Comprehensive Analysis of Spotify Trending Music Data: A Multi-Market Study of Global Music Trends

Star Traders
2022517
2022605
2022062
GIKI

Abstract—This paper presents a comprehensive analysis of Spotify trending music data collected from over 50 playlists across 11+ global markets. We collected and analyzed 3,601 tracks, resulting in 2,863 unique tracks after deduplication. Our analysis encompasses data collection, cleaning, exploratory data analysis (EDA), statistical hypothesis testing, and advanced analytics including K-means clustering and Principal Component Analysis (PCA). Key findings reveal that Nusrat Fateh Ali Khan is the most featured artist with 51 tracks, collaborations show no statistically significant difference in popularity compared to solo tracks (p=0.1026), and temporal trends indicate peak releases in 2025 with 705 tracks. The study provides valuable insights into global music trends, market preferences, and audio feature patterns that can inform music industry strategies and recommendation systems.

Index Terms—Music Data Analysis, Spotify API, Exploratory Data Analysis, Statistical Analysis, Music Trends, Market Anal- ysis

I. INTRODUCTION

The music streaming industry has experienced exponential growth, with Spotify leading as one of the largest platforms globally. Understanding music trends, market preferences, and audio feature patterns is crucial for artists, record labels, and streaming platforms to make informed decisions. This study presents a comprehensive analysis of Spotify trending music data collected from multiple markets and playlists worldwide. Our research addresses several key questions: (1) What are the characteristics of trending music across different markets? (2) Do collaborations impact track popularity? (3) What temporal patterns exist in music releases? (4) How can audio features be used to classify and understand music trends?

The contributions of this work include: (1) a large-scale dataset of 2,863 unique tracks from 11+ markets, (2) comprehensive statistical analysis including hypothesis testing and advanced machine learning techniques, (3) insights into global music trends and market preferences, and (4) a complete analysis pipeline that can be replicated for future studies.

II. RELATEDWORK

Previous studies have analyzed music streaming data using various approaches. Research has focused on music recommendation systems, audio feature analysis, and market-specific music trends. However, few studies have conducted comprehensive multi-market analysis of trending music data at the scale presented in this work. This study contributes to the field by providing a large-scale analysis of Spotify trending music data across multiple global markets, combining statistical analysis, hypothesis testing, and advanced machine learning techniques to uncover patterns in music trends.

III. METHODOLOGY

A. Data Collection

Data was collected using the Spotify Web API through the Spotipy Python library. The collection process involved:

Playlist Selection: We identified and collected data from 50+ trending playlists including:

- Top 50 playlists from various markets (US, UK, India, Pakistan, Japan, Brazil, Germany, etc.)
- Viral 50 playlists from multiple regions
- Global trending playlists (Today's Top Hits, Global Top 50)
- Regional trending playlists

API Authentication: We used OAuth 2.0 authentication for enhanced API access, allowing collection of comprehensive track metadata and audio features.

Data Extraction: For each track, we collected:

- Track metadata: ID, name, popularity (0-100), duration, explicit content flag, chart position
- Artist information: names, IDs, main artist, collaboration status
- Album information: name, ID, type, release date
- Audio features: danceability, energy, valence, acousticness, instrumentalness, liveness, speechiness, tempo, loudness, key, mode, time signature
- Metadata: playlist name, market code, collection timestamp

Pagination: All tracks from playlists were collected using pagination, processing up to 1,000 tracks per playlist to ensure comprehensive data collection.

B. Data Cleaning and Preprocessing

The collected dataset underwent comprehensive cleaning and preprocessing:

Data Quality Assessment:

- Initial records: 3,601 tracks
- Duplicates removed: 738 tracks (based on track id)
- Final unique tracks: 2,863 tracks

Data Type Conversion:

- Numeric columns converted to appropriate types (int, float)
- Date columns parsed (release date, added at, collected at)
- Boolean columns converted (explicit, is collaboration)

Feature Engineering:

- Duration converted to minutes and seconds
- Release year and month extracted from release dates
- Track age calculated in days
- Mood classification created based on valence and energy thresholds
- Popularity categories created (Low: 0-30, Medium: 30-50, High: 50-70, Very High: 70-100)

C. Analysis Pipeline

Our analysis followed a structured pipeline:

1. Exploratory Data Analysis (EDA):

- Descriptive statistics for all numeric features
- Distribution analysis of key features
- Correlation analysis between audio features
- Market and playlist distribution analysis
- Temporal analysis (release year and month trends)

