DATA ANALYTICS REPORT

EXPLORATORY DATA ANALYSIS OF SPOTIFY TRACK PERFORMANCE, PLAYLIST DYNAMICS & STREAMING POPULARITY

Listener Behavior, Playlist Dynamics & Track Popularity Insights

Spotify

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

GITHUB

TABLE OF CONTENTS

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

Spotify playlists and tracks serve a diverse listener base with varying engagement, discovery, and streaming behaviors, making it challenging to identify the key drivers of track popularity and long-term audience growth.

Objective:

To explore the Spotify tracks and playlists dataset to uncover patterns, insights, and trends that can help businesses make data-driven decisions around listener behavior, playlist performance, and streaming popularity.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Spotify Dataset from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: e.g Top 10 Tracks, Genre Distribution, Track Popularity etc
- **Bivariate:** Track_popularity vs danceability, Average Track Popularity by Playlist Genre
- Multivariate Analysis: Track Popularity vs Duration vs Playlist Name

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED

- Which playlist types ("Top," "Hits," "Global") maximize exposure and engagement for featured tracks?
- How do mid-popularity versus breakout tracks contribute to sustained listener engagement and incremental growth?
- What role does danceability play in baseline playlist inclusion versus overall track popularity?
- How do energy levels and tempo ranges influence playlist fit and mainstream appeal?
- To what extent does genre balance and subgenre positioning affect playlist curation and audience reach?
- Which genres and artist features drive higher engagement and can be leveraged for targeted marketing?
- How does track duration align with listener attention spans and replay behavior to optimize commercial performance?
- What strategies are most effective for revitalizing catalog tracks to compete with fresh releases on streaming platforms?

DATA OVERVIEW

The dataset provides insights into Spotify listener behavior, covering track metadata, playlist characteristics, audio features, and streaming activity. It also captures engagement signals, popularity metrics, and playlist positioning indicators that are relevant for understanding track performance, audience growth, and content discovery dynamics.

Data Source: Kaggle

Dataset Size

32,833

Records

23

Features

Purchase Diversity

23,450

Tracks

10,693

Track Artists

6

Genre

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
track_id	Categorical	Unique identifier of the Spotify track
track_name	String	Name of the song
track_artist	String	Name of the artist who performed the song
track_popularity	Numerical (Discrete)	Popularity score of the track (0 to 100), higher values indicate greater popularity
track_album_id	String	Unique identifier of the album containing the track
track_album_name	Categorical	Name of the album to which the track belongs
track_album_release_date	Date / Time	Release date of the album
playlist_name	Categorical	Name of the playlist in which the track appears
playlist_id	String	Unique identifier of the playlist
playlist_genre	Categorical	Genre category assigned to the playlist
playlist_subgenre	Categorical	Subgenre classification of the playlist
danceability	Numerical (Continuous)	Measure from 0.0 to 1.0 indicating how suitable a track is for dancing, higher values mean more danceable

DATA OVERVIEW

Continued...

Dataset Features	Type	Feature Description
energy	Numerical (Continuous)	Measure from 0.0 to 1.0 representing track intensity and activity level
key	Numerical (Discrete)	Estimated musical key of the track as an integer (0 = C, 1 = C♯/D♭, ..., -1 means no key detected)
loudness	Numerical (Continuous)	Overall loudness of the track in decibels (dB), typically between -60 dB and 0 dB
mode	Numerical (Discrete)	Indicates if track is in a major (1) or minor (0) key
speechiness	Numerical (Continuous)	Measure from 0.0 to 1.0 indicating presence of spoken words, higher values mean more spoken content
acousticness	Numerical (Continuous)	Confidence measure (0.0 to 1.0) whether the track is acoustic, higher values mean more acoustic
instrumentalness	Numerical (Continuous)	Probability (0.0 to 1.0) that track contains no vocals, higher means more likely instrumental
liveness	Numerical (Continuous)	Measure (0.0 to 1.0) indicating presence of an audience, higher values suggest live performance
valence	Numerical (Continuous)	Measure (0.0 to 1.0) describing musical positivity, higher values indicate happier or more positive tracks
tempo	Numerical (Continuous)	Estimated tempo of the track in beats per minute (BPM)
duration_ms	Numerical (Discrete)	Duration of the track in milliseconds

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- Several fields (track_name, track_artist, track_album_name, playlist_name, playlist_genre, playlist_subgenre) are stored as strings and should be converted to categorical types for more efficient analysis.
- track_album_release_date is currently stored as a string but should be converted to datetime for proper time-based and recency analysis.

DATA CLEANING & TREATMENT

Inconsistencies were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Categorical fields such as track_name, track_artist, track_album_name, playlist_name, playlist_genre, and playlist_subgenre were converted to appropriate categorical data types.
- track_album_release_date was standardized to datetime format to support efficient and accurate time-based analysis.

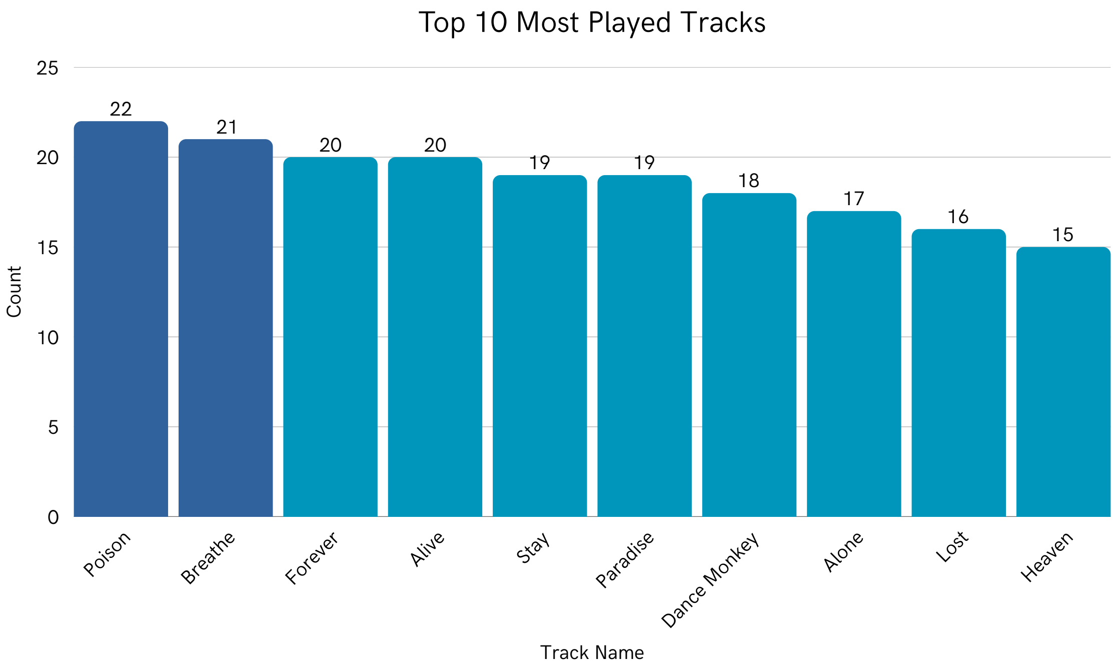
INSIGHTS

Listener Profile & Engagement

A small set of repeat-featured tracks dominate playlist exposure, signaling strong curator bias toward proven hits

23,450

Tracks across 449 Playlists



Key observations

- The top tracks appear 15–22 times across playlists, with a clear drop-off after the top 3–4 songs.
- Playlist representation is highly concentrated, suggesting repeated inclusion of the same popular tracks rather than broad discovery.

