracks, but it also produces many false positives.