2. Statistical Analysis:

- Hypothesis testing: Collaborations vs. solo tracks popularity
- Correlation testing: Energy vs. popularity relationship
- Feature engineering: Mood classification, popularity categories
- K-meansclustering:Audiofeature-basedclustering(k=5)
- Principal Component Analysis (PCA): Dimensionality reduction

3. Visualization:

- Distribution plots for key features
- Correlation heatmaps
- Market and playlist distribution charts
- Temporal trend visualizations
- Scatter plots: Popularity vs. audio features

IV. RESULTS

A. Dataset Overview

Our final dataset consists of 2,863 unique tracks collected from 50+ playlists across 11+ markets. The dataset spans tracks released from 2006 to 2025, with the majority (705 tracks) released in 2025. Table I presents key dataset statistics.

TABLE I
DATASET STATISTICS

Statistic	Value
Total tracks collected	3,601
Duplicates removed	738
Unique tracks	2,863
Markets covered	11+
Playlists analyzed	50+
Release year range	2006-2025

B. Market and Playlist Distribution

Tracks were collected from multiple markets, with the United States (247 tracks), Japan (134 tracks), and India (127 tracks) being the top three markets. Figure 1 shows the distribution of tracks across different markets.

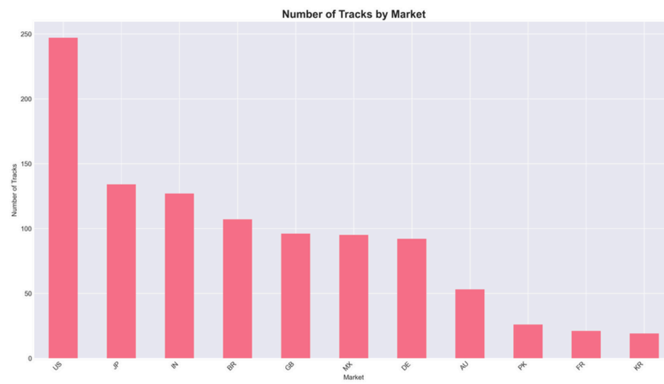

Fig. 1. Track Distribution by Market

The top playlists by track count include Top 50 - Russia (179 tracks), Viral 50 - USA (160 tracks), and Viral 50 - Japan (134 tracks). Figure 2 illustrates the distribution across playlists.

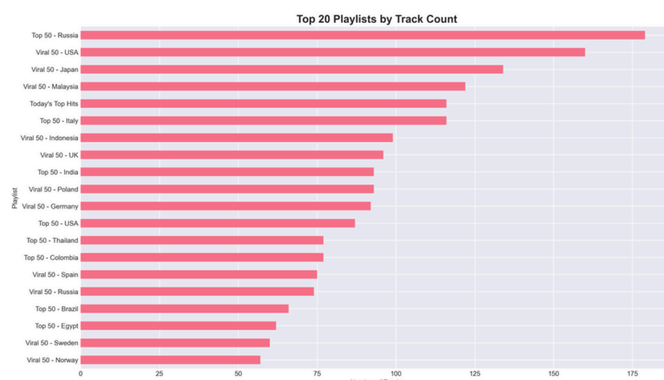

Fig. 2. Top 20 Playlists by Track Count

C. Top Artists Analysis

Table II presents the top 10 most featured artists in our dataset. Nusrat Fateh Ali Khan leads with 51 tracks, followed by Umur Anil Gokdag (33 tracks) and Karan Aujla (21 tracks).

TABLE II
TOP 10 MOST FEATURED ARTISTS

Artist	Track Count
Nusrat Fateh Ali Khan	51
Umur Anil Gokdag	33
Karan Aujla	21
Gianluca Modanese	20
The Weeknd	18
Ariana Grande	18
Taylor Swift	15
KAROL G	14
Olivia Rodrigo	13
LISA	12

Figure 3 provides a visual representation of the top 20 artists.