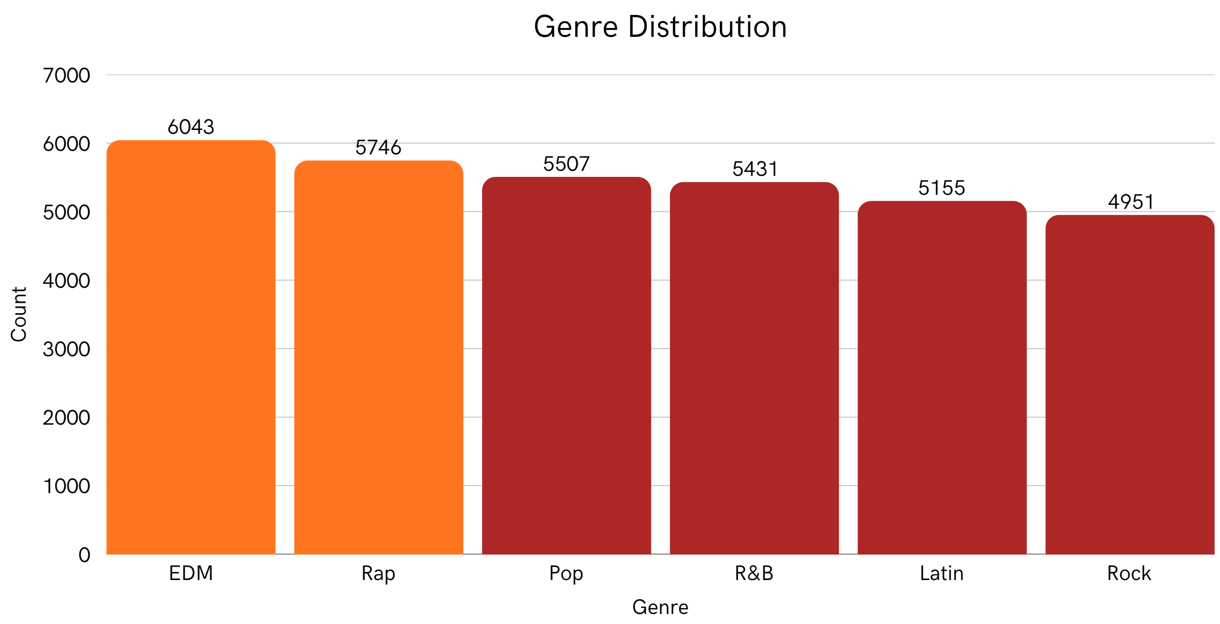
Business Insights

- Playlist visibility is being driven by a winner-takes-most dynamic, where a few tracks capture disproportionate exposure and listening opportunities.
- Increasing playlist rotation for less-featured tracks could improve content discovery without significantly diluting hit-track presence.

Playlist curation is broadly diversified, with EDM and Rap holding a slight but consistent edge in track representation

37.03 %

Of the Tracks EDM & Rap



Key observations

- EDM has the highest playlist representation, closely followed by Rap, Pop, and R&B, with no extreme dominance by any single genre.
- Track counts across genres fall within a relatively narrow range (~5K–6K), indicating balanced genre coverage.

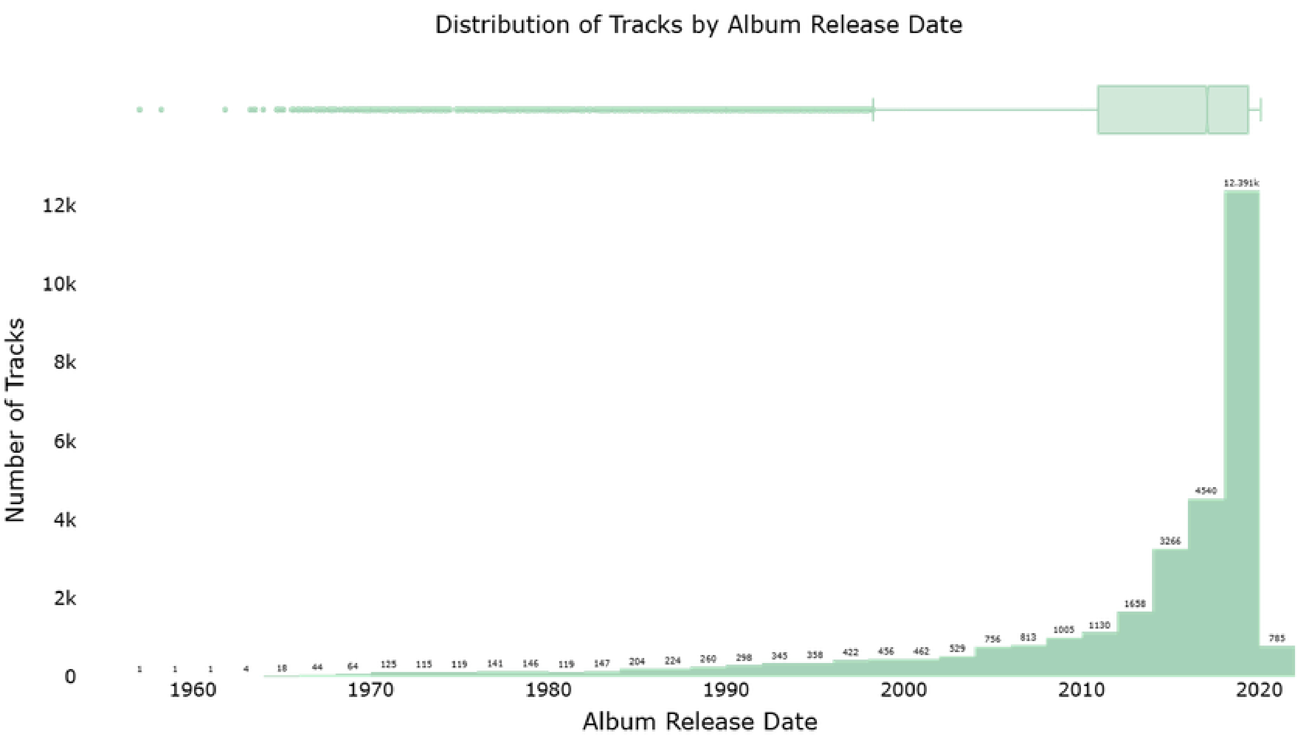
Business Insights

- The absence of a runaway genre suggests curator intent to maintain cross-genre variety, rather than over-indexing on one sound.
- Emerging or niche genres may require distinct positioning or subgenre-focused playlists to gain visibility in an already well-balanced genre landscape.

Playlist catalogs are heavily skewed toward recent releases, with a sharp acceleration after the mid-2010s

75.54 %

Of the Tracks released after 2015



Key observations

- Track counts rise gradually from older decades, then surge sharply after ~2015, peaking in the late 2010s.
- Pre-2000 releases form a small minority of playlisted tracks, indicating limited legacy representation.

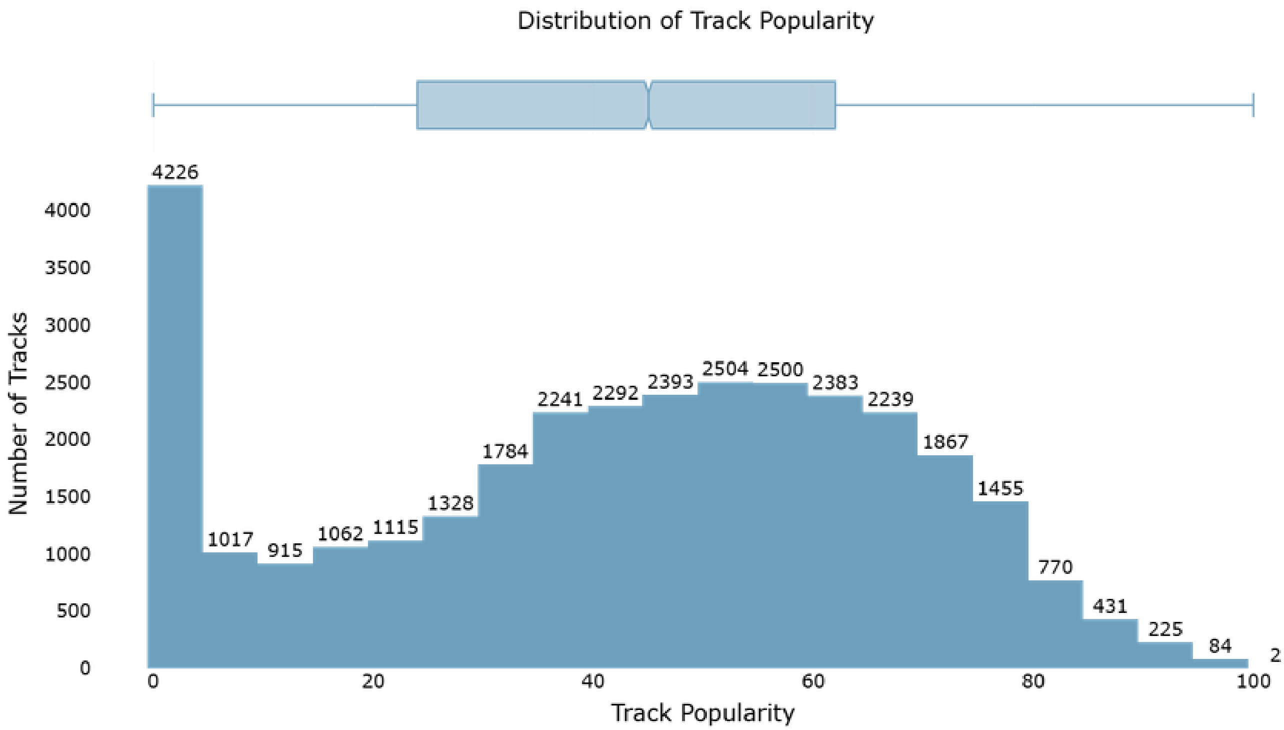
Business Insights

- Playlist curation strongly favors recency, aligning discovery and promotion toward newer music over back catalog depth.
- Older tracks may require explicit throwback or nostalgia-themed playlists to gain visibility in a freshness-driven ecosystem.

