

What Makes A Hit Song?

Analyzing Spotify Data to Find the Traits of Popular Music



CONCORDIA
INTERNATIONAL SCHOOL HANOI



Jonathan Lubaton 

jonathan.lubaton@concordiahanoi.org

INTRODUCTION

Music plays a major role in daily life, yet only some songs become highly popular while others go unnoticed. Although taste is subjective, songs share measurable features like danceability, energy, tempo, loudness, and valence that can influence listener response. With over 100 million tracks on Spotify, competition is intense. This project compares high- and low-popularity songs using Kaggle datasets to identify patterns in these features and determine how they may contribute to a song's success.

DATA & METHODS

Source: Spotify via Kaggle (Aug 2025)

- High Popularity: 1,687 tracks
- Low Popularity: 3,146 tracks
- 29 columns each (same structure)
- No cleaning needed - all valid

Features analyzed: danceability, energy, tempo, loudness, valence, duration, etc.

VISUALIZATIONS & ANALYSIS

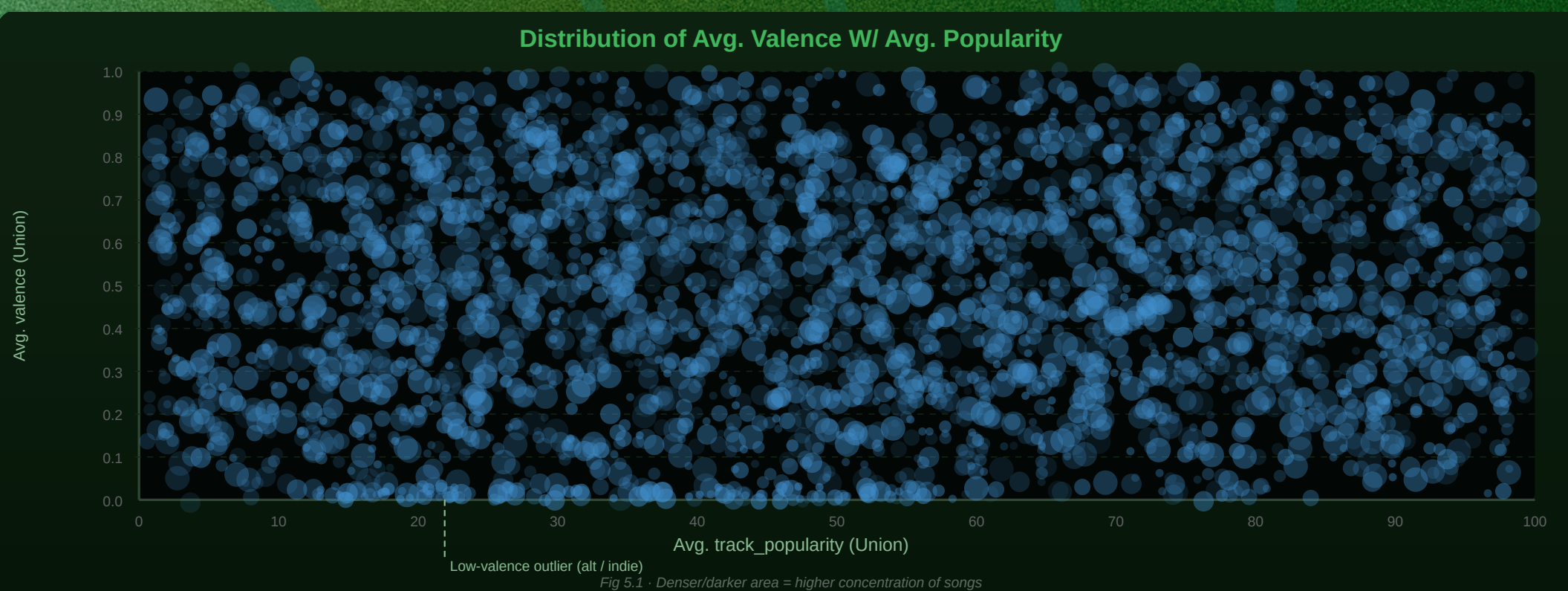


Fig 5.1 - Valence Unlike energy and tempo, valence shows a more scattered, uniform distribution — emotional tone alone doesn't strictly determine popularity. However, tracks with mid-to-positive valence (0.4–0.8) appear more densely in higher popularity ranges, while the low-valence outlier cluster shows darker songs can still find a consistent audience, just less predictably.

