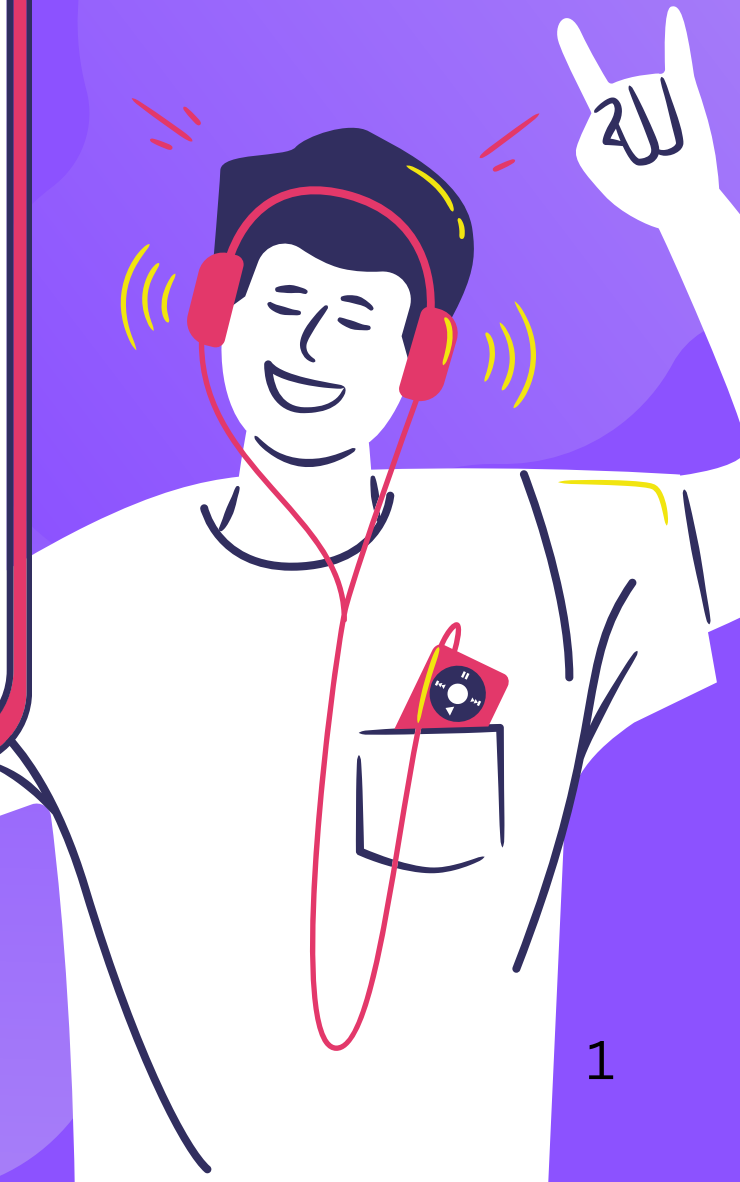
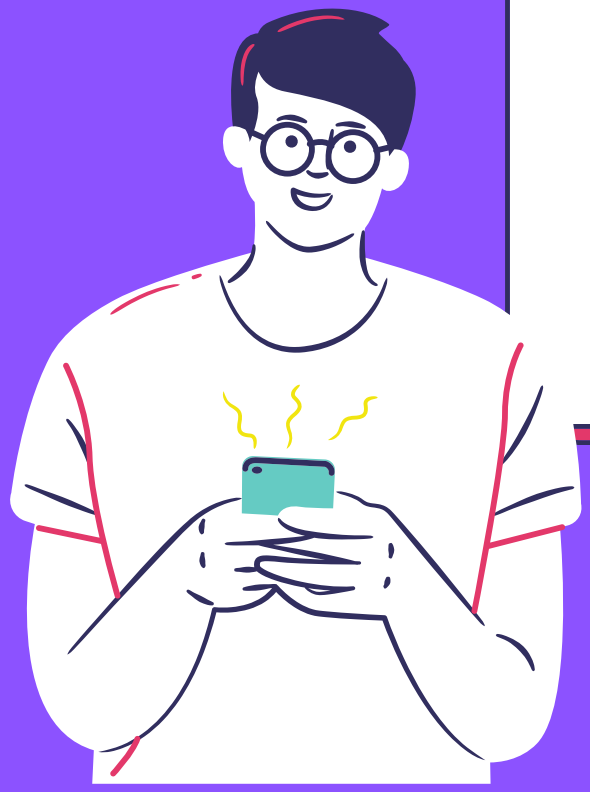
Spotify / YouTube Data Analysis

Select * From HITS



Presented By:
Lisa Yu
Chenxi Ye
Nik King Fredel

Question: How can a music studio prioritize investment?



Objective: Help music studios prioritize track investments using Spotify and YouTube data by **identifying high-performing songs** and **key audio features** that **signal potential breakout success.**

Data Overview

- 20,000+ tracks
- Featuring 2,000+ artists
- Includes:
  - Basic track info
  - Core audio features
  - Spotify streams and YouTube engagement

Data Overview

TRACK DATA

- artist: Primary artist name
- track: Track title
- album: Album name
- album_type: Type of release (e.g., single, album, compilation)
- most_playedon: Dominant platform (“Spotify” or “YouTube”)

YOUTUBE DATA

- title: YouTube video title
- channel: Channel name
- views: Total views
- likes: Total likes
- comments: Total comments
- licensed: Flag indicating if the video is officially licensed (1 = Yes, 0 = No)
- official_video: Flag for whether it’s an official music video (1 = Yes, 0 = No)