Most playlisted tracks achieve moderate popularity, while extreme hits and low visibility songs remain relatively scarce

45.0

Typical Track Popularity (Median)



Key observations

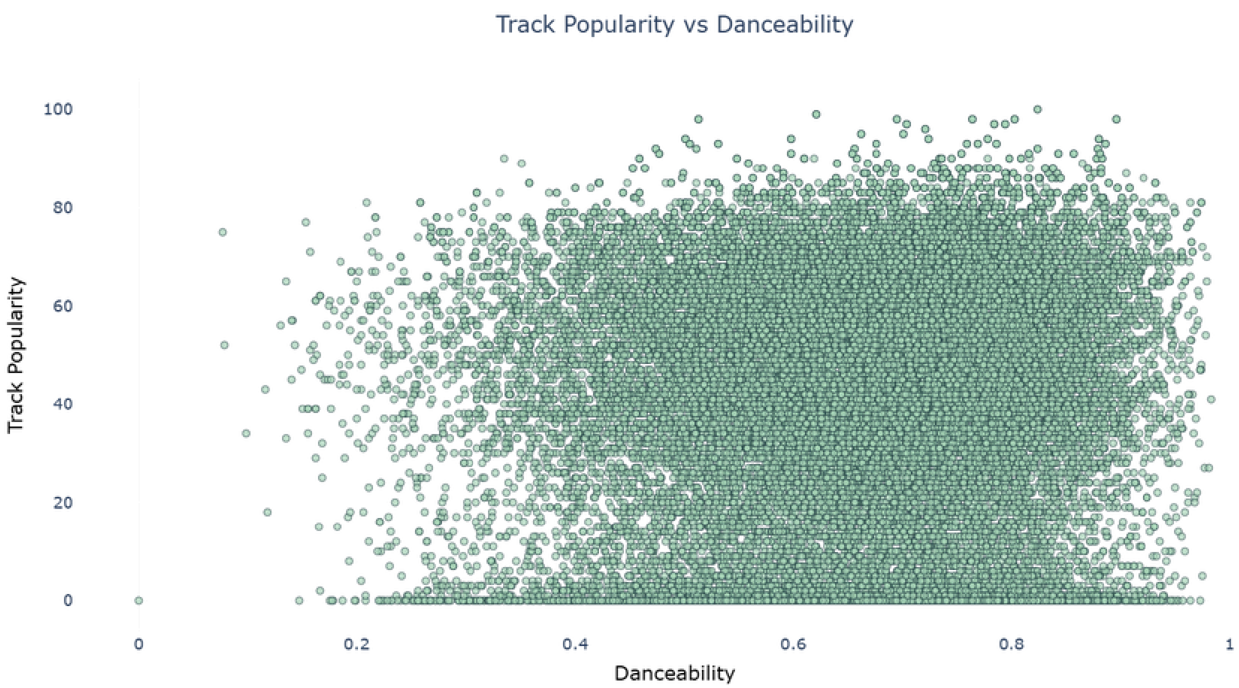
- Track popularity clusters heavily in the 40–65 range, indicating that most playlisted songs are neither niche nor viral extremes.
- Very high-popularity tracks (80+) and very low-popularity tracks (<10) form thin tails of the distribution.

Business Insights

- Playlists tend to favor reliably performing tracks over breakout hits, balancing familiarity with diversity.
- Mid-popularity tracks represent the core opportunity set for sustained playlist exposure and incremental audience growth.

Listening Behavior & Track Performance

Higher danceability supports popularity, but does not guarantee hit-level success



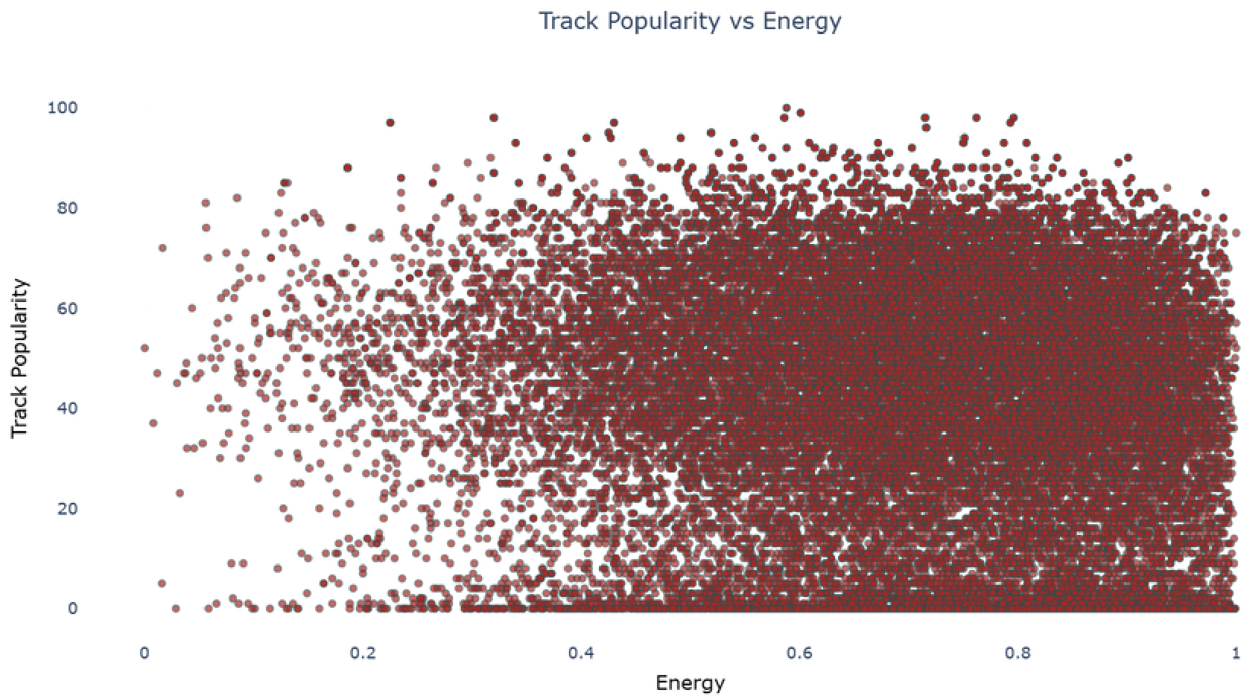
Key observations

- Tracks with moderate to high danceability (≈0.5–0.8) dominate the popularity range, including most high-performing songs.
- Very low danceability tracks rarely achieve high popularity, but high danceability alone shows wide popularity spread.

Business Insights

- Danceability appears to be a necessary but not sufficient condition for popularity, other factors drive breakout success.
- Optimizing for danceable characteristics can improve baseline appeal, but playlist placement and artist reach remain critical for top-tier popularity.