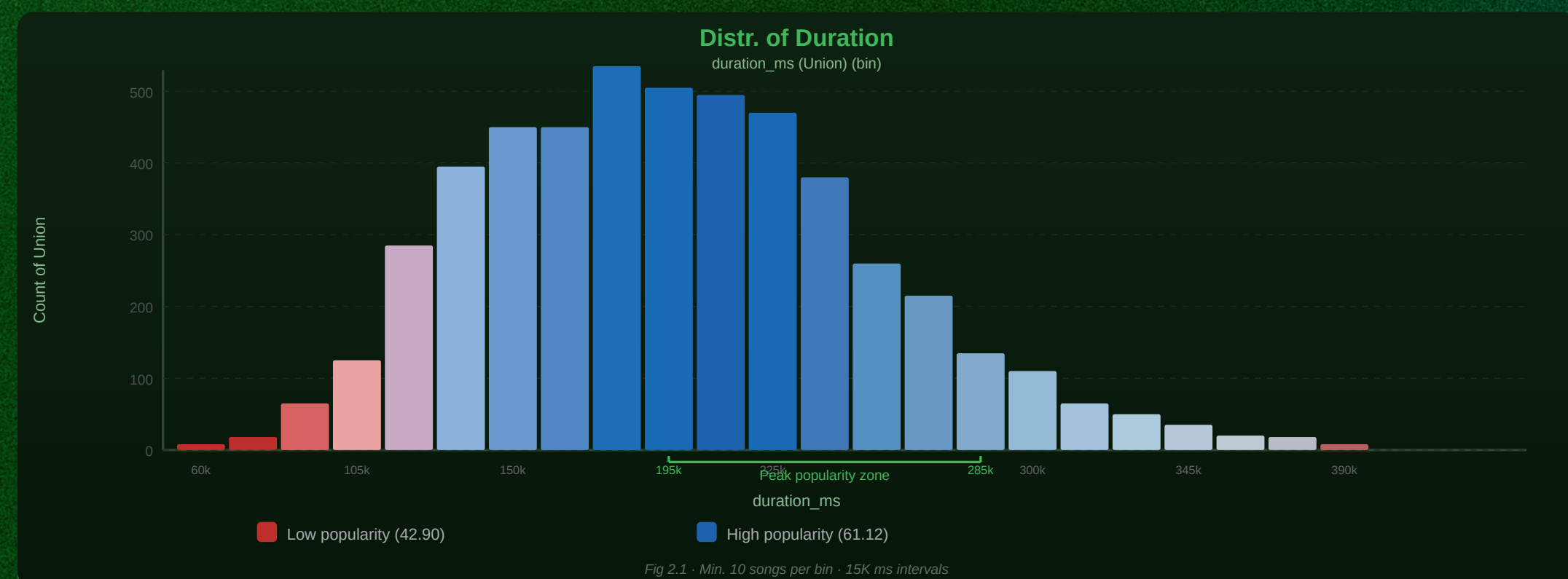


Fig 2.1 - Duration Songs between 195k–285k ms (3:15 to 4:45 min) score the highest average popularity, aligning with Billboard Hot 100 trends. Tracks shorter than 3 minutes or longer than 5 minutes show a clear drop, suggesting both listeners and algorithms favor mid-length songs.