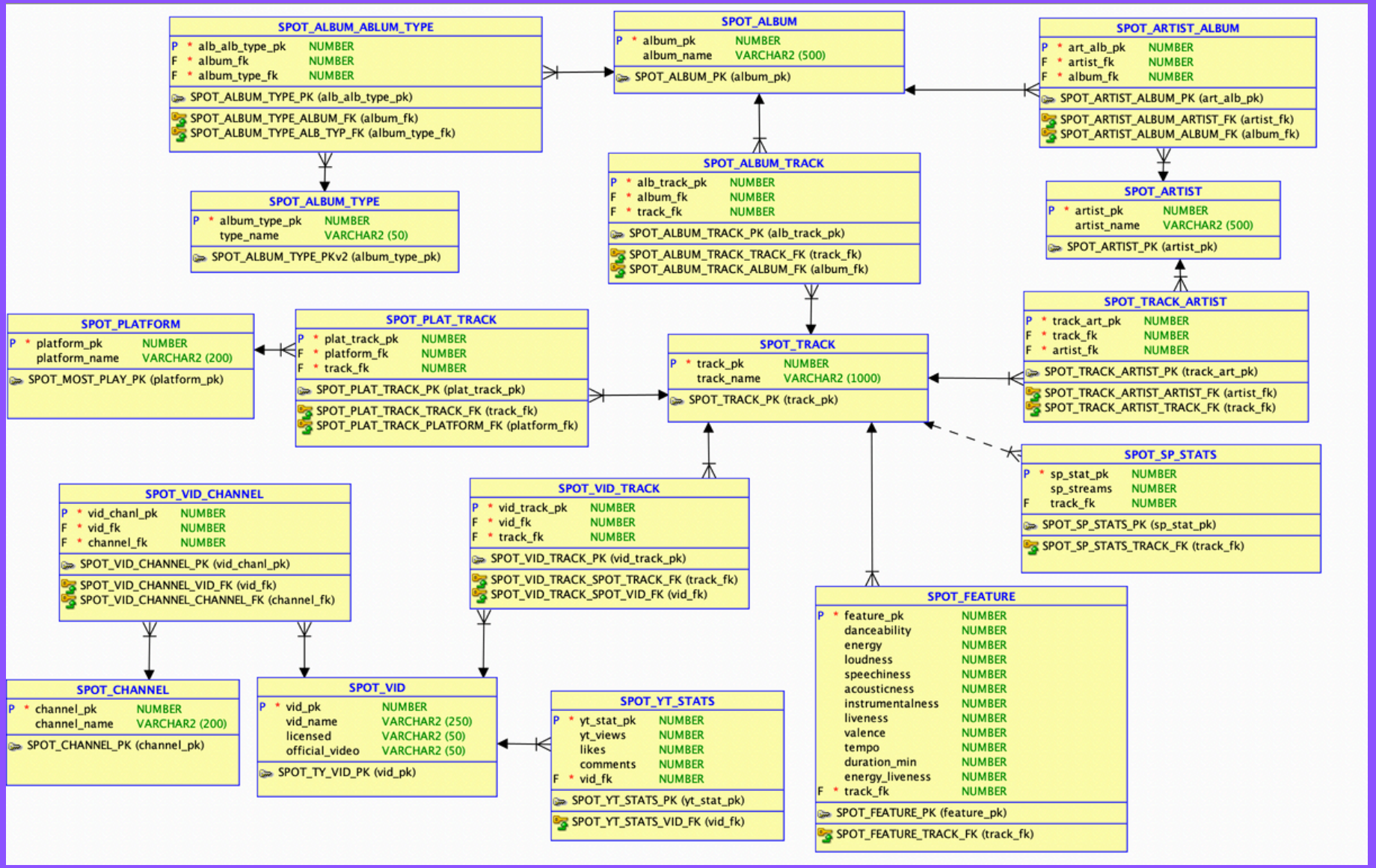
SPOTIFY DATA

- stream: Total number of Spotify streams for the track

AUDIO FEATURES

- danceability: Suitability for dancing, based on rhythm and musical elements (0–1)
- energy: Perceived intensity and activity level (0–1)
- loudness: Overall track loudness in decibels (dB)
- speechiness: Presence of spoken words (0–1)
- acousticness: Confidence the track is acoustic (0–1)
- instrumentalness: Likelihood the track has no vocals (0–1)
- liveness: Detects audience presence, indicating live performance (0–1)
- valence: Musical positivity or mood (0–1)
- tempo: Estimated tempo in beats per minute (BPM)
- duration: Track duration in minutes (converted from milliseconds)
- energyliveness: Combined feature representing both energy and liveness