High-Energy Tracks Dominate Volume, But Popularity Is Not Energy-Driven



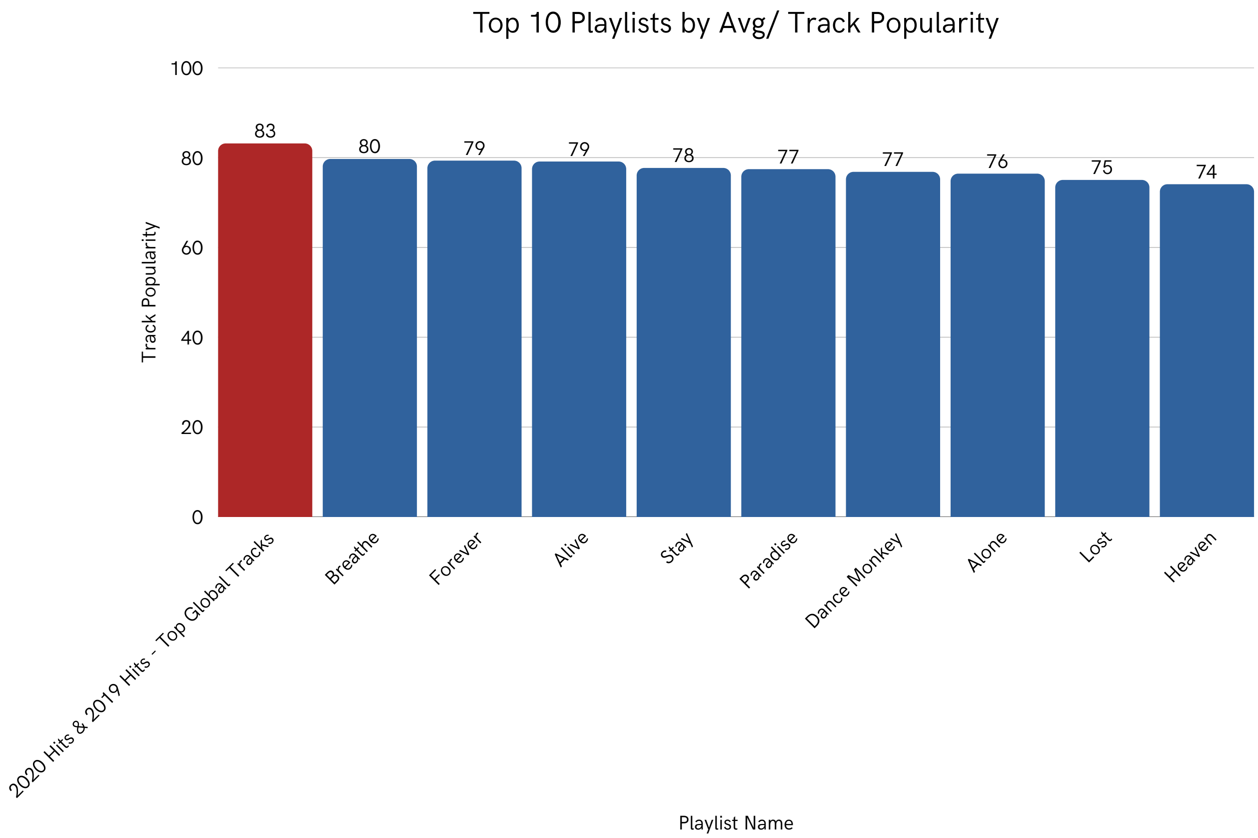
Key observations

- Popularity is widely distributed across all energy levels, showing no strong direct correlation between energy and track success.
- Most tracks cluster in the mid-to-high energy range (0.5–0.9), yet high popularity appears across both moderate and high energy levels.

Business Insights

- Increasing a track’s energy is unlikely to significantly boost popularity on its own; success depends on a combination of audio features and positioning.
- Since the market is saturated with high-energy tracks, differentiation through elements like mood, uniqueness, or playlist targeting is more critical than simply maximizing intensity.

Global & Trend-Based Playlists Lead in Average Track Popularity



Key observations

- “2020 Hits & 2019 Hits - Top Global Tracks” records the highest average track popularity, closely followed by other “Top” and “Most Popular” themed playlists.
- The difference in average popularity among the top 10 playlists is relatively small, indicating consistently high-performing tracks across major curated playlists.

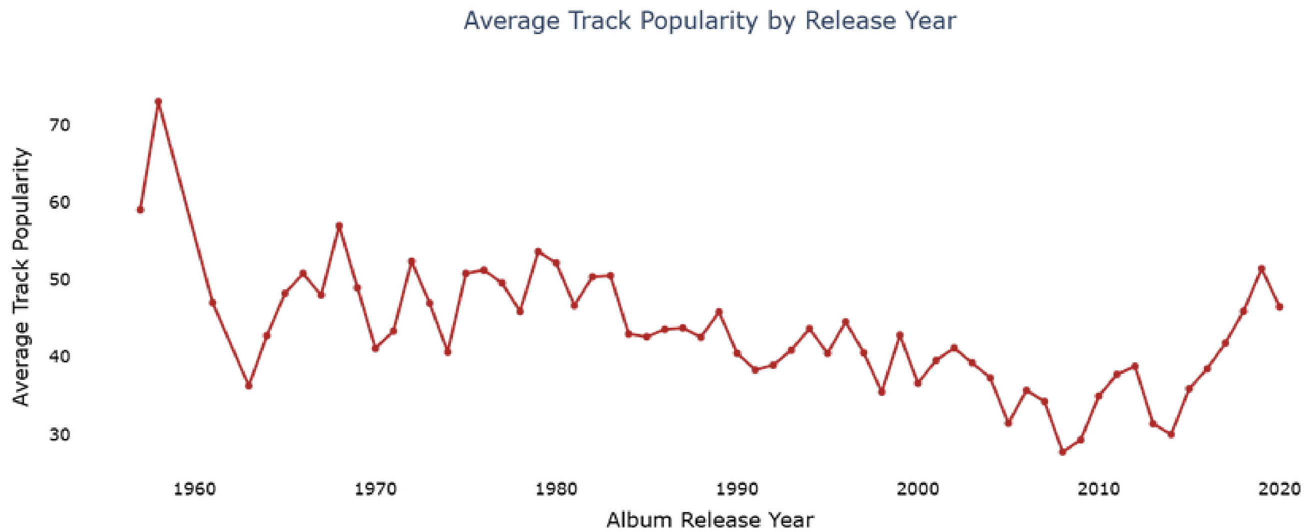
Business Insights

- Playlists positioned around “Top,” “Hits,” and “Global” themes attract and maintain higher average popularity, making them prime targets for label pitching and promotional placement.
- Since top playlists show similar high averages, securing placement in any major flagship playlist can significantly enhance exposure and streaming performance.

Recent Releases Regain Popularity Momentum After Early 2000s Decline

43.24

Typical Popularity over the years (Mean)



Key observations

- Average track popularity declines from the late 1990s through the late 2000s, reaching its lowest levels around 2008–2012.
- Popularity shows a clear upward trend from the mid-2010s onward, with recent albums (2018–2020) achieving significantly higher average popularity scores.

Business Insights

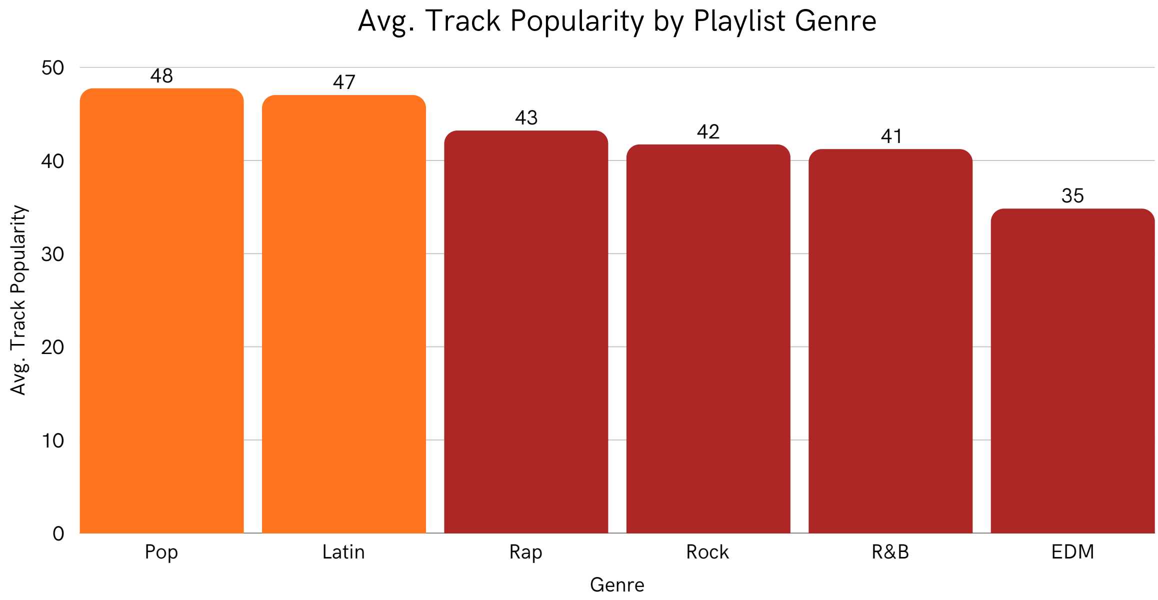
- Newer releases benefit from stronger streaming visibility, algorithmic promotion, and recency bias, making fresh content a key driver of platform success.
- Catalog tracks from the early 2000s may require strategic re-marketing (remasters, playlist reintroduction, social media revival) to compete with the dominance of recent releases.