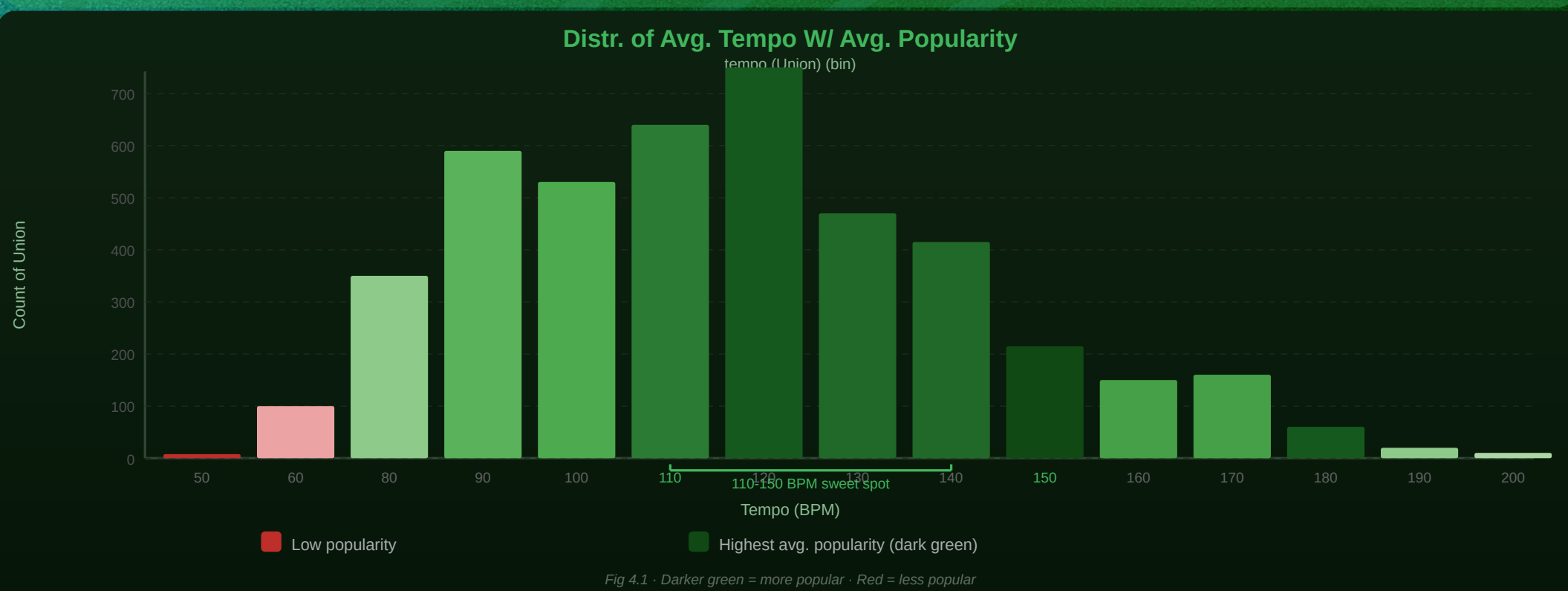


Fig 4.1 - Tempo Tracks in the 110–150 BPM range, the natural home of pop, hip-hop, and dance, consistently show the darkest green bars and highest popularity scores. The 150 BPM bin especially punches above its weight with one of the highest average popularity values (57.83) despite having fewer total tracks.