Data Model


SPOT_ALBUM_ABLUM_TYPE
P * alb_alb_type_pk NUMBER
F * album_fk NUMBER
F * album_type_fk NUMBER
SPOT_ALBUM_TYPE_PK (alb_alb_type_pk)
SPOT_ALBUM_TYPE_ALBUM_FK (album_fk)
SPOT_ALBUM_TYPE_ALB_TYP_FK (album_type_fk)
SPOT_ALBUM
P * album_pk NUMBER
album_name VARCHAR2 (500)
SPOT_ALBUM_PK (album_pk)
SPOT_ARTIST_ALBUM
P * art_alb_pk NUMBER
F * artist_fk NUMBER
F * album_fk NUMBER
SPOT_ARTIST_ALBUM_PK (art_alb_pk)
SPOT_ARTIST_ALBUM_ARTIST_FK (artist_fk)
SPOT_ARTIST_ALBUM_ALBUM_FK (album_fk)
SPOT_ALBUM_TYPE
P * album_type_pk NUMBER
type_name VARCHAR2 (50)
SPOT_ALBUM_TYPE_PKv2 (album_type_pk)
SPOT_ALBUM_TRACK
P * alb_track_pk NUMBER
F * album_fk NUMBER
F * track_fk NUMBER
SPOT_ALBUM_TRACK_PK (alb_track_pk)
SPOT_ALBUM_TRACK_TRACK_FK (track_fk)
SPOT_ALBUM_TRACK_ALBUM_FK (album_fk)
SPOT_ARTIST
P * artist_pk NUMBER
artist_name VARCHAR2 (500)
SPOT_ARTIST_PK (artist_pk)
SPOT_PLATFORM
P * platform_pk NUMBER
platform_name VARCHAR2 (200)
SPOT_MOST_PLAY_PK (platform_pk)
SPOT_PLAT_TRACK
P * plat_track_pk NUMBER
F * platform_fk NUMBER
F * track_fk NUMBER
SPOT_PLAT_TRACK_PK (plat_track_pk)
SPOT_PLAT_TRACK_TRACK_FK (track_fk)
SPOT_PLAT_TRACK_PLATFORM_FK (platform_fk)
SPOT_TRACK
P * track_pk NUMBER
track_name VARCHAR2 (1000)
SPOT_TRACK_PK (track_pk)
SPOT_TRACK_ARTIST
P * track_art_pk NUMBER
F * track_fk NUMBER
F * artist_fk NUMBER
SPOT_TRACK_ARTIST_PK (track_art_pk)
SPOT_TRACK_ARTIST_ARTIST_FK (artist_fk)
SPOT_TRACK_ARTIST_TRACK_FK (track_fk)
SPOT_SP_STATS
P * sp_stat_pk NUMBER
sp_streams NUMBER
F track_fk NUMBER
SPOT_SP_STATS_PK (sp_stat_pk)
SPOT_SP_STATS_TRACK_FK (track_fk)
SPOT_VID_CHANNEL
P * vid_chanl_pk NUMBER
F * vid_fk NUMBER
F * channel_fk NUMBER
SPOT_VID_CHANNEL_PK (vid_chanl_pk)
SPOT_VID_CHANNEL_VID_FK (vid_fk)
SPOT_VID_CHANNEL_CHANNEL_FK (channel_fk)
SPOT_VID_TRACK
P * vid_track_pk NUMBER
F * vid_fk NUMBER
F * track_fk NUMBER
SPOT_VID_TRACK_PK (vid_track_pk)
SPOT_VID_TRACK_SPOT_TRACK_FK (track_fk)
SPOT_VID_TRACK_SPOT_VID_FK (vid_fk)
SPOT_FEATURE
P * feature_pk NUMBER
danceability NUMBER
energy NUMBER
loudness NUMBER
speechiness NUMBER
acousticness NUMBER
instrumentalness NUMBER
liveness NUMBER
valence NUMBER
tempo NUMBER
duration_min NUMBER
energy_liveness NUMBER
F * track_fk NUMBER
SPOT_FEATURE_PK (feature_pk)
SPOT_FEATURE_TRACK_FK (track_fk)
SPOT_CHANNEL
P * channel_pk NUMBER
channel_name VARCHAR2 (200)
SPOT_CHANNEL_PK (channel_pk)
SPOT_VID
P * vid_pk NUMBER
vid_name VARCHAR2 (250)
licensed VARCHAR2 (50)
official_video VARCHAR2 (50)
SPOT_TY_VID_PK (vid_pk)
SPOT_YT_STATS
P * yt_stat_pk NUMBER
yt_views NUMBER
likes NUMBER
comments NUMBER
F * vid_fk NUMBER
SPOT_YT_STATS_PK (yt_stat_pk)
SPOT_YT_STATS_VID_FK (vid_fk)

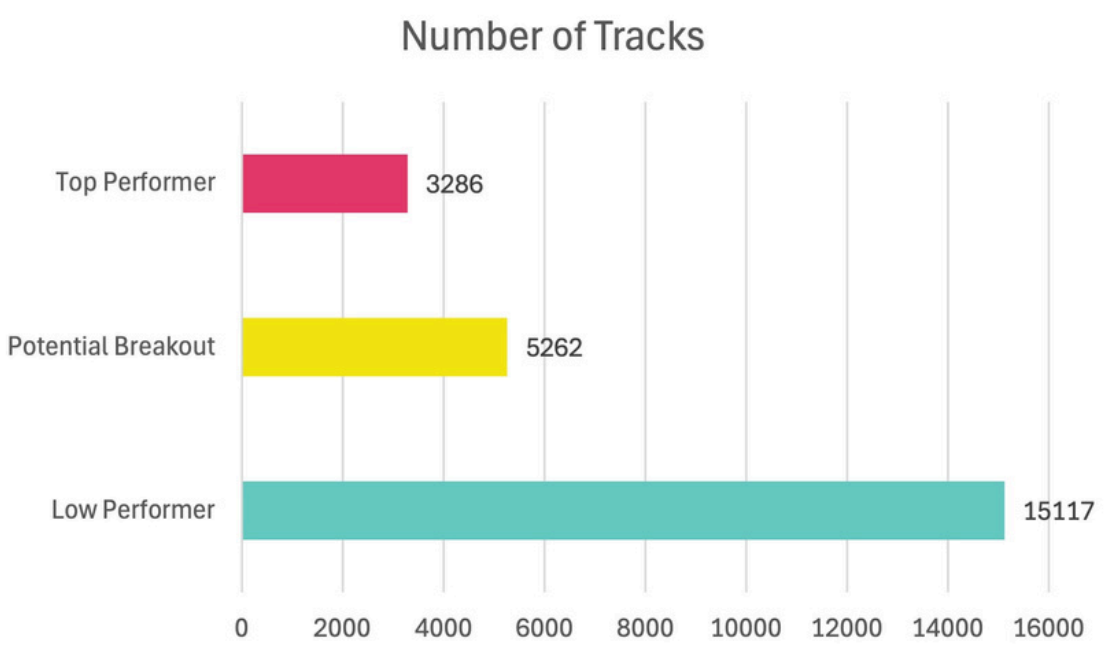
Performance Analysis Part 1 :

Clusters of audio features that are common in trending tracks

Classification Approach

- **Top Performer:** Top 25% in both streams and views
- **Potential Breakout:** Top 25% in either streams or views, but not both
- **Low Performer:** any song not in the top quartile

Classification based on NTILE(4) percentile ranking of streams and views.



Average Duration of Songs

Category	Avg Duration (mins)
Top Performer	3.78
Potential Breakout	3.74
Low Performer	3.65

Top performers tend to be slightly longer - suggesting a stronger built-up and catchy chorus or bridge.

Musical Characteristics - Danceability

Category	Avg Danceability
Top Performer	0.643
Potential Breakout	0.62
Low Performer	0.606

Higher danceability is consistently associated with top-performing tracks – suggesting that rhythm and groove matter.

Musical Characteristics - Energy

Category	Avg Energy
Top Performer	0.665
Potential Breakout	0.635
Low Performer	0.617

Top-performing tracks are slightly more energetic – reinforcing the value of intensity and drive.