Role of Genre & Subgenre

Pop and Latin Genres Lead in Average Track Popularity, While EDM Trails Behind

42.62

Typical Popularity per Genre (Mean)



Key observations

- Tracks in the Pop and Latin playlists consistently achieve the highest average popularity scores, indicating strong listener preference.
- In contrast, EDM tracks exhibit noticeably lower average popularity, suggesting less engagement compared to other genres.

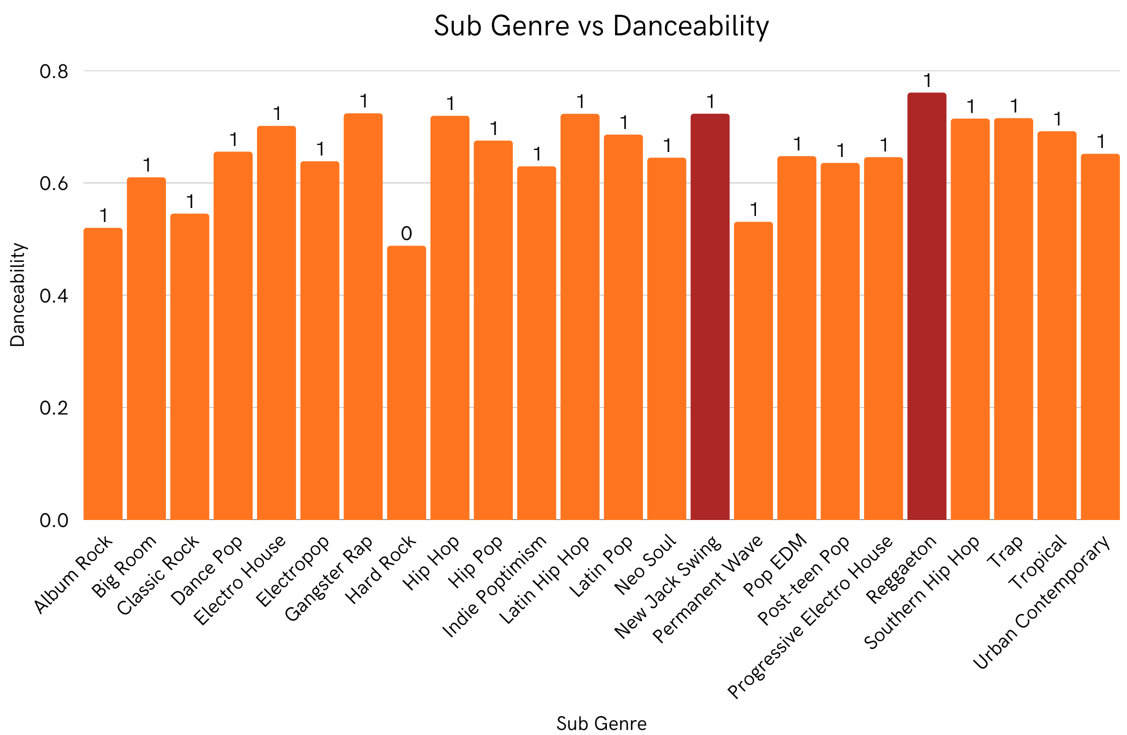
Business Insights

- Investing in marketing and playlist curation for Pop and Latin genres can drive higher user engagement and streaming performance.
- To capture more audience interest, it is important to re-evaluate and enhance EDM playlists with strategies that increase their appeal and popularity.

Reggaeton and New Jack Swing Subgenres Exhibit the Highest Danceability, While Hard Rock and Permanent Wave Lag Behind

0.653

Typical Danceability factor per Sub-Genre (Mean)



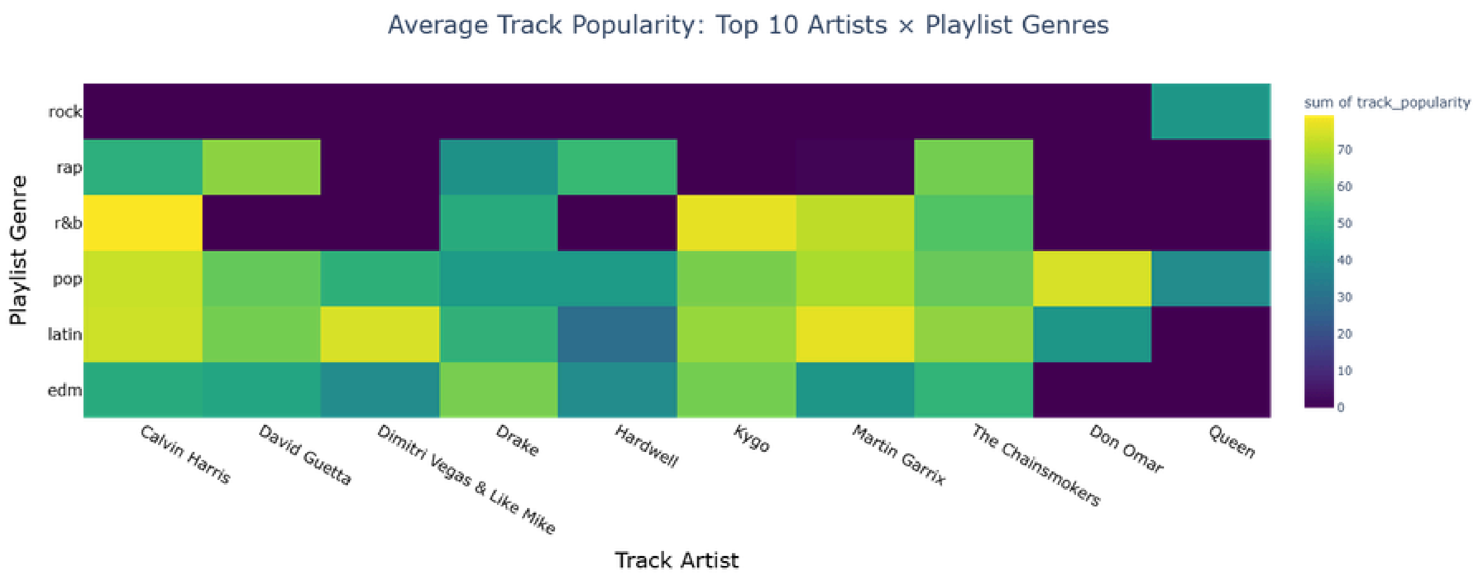
Key observations

- Reggaeton and New Jack Swing subgenres lead with the highest average danceability scores, indicating strong rhythmic appeal for dancing.
- Hard Rock and Permanent Wave show the lowest danceability, suggesting these subgenres are less suited for dance-focused playlists.

Business Insights

- Emphasizing Reggaeton and New Jack Swing in dance-centric playlists can enhance listener engagement and satisfaction.
- Consider repositioning or blending harder-to-dance subgenres like Hard Rock in less dance-focused contexts to optimize playlist performance.

Top Artists Dominate Popularity Differently Across Playlist Genres, Highlighting Genre-Specific Star Power