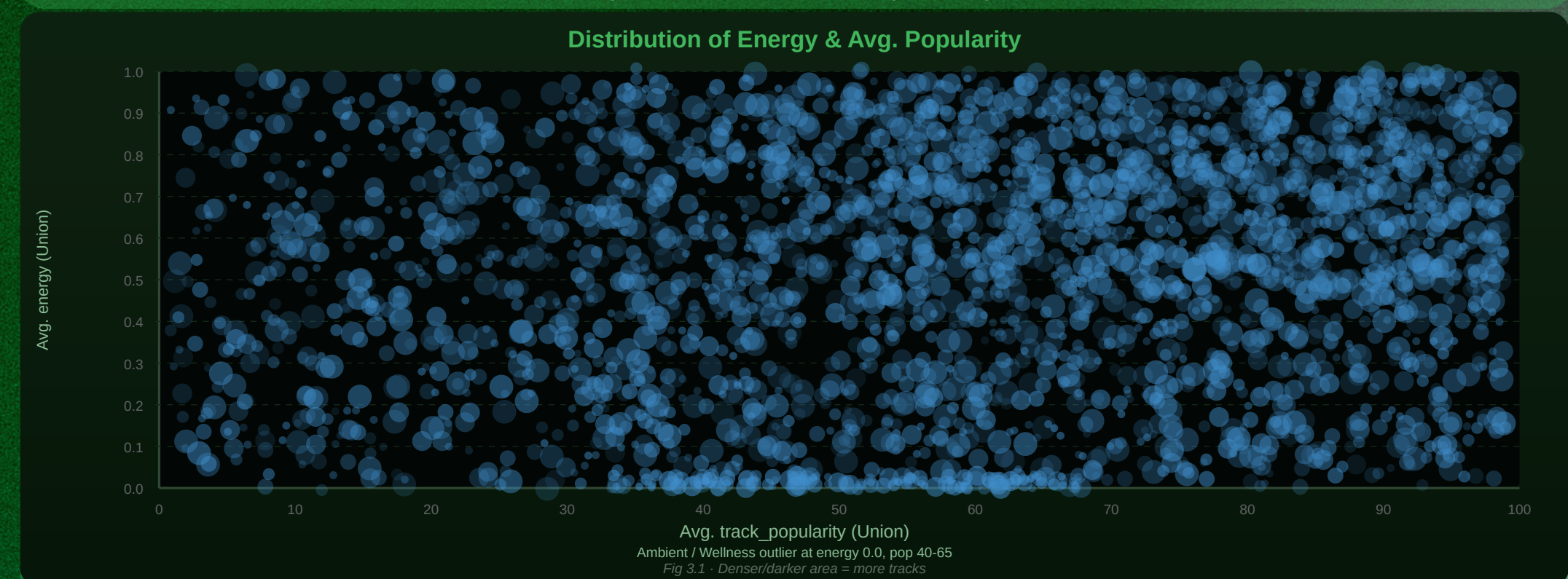


Fig 3.1 - Energy The dense blue cloud shifts heavily toward the upper-right. High-popularity tracks almost always sit at energy 0.5–1.0, meaning more intense songs dominate. The small outlier cluster at energy 0.0 represents ambient and wellness music, which earns moderate popularity through niche, purpose-driven listening rather than mainstream appeal.

KEY FINDINGS SUMMARY

195-285k ms

Optimal Duration Range

110-150 BPM

Top-Performing Tempo

0.5-1.0 Energy

High-Popularity Range

0.4-0.8 Valence

Mid-Positive Tone

DISCUSSION & LIMITATIONS

The analysis shows that song popularity is shaped by multiple musical features rather than a single factor. Mid-length songs (around 3–4 minutes), along with tracks that have moderate-to-high energy and faster tempos, appear more often among popular songs, while valence shows a weaker and less consistent relationship. Some outliers, such as low-energy or ambient tracks, still reach moderate popularity, showing that success does not follow one pattern. However, the study has several limitations, as the dataset only reflects Spotify data and excludes external influences like marketing, playlist placement, and artist popularity. In addition, using averaged values can hide variation between individual tracks, and density graphs may obscure unique or extreme cases, meaning the results show correlation rather than causation.

CONCLUSION & FUTURE WORK

This study shows that song popularity is influenced by a combination of features like duration, energy, tempo, and valence rather than a single factor. While clear patterns appear, they are not absolute, highlighting the complexity of musical success. Future research should expand the dataset by including variables such as genre, release year, and artist popularity, along with external influences like playlist placement and social media trends, to improve accuracy and make the findings more widely applicable.



LINK TO
RESEARCH
PAPER