Musical Characteristics - Loudness

Category	Avg Loudness
Top Performer	-6.3
Potential Breakout	-7.05
Low Performer	-8.06

Top performers are consistently louder — suggesting greater polish or post-production intensity.

Musical Characteristics - Speechiness

Category	Avg Speechines
Top Performer	0.089
Potential Breakout	0.08
Low Performer	0.092

Low-performing tracks had slightly more speech-like content — suggesting that hits may favor melodic over spoken-word vocals.

Musical Characteristics - Acousticness

Category	Avg Acousticness
Top Performer	0.089
Potential Breakout	0.08
Low Performer	0.092

Hit songs skew less acoustic – suggesting heavier instrumentation and production.

Musical Characteristics - Instrumentalness

Category	Avg Instrumentalness
Top Performer	0.022
Potential Breakout	0.034
Low Performer	0.065

Instrumental tracks are less likely to succeed – Top Performers tend to be more vocal-heavy.

Musical Characteristics - Liveness

Category	Avg Liveness
Top Performer	0.18
Potential Breakout	0.182
Low Performer	0.188

Lower liveness correlates with higher performance – hits sound more polished and less like live recordings.

Musical Characteristics - Valence

Category	Avg Valence
Top Performer	0.52
Potential Breakout	0.514
Low Performer	0.519

Slight uptick in positivity among top performers, but valence shows only weak correlation with success.

Musical Characteristics - Tempo

Category	Avg Tempo
Top Performer	121.55
Potential Breakout	121.35
Low Performer	120.36

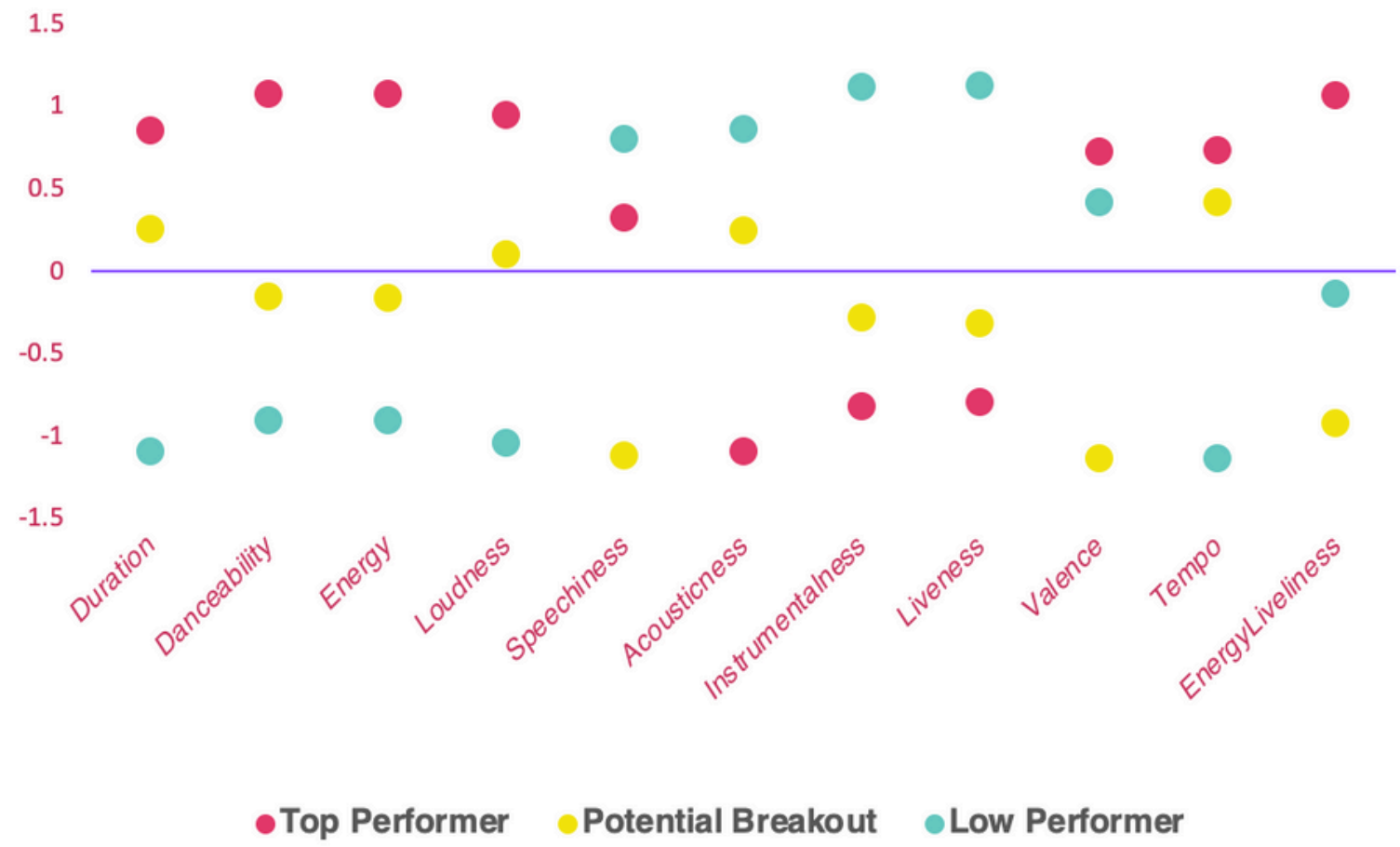
Top-performing tracks trend slightly faster – tempo may contribute to listener engagement.

Musical Characteristics - Energy Liveness

Category	Avg Energy Liveness
Top Performer	5.496
Potential Breakout	5.007
Low Performer	5.2

Higher combined energy and liveness scores are associated with stronger performance.

Characteristics Analysis



*All feature values are standardized using Z-score normalization

Insights & Recommendations

Characteristics of top tracks:

- More upbeat, energetic, and emotionally positive.
- More vocal-driven and less acoustic and instrumental-heavy.