Key observations

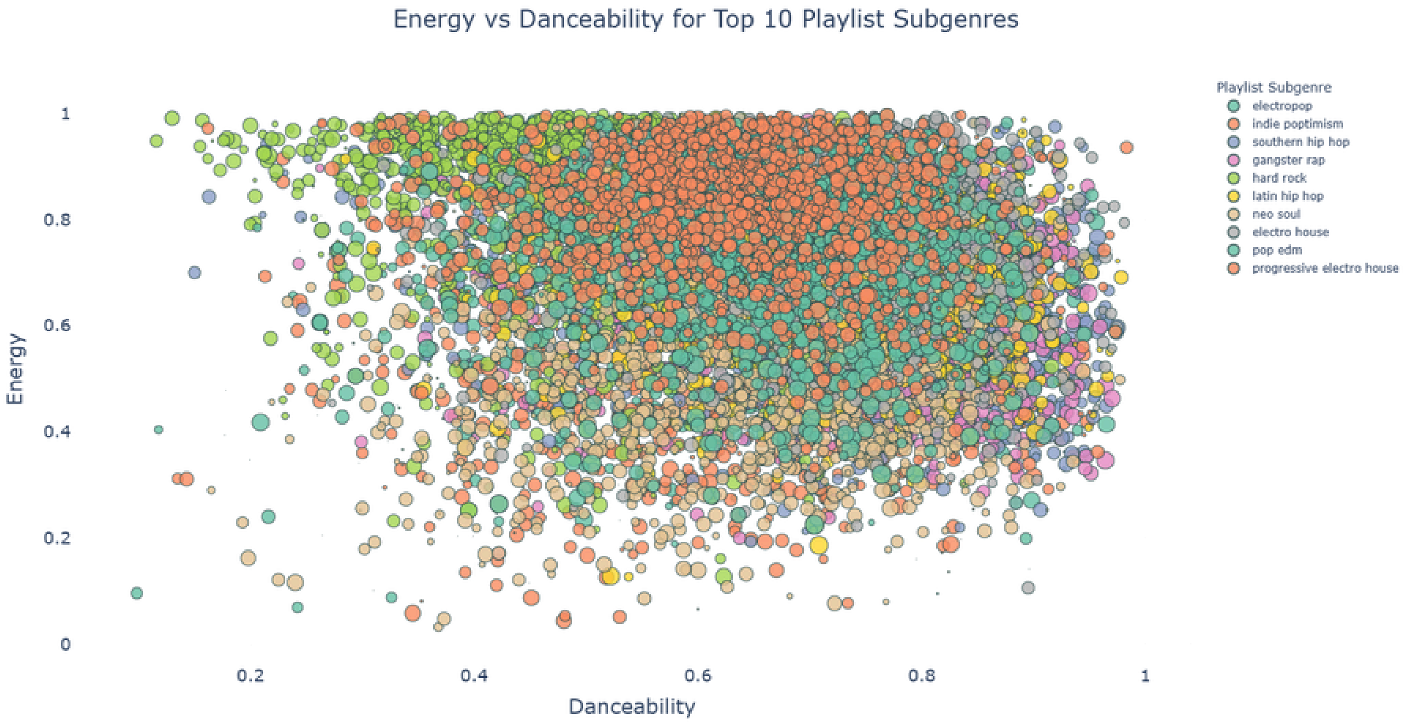
- Artists like Calvin Harris and David Guetta lead in EDM popularity, while Queen dominates the Rock genre.
- Rap and Latin genres show strong popularity for artists like Drake and Don Omar, respectively, reflecting genre-aligned fan bases.

Business Insights

- Leveraging genre-specific top artists in playlists can drive higher listener engagement and targeted marketing success.
- Collaborations or featured tracks with these key artists may boost playlist popularity within their respective genres.

Track Features & Engagement (ROI Drivers)

High Danceability Is the Baseline for Playlist Success, While Energy Defines Genre Positioning



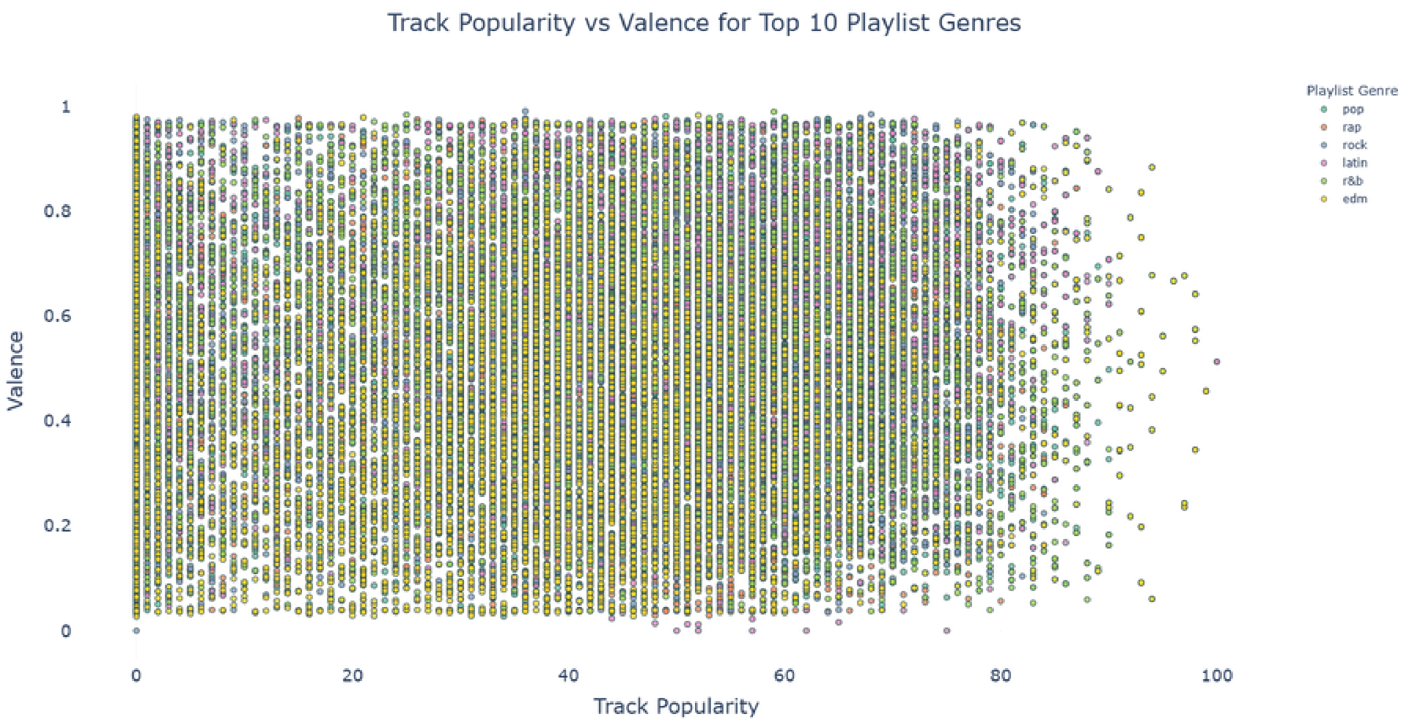
Key observations

- Most top playlist subgenres cluster at moderate-to-high danceability (0.5–0.9), indicating rhythm-driven tracks dominate curated playlists.
- EDM-related genres (progressive electro house, pop EDM, electro house) exhibit consistently higher energy levels, while genres like neo soul and indie poptimism trend toward lower-to-mid energy ranges.

Business Insights

- Danceability acts as a gateway metric for playlist inclusion, making it a critical attribute for artists targeting mainstream streaming growth.
- Energy serves as a strategic differentiator — high-energy tracks align with workout/party playlists, while moderate-energy tracks are better positioned for mood-based or lifestyle playlists, enabling sharper content positioning strategies.