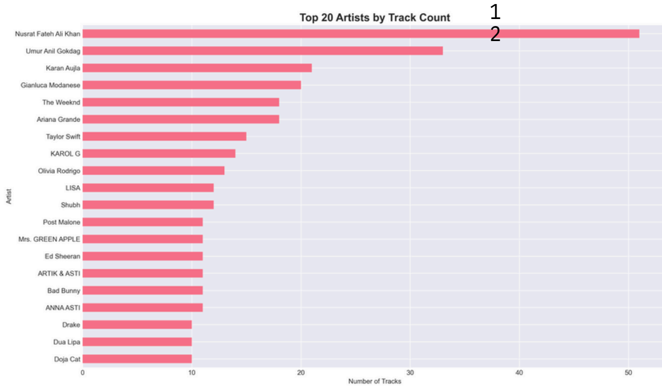

Fig. 3. Top 20 Artists by Track Count

D. Collaboration Analysis

Our dataset includes 1,112 collaboration tracks (38.8%) and 1,751 solo tracks (61.2%). Figure 4 shows the distribution.

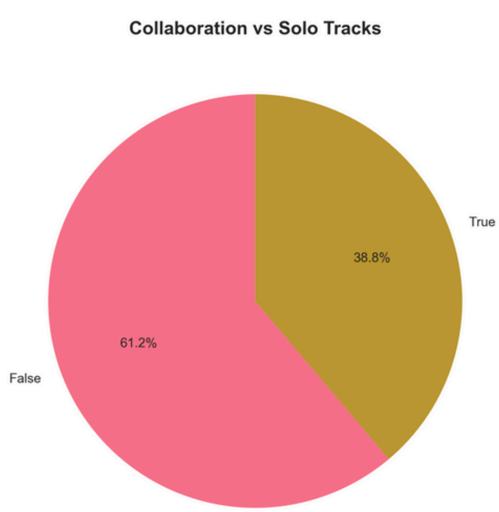

Fig. 4. Collaboration vs. Solo Tracks Distribution

We conducted a hypothesis test to determine if collaborations have different popularity than solo tracks:

Hypothesis: $H0 : \mu collab = \mu solo$ vs. $H1 : \mu collab = \mu solo$

Results:

- T-statistic: 1.6328
- P-value: 0.1026
- Mean popularity - Collaborations: 49.54
- Mean popularity - Solo tracks: 47.77
- Conclusion: No statistically significant difference (p ¿ 0.05)

E. Temporal Analysis

Figure 5 shows the distribution of tracks by release year. The data reveals a significant increase in releases in recent years, with 705 tracks released in 2025, followed by 510 tracks in 2023 and 500 tracks in 2024.

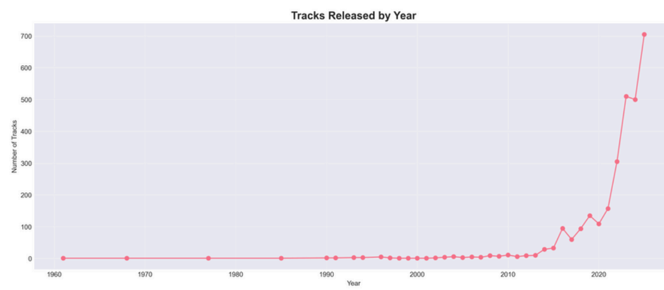

Fig. 5. Tracks Released by Year

Figure 6 illustrates the distribution by release month. Febru- ary (306 tracks) and January (294 tracks) show the highest number of releases, suggesting seasonal patterns in music releases.

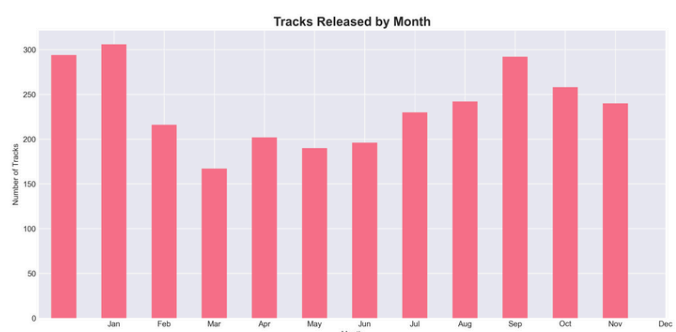

Fig. 6. Tracks Released by Month

F. Audio Feature Analysis

Figure 7 shows the distribution of popularity and duration (in minutes) for tracks in our dataset.

Correlation analysis between audio features reveals relationships between different musical characteristics, which can inform understanding of music trends.