Prioritize tracks that:

- **score high on**
  - Danceability (>0.62)
  - Energy (>0.635)
  - Loudness (>7.05)
  - Tempo (>121.35)
- **score low on**
  - Acousticness (<0.089)
  - Instrumentalness (<0.034)
  - Speechness (<0.09)

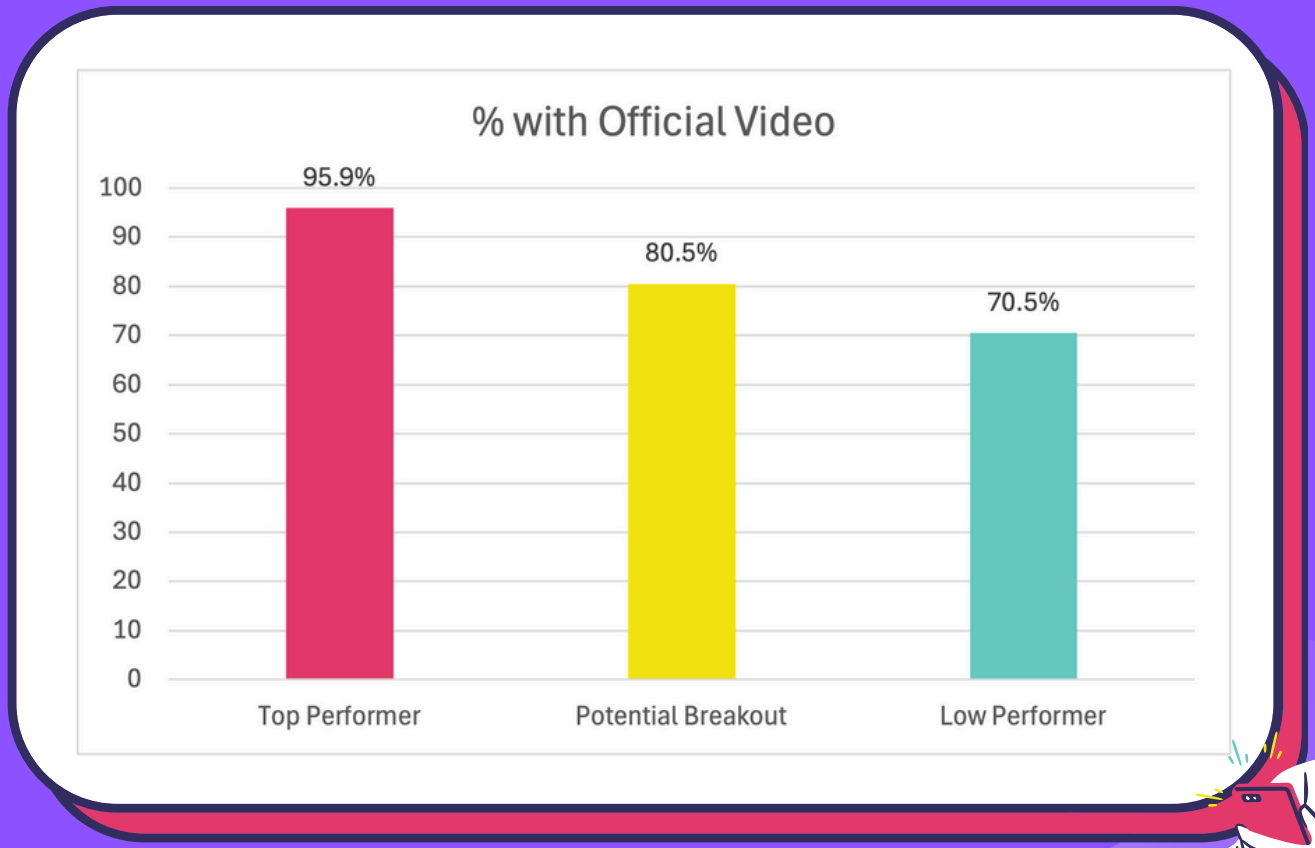
Performance Analysis Part 2 :