High Popularity Is Achieved Across All Emotional Tones, Not Just Positive Tracks



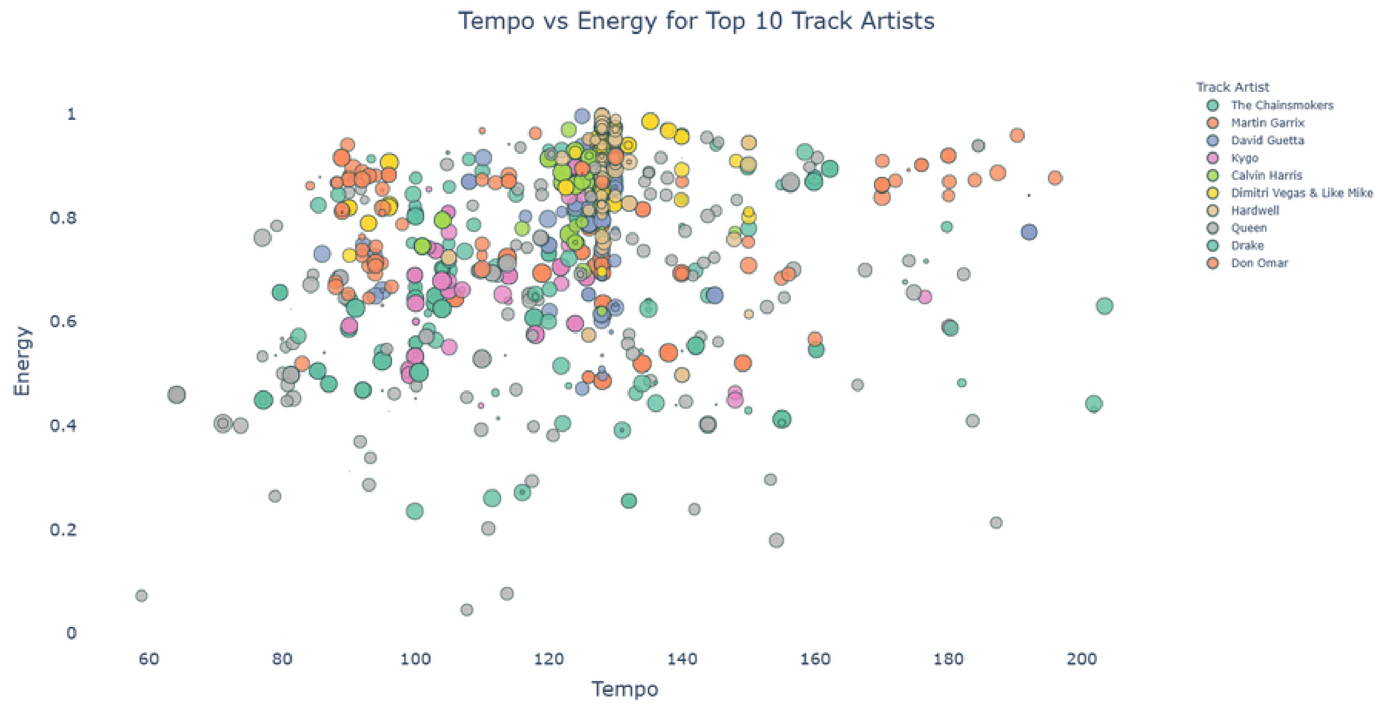
Key observations

- Tracks with high popularity (80–100) are distributed across the full valence spectrum (0–1), indicating both positive and emotionally intense tracks can succeed commercially.
- EDM and pop genres show stronger clustering in mid-to-high popularity ranges, while other genres display more dispersed popularity across valence levels.

Business Insights

- Emotional positivity (high valence) is not a prerequisite for commercial success — labels and artists can confidently promote both upbeat and emotionally heavy tracks.
- Since popularity is not strongly correlated with valence, genre positioning and marketing strategy likely play a larger role than mood alone in driving streaming performance, emphasizing the importance of brand, playlist placement, and audience targeting.

Top Artists Compete in the 100–130 BPM Sweet Spot, Where High Energy Drives Mainstream Appeal



Key observations

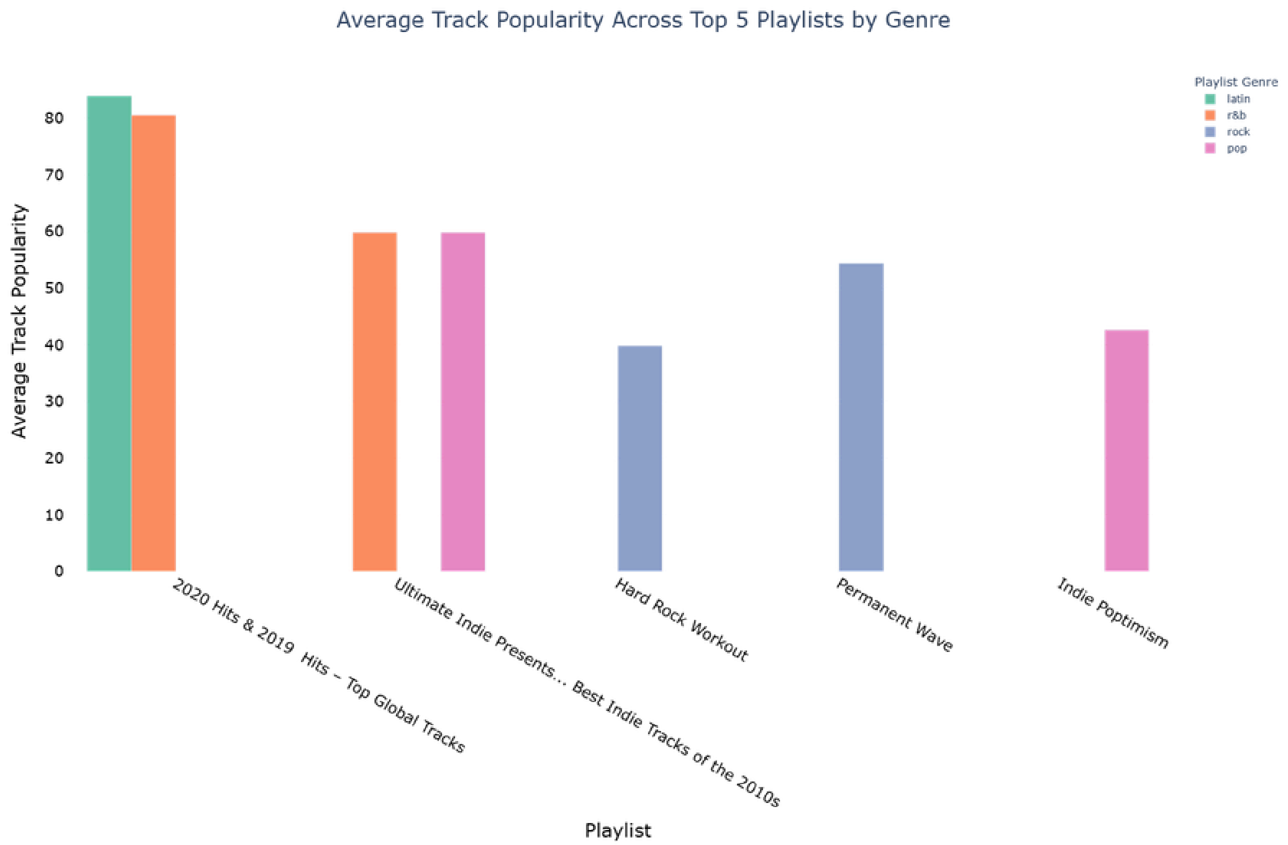
- The majority of tracks from top artists cluster between 100–130 BPM with medium-to-high energy (0.6–1.0), indicating a dominant commercial tempo range.
- High-tempo tracks (>150 BPM) exist but are less concentrated, suggesting extreme tempo is less common among top-performing mainstream artists.

Business Insights

- The 100–130 BPM range appears to be the commercial sweet spot, balancing danceability and energy — making it a strategic benchmark for producers targeting mainstream streaming success.
- Extremely high or low tempos may serve niche audiences, but consistent chart presence favors controlled tempo with strong energy, reinforcing the importance of optimized production formulas for mass appeal.

Satisfaction & Improvement

Mainstream Hit Playlists Significantly Outperform Niche Genre Collections in Track Popularity