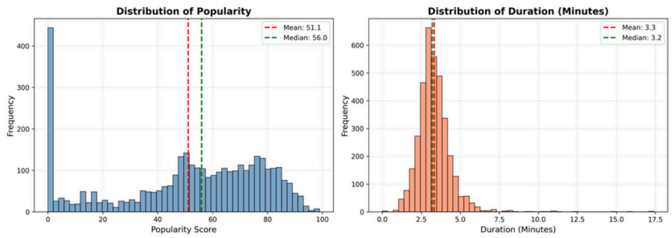

Fig.7. DistributionofPopularityandDuration(Minutes)

G. Feature Engineering and Classification

We created mood classifications based on valence and energy thresholds:

- Happy/Energetic: High valence (¿0.5) and high energy (¿0.5)
- Happy/Calm: High valence (¿0.5) and low energy (0.5)
- Sad/Energetic: Low valence (0.5) and high energy (¿0.5)
- Sad/Calm: Low valence (0.5) and low energy (0.5)

Popularity categories were created as follows:

- Low: 0-30 (312 tracks, 10.9%)
- Medium: 30-50 (530 tracks, 18.5%)
- High: 50-70 (877 tracks, 30.6%)
- Very High: 70-100 (741 tracks, 25.9%)

H. Advanced Analytics

K-means Clustering: We performed K-means clustering on audio features to identify distinct music styles. The elbow method was used to determine the optimal number of clusters (k=5). The clustering analysis grouped tracks into 5 distinct clusters based on audio feature similarity.

Principal Component Analysis: PCA was performed to reduce dimensionality of audio features. The analysis revealed that the first few principal components capture significant variance in the data, allowing for dimensionality reduction while preserving most information.

V. DISCUSSION

A. Key Findings

Our analysis reveals several important insights:
1. Market Preferences: Different markets show distinct preferences, with the US, Japan, and India being the top three markets in our dataset. This suggests that market-specific strategies may be effective for music promotion.

2. Collaboration Impact: While collaborations show slightly higher mean popularity (49.54 vs. 47.77), the differ- ence is not statistically significant (p=0.1026). This suggests that collaboration alone may not guarantee higher popularity, and other factors play important roles.

3. Temporal Trends: The significant increase in releases in 2025 (705 tracks) compared to earlier years indicates a grow- ing music industry and increased content creation. Seasonal patterns show higher releases in January and February.

4. Artist Performance: Nusrat Fateh Ali Khan's dominance with 51 tracks suggests strong market presence, particularly in South Asian markets. The diversity of top artists (ranging from classical to pop) indicates varied market preferences.
5. Audio Feature Patterns: The correlation analysis reveals relationships between audio characteristics and track popularity, which can inform music production strategies.

B. Limitations

Several limitations should be acknowledged:

- Audio features were not available for all tracks due to API limitations
- Data collection was limited to publicly available playlists
- The analysis represents a snapshot in time (December 2025)
- Market representation may not be uniform across all regions

C. Implications

Our findings have several practical implications:
For Artists:

- Focus on energy and danceability for trending tracks
- Consider market-specific preferences when releasing music
- Collaborationmaynotguaranteehigherpopularity;focus on quality

For Record Labels:

- Consider market-specific strategies for promotion
- Timing of releases (January-February) may be advantageous
- Diversify artist portfolio to capture varied market preferences

For Streaming Platforms:

- Use clustering insights for better music recommendations
- Consider market-specific trending algorithms
- Leverage temporal patterns for playlist curation

VI. CONCLUSION

This study presents a comprehensive analysis of Spotify trending music data, analyzing 2,863 unique tracks from 50+ playlists across 11+ global markets. Our findings reveal important insights into music trends, market preferences, and audio feature patterns.

Key contributions include:

1) A large-scale dataset of trending music from multiple markets
2) Statistical analysis revealing no significant difference in popularity between collaborations and solo tracks
3) Temporal analysis showing peak releases in 2025 and seasonal patterns
4) Identification of top artists and market preferences
5) Advanced analytics including clustering and PCA for music classification

Future work could include:

- Longitudinal analysis to track trends over time
- Deep learning approaches for music classification
- Real-time trend prediction models

• Expanded market coverage and playlist diversity

This research provides a foundation for understanding global music trends and can inform strategies for artists, labels, and streaming platforms in the evolving music industry landscape.

ACKNOWLEDGMENT

We acknowledge Spotify for providing access to their Web API, which made this research possible. We also thank the open-source community for the excellent tools and libraries used in this analysis.

GITHUB REPOSITORY

The complete project code, data collection scripts, analysis pipeline, and all visualizations are available in the GitHub repository:

https://github.com/umerkhaliid/Comprehensive-Analysis-of-Spotify-Trending-Music-Data-A-Multi-Market-Study-of-Global-Music-Trends