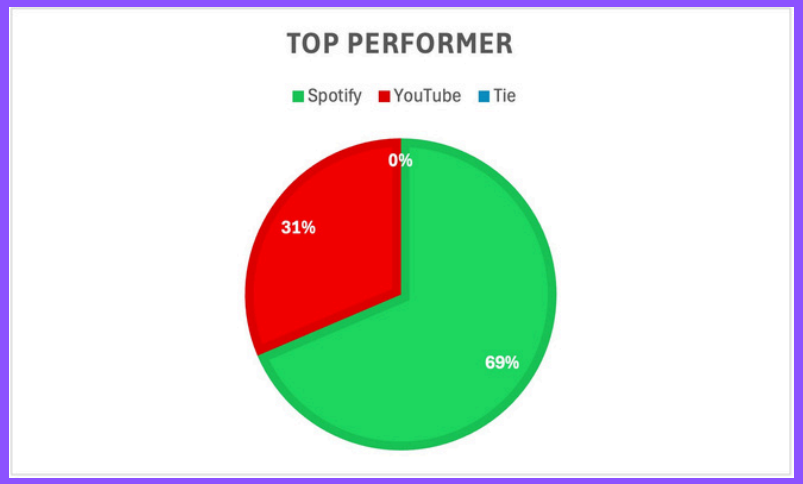
Impact of official videos and platforms

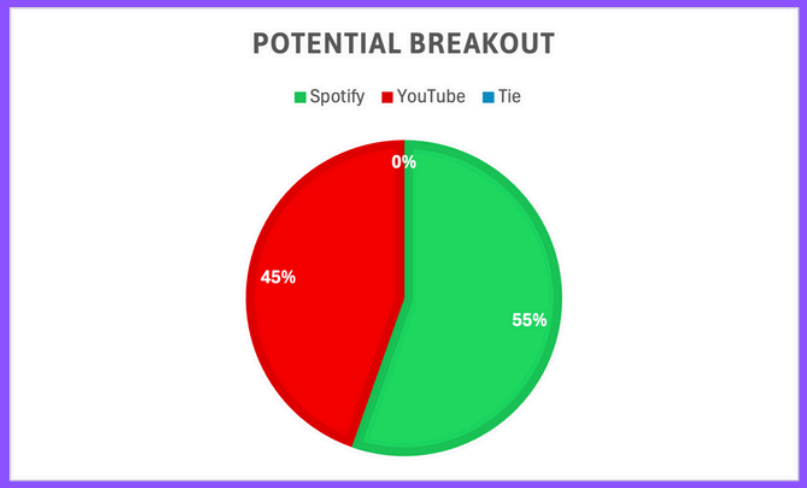
Impact of Official Video

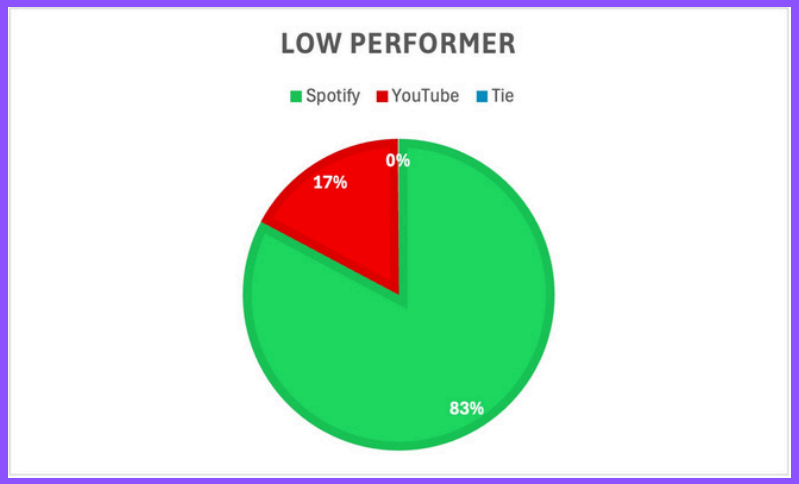
Top Performer: Top 25% in both streams and views

- **Potential Breakout:** Top 25% in either, but not both
- **Low Performer:** Any song not in the top quartile









Dominant Platform Analysis

Categorization Criteria:

- YouTube if views > streams
- Spotify if streams > views.

Insights and Recommendations:

- Spotify is the **stronger** platform for songs
- Youtube serves as a **launchpad** for emerging songs.
- If a track is gaining traction on Youtube but not yet on Spotify, treat that as a **“potential breakout” candidate**.

Performance Analysis Part 3 :

Platform specific features + popular examples

Audio features: Spotify vs YouTube

Instrumentalness

Acousticness

Loudness

Energy

Speechiness

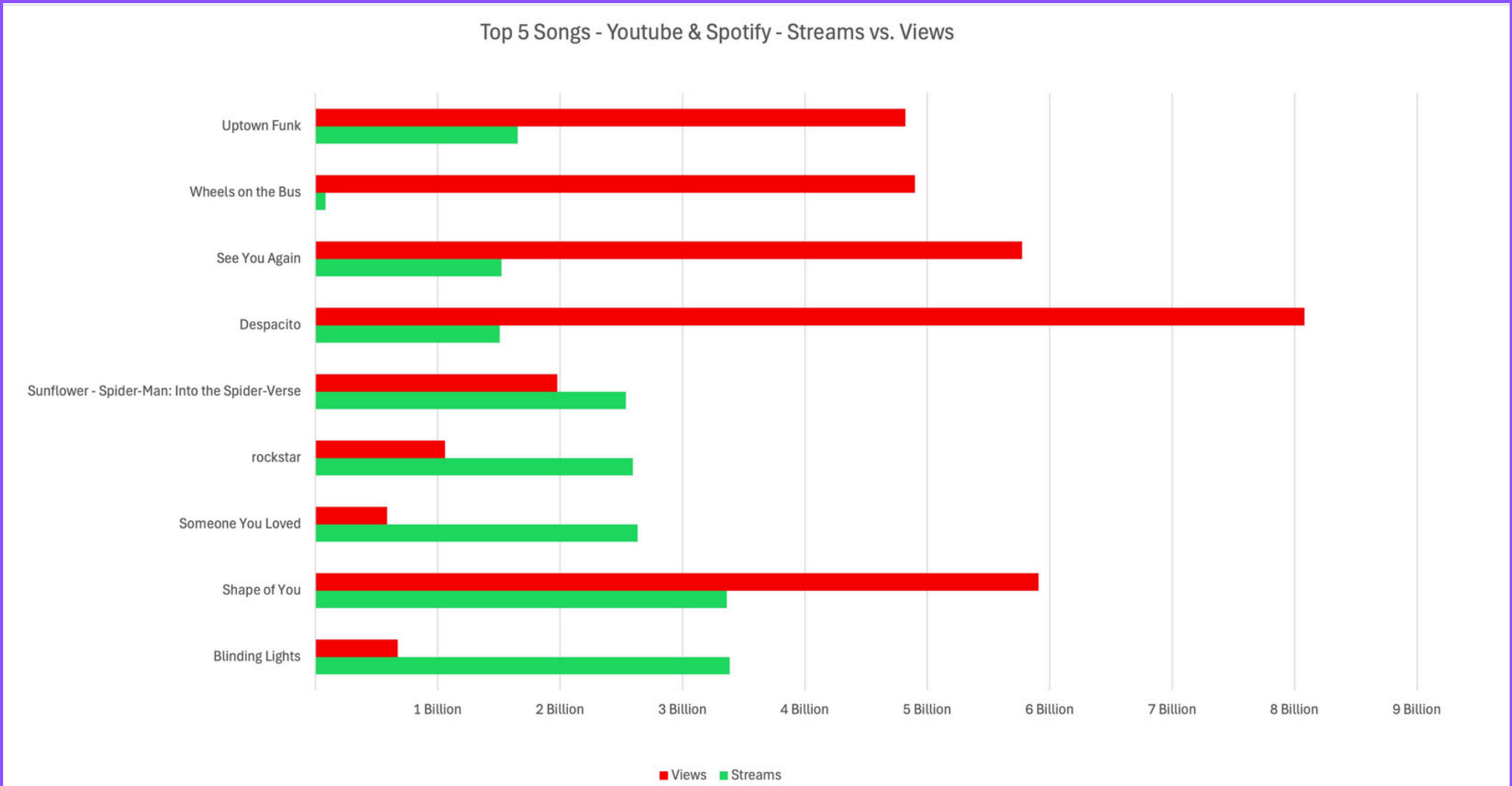
Liveness

Valence

Insights and Recommendations:

- **Quiter and non-vocal** songs are more popular on **Spotify**
- **Energetic and high-tempo** songs are more popular on **Youtube**.