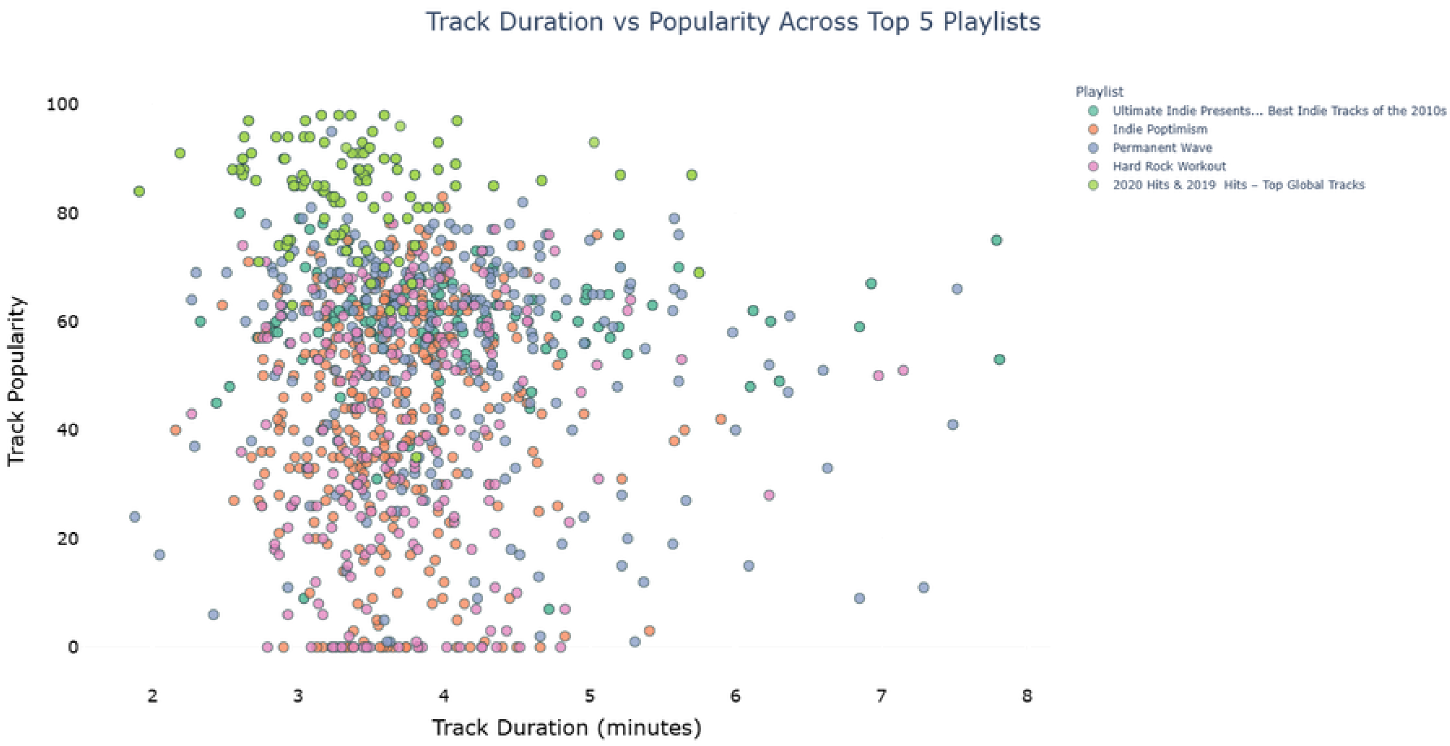
Key observations

- "2020 Hits & 2019 Hits – Top Global Tracks" records the highest average popularity (~80+), outperforming all other playlists by a wide margin.
- Rock and Indie-focused playlists show comparatively lower average popularity (40–60 range), indicating weaker mainstream traction.

Business Insights

- Prioritizing global, multi-genre hit playlists can maximize listener engagement and streaming performance.
- Niche genre playlists (Rock/Indie) may require targeted promotion or repositioning to improve competitiveness and audience reach.

Track Duration Shows Minimal Impact on Popularity, With Most Hits Concentrated Around 3-4 Minutes



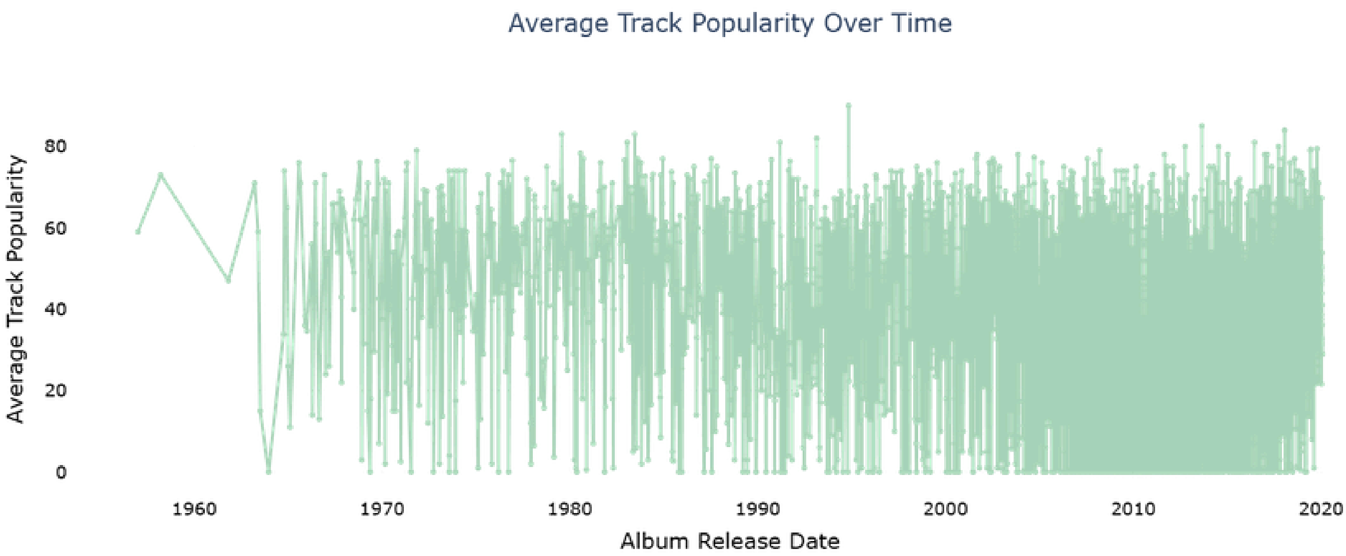
Key observations

- The majority of highly popular tracks (70+ popularity) are clustered between 3 to 4 minutes, indicating a common optimal duration range.
- There is no strong linear relationship between duration and popularity; both shorter and longer tracks show mixed performance outcomes.

Business Insights

- The 3–4 minute duration window appears to be the commercial sweet spot, aligning with listener attention spans and streaming behavior.
- Since longer tracks do not consistently yield higher popularity, optimizing for engagement and replay value may be more impactful than extending track length.

Recent Releases Dominate Popularity, While Older Tracks Show High Volatility



Key observations

- Tracks released after the 2000s show consistently moderate-to-high popularity levels, with fewer extreme low values compared to earlier decades.
- Older releases (pre-1980s) display high variability in popularity, with both very low and occasional high-performing tracks.

Business Insights

- Newer releases benefit from stronger streaming momentum and platform algorithms, indicating that recency plays a key role in driving popularity.
- Catalog (older) music requires strategic curation, remastering, or playlist reintegration to compete effectively with contemporary releases.

BUSINESS / DEVELOPER TAKEAWAYS

- **Hit Concentration Creates Discovery Imbalance:** Playlist exposure follows a winner-takes-most pattern, limiting visibility for emerging tracks and constraining broader catalog monetization.
- **Recency Bias Structurally Favors New Releases:** Newer tracks benefit from algorithmic promotion and curated placement, making freshness a primary driver of streaming success.
- **Flagship "Hits" Playlists Act as Growth Multipliers:** Global and Hits-branded playlists consistently sustain higher popularity averages, making them high-ROI placement targets.
- **Mid-Popularity Tracks Represent the Largest Scalable Opportunity:** Playlists favor reliable performers over viral spikes, positioning mid-tier tracks as the strongest candidates for sustained growth.
- **Danceability Is a Gateway Metric for Inclusion:** High danceability improves baseline playlist eligibility, but breakout success depends more on positioning, brand strength, and reach.

BUSINESS / DEVELOPER TAKEAWAYS

- **Energy and Tempo Drive Contextual Fit, Not Automatic Popularity**: The 100 to 130 BPM range and optimized energy levels support mainstream appeal, but differentiation rather than intensity alone drives top-tier success.
- **Genre Balance Reflects Curated Diversification Strategy**: No runaway genre dominance suggests intentional cross-genre variety, requiring niche genres to pursue sharper positioning strategies.
- **Pop and Latin Show Stronger Commercial Momentum:** These genres demonstrate higher engagement potential, while EDM and harder subgenres may require strategic repositioning.
- **Optimal Track Length Aligns With Streaming Behavior:** The 3–4 minute duration window best matches listener attention patterns and repeat consumption dynamics.
- **Catalog Music Requires Active Re-Marketing:** Older tracks lack algorithmic momentum and must rely on nostalgia positioning, remasters, or curated reintroduction to remain competitive.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset is based on tracked streaming activity and playlist interactions on Spotify and may not capture offline listening or activity on other platforms.
- Listener behavior reflects only observed interactions within the dataset timeframe, potentially underrepresenting long-term trends or infrequent listeners.
- Timestamped track and playlist events provide a snapshot of activity during the observed period, limiting the ability to infer long-term shifts in popularity or engagement beyond the available data window.

Future Work / Opportunities:

- Incorporate listener feedback, ratings, or social engagement metrics (if available) to analyze sentiment and better understand track and playlist perception.
- Extend time-series analysis to longer observation windows to study how track popularity, playlist engagement, and streaming patterns evolve over time.
- Perform deeper comparisons between emerging tracks and established hits to identify characteristics associated with sustained popularity and growth.
- Enrich the dataset with recommendation exposure, playlist placement frequency, or algorithmic promotion metrics to assess their impact on track discovery and streaming performance.

CONCLUSION

This analysis of the Spotify tracks and playlists dataset provides insights into listener behavior, playlist dynamics, and track popularity across genres, audio features, and release timelines. The findings demonstrate how playlist positioning, track characteristics, recency, and genre influence overall streaming performance and audience growth, while future work could extend this through longer-term time-series analysis, listener feedback incorporation, and deeper comparisons between emerging tracks and established hits.

This analysis was conducted using Python, Pandas, Matplotlib, Seaborn and Plotly.

Author:
Tejas Jadhav

GitHub: **@tejas-jadhav**
LinkedIn: **Tejas Jadhav**