Top 5 songs on Spotify and YouTube



Top 5 Youtube Videos

Despacito
Daddy Yankee & Luis Fonsi

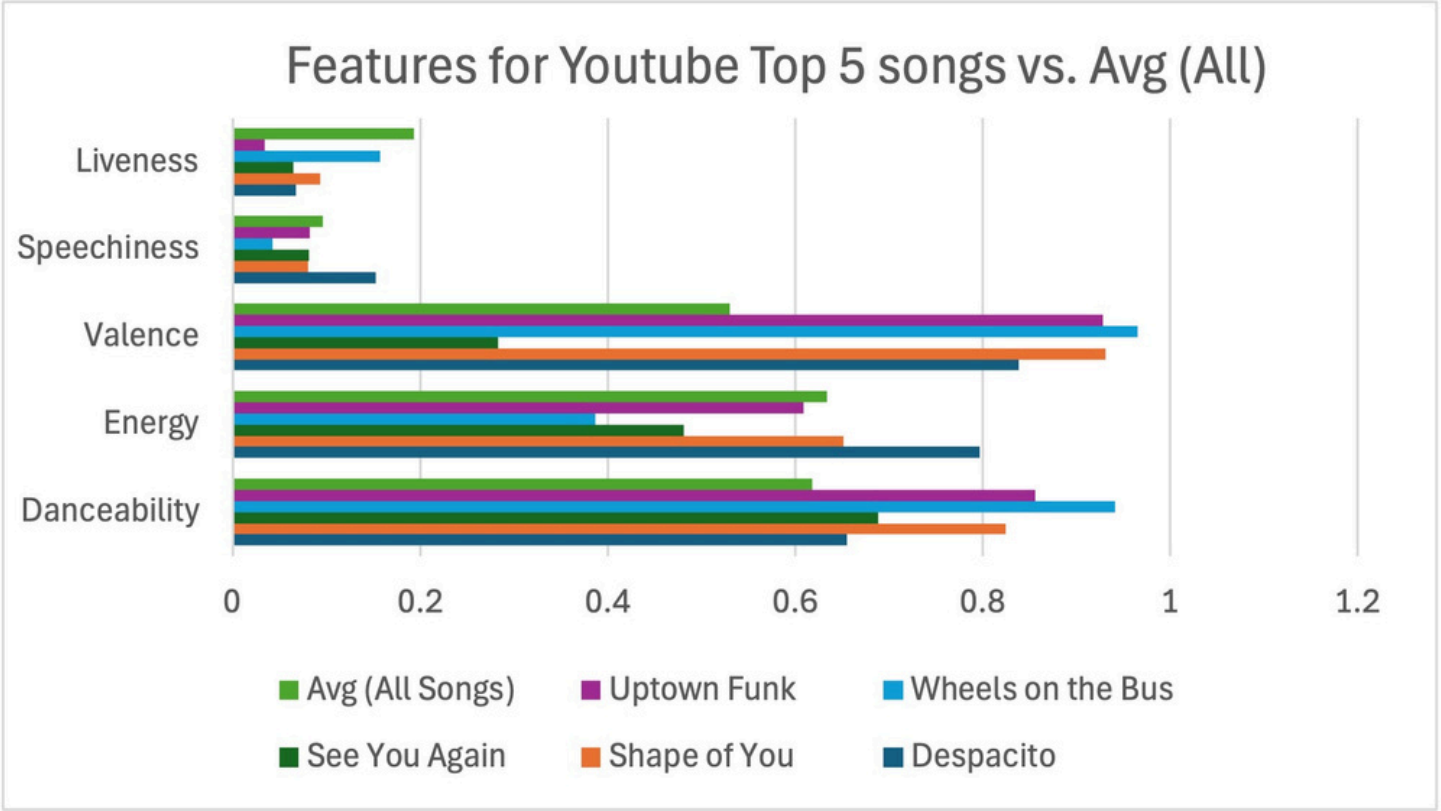
Shape of You
Ed Sheeran

See You Again
Wiz Khalifa & Charlie Puth

Wheels on the Busy
CoCoMelon

Uptown Funk
Mark Ronson & Bruno Mars

Top 5 Youtube Videos



YouTube's hits lean into high positivity and features like liveness and clarity, suggesting that emotional appeal and performative qualities play a bigger role in driving success.

Top 5 Spotify Songs

Blinding Lights
The Weeknd

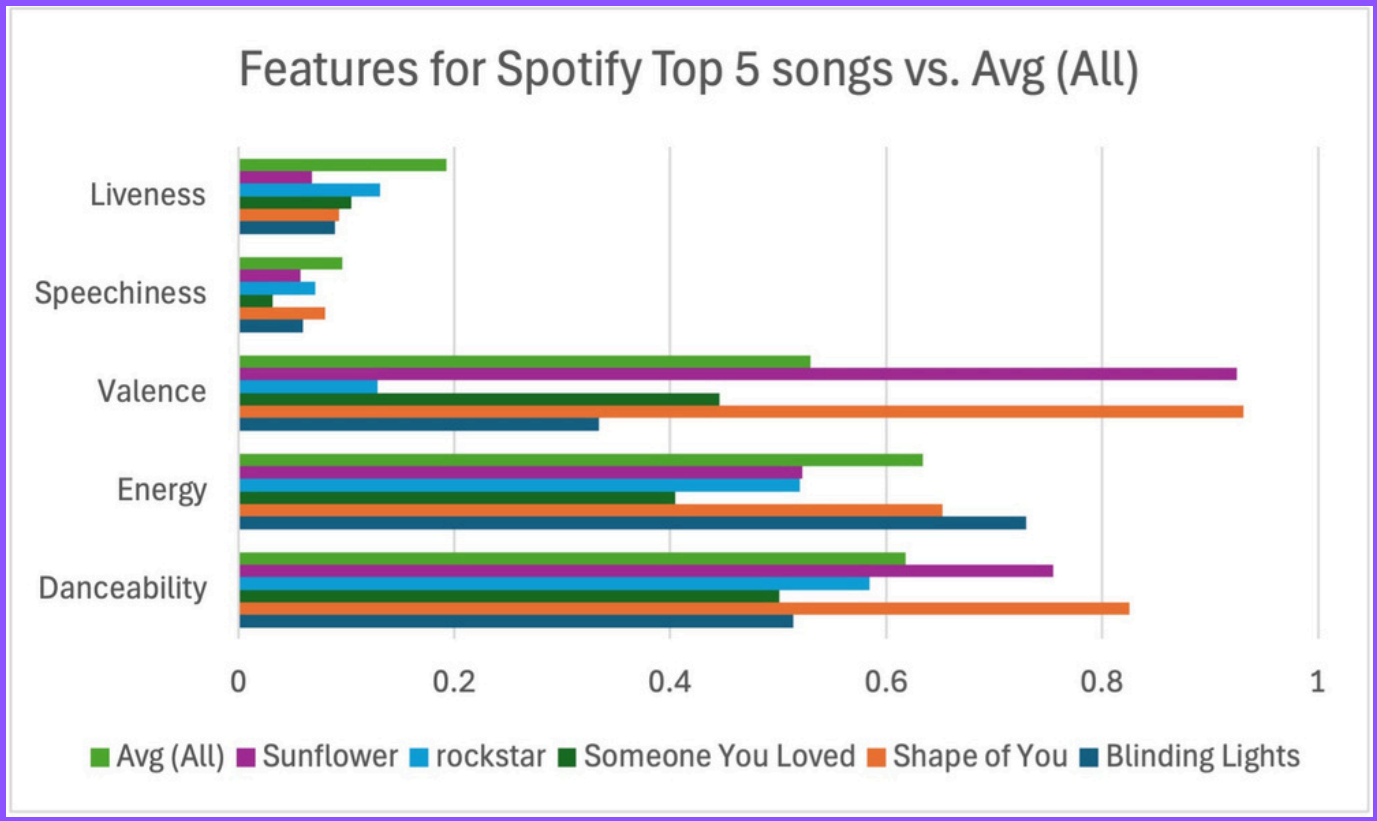
Shape of You
Ed Sheeran

Someone You Loved
Lewis Capaldi

Rockstar
Post Malone & 21 Savage

Sunflower
Post Malone & Swae Lee

Top 5 Spotify Tracks



Spotify's top tracks cluster around rhythm and energy – ideal for passive, playlist-based listening.

Overall Recommendations

About Music

1. Focus on more **energetic and high-tempo** songs
2. Less prioritize **instrumental and acoustic** songs

About Platform

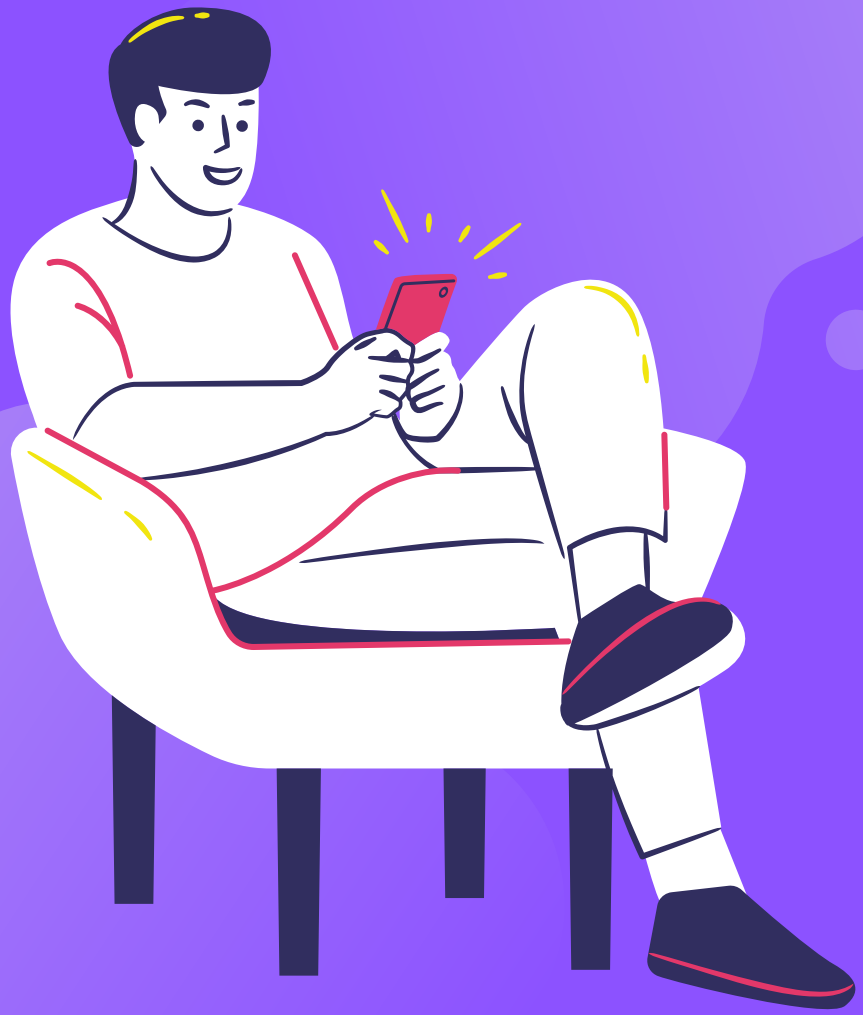
1. Prioritize releasing songs on **Spotify**
2. Use **Youtube** as a jump pad to **attract audience** and to **identify potential breakout** songs

About Long-term Effort

1. Build a **predictive scoring systems** that can systematically evaluate audio features + platform of songs to determine worthiness of investing the song

Limitations

- No genre classification
- No visual analysis
- Timeframe not considered
- No user behavior data
- Limited platform scope



Thank You