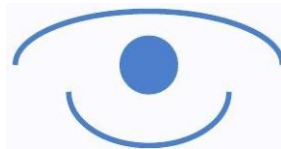# The Sound of Success

Insights into Genre Popularity and Listener Preferences

Author: Ofer Assaf

Title: Junior Data Analyst

Certification: Google Data Analytics Professional Certificate

Date: January 10th, 2025

Contact: oferassaf@gmail.com

Table of Contents

1.   <u>Executive Summary</u>

My case study examines Spotify track data to uncover patterns in genre popularity and audio features. Key findings include the strong correlation between high energy and danceability with genre popularity, and the role of cultural and niche targeting in driving listener engagement. Recommendations provide actionable insights for producers, artists, marketers, and playlist curators to maximize their impact.

Task:

This analysis bridges Spotify track data with actionable insights for the music industry. By identifying trends in genre popularity and uncovering the role of audio features like danceability and energy, this case study offers guidance for producers, marketers, and playlist curators. Key questions include:

- What makes a genre successful on Spotify?
- How do audio features like danceability and tempo vary across genres?
- How can these insights inform strategic decisions?

2.   Key Components

2.1   Introduction

An overview of the dataset, objectives, and the importance of understanding genre popularity and audio features in the music industry.

2.2   Cleaning Process

A detailed summary of the cleaning steps, including:
- Removal of duplicate entries to retain the most popular track versions.
- Handling of missing values in critical columns like genre and popularity.
- Standardization of data types and validation of genre consistency to ensure reliable and accurate analysis.

2.3   Analysis Results

- Insights into genre trends, including the identification of top-performing genres and the role of audio features like danceability and energy.
- Charts and tables are used to visualize trends in genre popularity, feature correlations, and the characteristics of both high- and low-performing genres.

2.4   Conclusions

- A synthesis of findings highlighting key trends, such as the dominance of energetic and danceable tracks and the unique appeal of niche genres.
- Discussion of the broader factors influencing genre success, including cultural and branding elements.

2.5   Recommendations

Tailored strategies for mainstream and niche genres to enhance performance and engagement, along with general insights for identifying emerging trends and refining music production strategies.

3.   Cleaning Section

Introduction to Cleaning:

The dataset contains Spotify track information, including genres, popularity, and audio features. Cleaning is essential to ensure reliable analysis and meaningful insights.

3.1   Remove Duplicate Rows

Duplicate rows can skew results, especially if a track is listed multiple times under the same genre. I removed duplicates while keeping the row with the highest popularity.

Query:

```
CREATE TABLE
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_cleaned` AS
WITH ranked_tracks AS ( SELECT *,
ROW_NUMBER() OVER (PARTITION BY track_id, track_genre ORDER BY popularity
DESC) AS row_num
FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks`)
SELECT *
FROM ranked_tracks WHERE row_num = 1;
```

Why This is Important:

- Keeps only one version of each track_id and track_genre combination.
- Prioritizes the row with the highest popularity.

3.2   Check and Remove Missing Values

I will identify and handle missing values in critical columns, such as track_genre or popularity that can disrupt the analysis.

Query to Identify Missing Values:

```
SELECT
COUNT(*) AS total_rows,
SUM(CASE WHEN track_genre IS NULL THEN 1 ELSE 0 END) AS
null_genres,
SUM(CASE WHEN popularity IS NULL THEN 1 ELSE 0 END) AS
null_popularity
FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_cleaned`;
```

Query to Remove Rows with Missing Critical Values:

```
CREATE TABLE
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_no_missing` AS
SELECT * FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_cleaned`
WHERE track_genre IS NOT NULL AND popularity IS NOT NULL
```

Why This is Important:

- Ensures all rows have the necessary data for analysis.
- Prevents distortion in genre and popularity trends.

3.3   Standardize Data Types

Standardizing data types ensures consistency and prevents issues during analysis.

Query:

```
CREATE TABLE
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_standardized` AS
SELECT
SAFE_CAST(track_id AS STRING) AS track_id, SAFE_CAST(track_genre AS STRING)
AS track_genre, CAST(popularity AS FLOAT64) AS popularity, CAST(danceability AS
FLOAT64) AS danceability, CAST(tempo AS FLOAT64) AS tempo, CAST(duration_ms
AS FLOAT64) AS duration_ms, CAST(energy AS FLOAT64) AS energy,
CAST(valence AS FLOAT64) AS valence,

-- Include all other columns except the ones standardized

* EXCEPT(track_id, track_genre, popularity, danceability, tempo, duration_ms, energy,
valence)
FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_no_missing`;
```

Why This is Important:

- Ensures numeric features like popularity and tempo are correctly formatted as FLOAT64.
- Treats track_id and track_genre as categorical data (STRING)

Results: Columns standardized without affecting row count.

Cleaning Summary:

Step	Rows Before	Rows After
Original Dataset	114.000	114,000
After Removing Duplicates	114.000	113,556
After Removing Missing Values	113,556	113,550
Final Cleaned Dataset	113,550	113,550

After the cleaning process, the dataset was reduced from **114,000 rows to 113,550 rows**.

- Duplicate removal eliminated **444 rows**.
- Missing value handling removed **6 rows**.
- Data type standardization preserved all rows.

3.4   Validate Genre Consistency

Ensure the track_genre column does not contain misspellings or inconsistent formatting.

Query:

SELECT DISTINCT track_genre FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_standardized`
ORDER BY track_genre;

Results: No inconsistencies. The query results showed a complete list of unique genre names, all correctly spelled and consistently formatted, without any duplicates, null values, or unexpected entries. This confirms that the track_genre column is clean and free of any inconsistencies, ensuring reliable and accurate analysis.

Query results

JOB INFORMATION | RESULTS

Row	track_genre ▼
1	acoustic
2	afrobeat
3	alt-rock
4	alternative
5	ambient
6	anime
7	black-metal
8	bluegrass
9	blues
10	brazil
11	breakbeat
12	british
13	cantopop
14	chicago-house

4.   Organizing and Analysis Stage

Introduction

The analysis phase aims to uncover patterns in genre popularity and explore how audio features vary across genres. Visualizations are used to complement the findings, making trends more accessible and actionable.

4.1   Overall Genre Popularity

Goal: Identify the most popular genres across the dataset by calculating their average popularity.

Action:
- I grouped the data by track_genre and computed metrics, such as average popularity and the number of tracks for each genre.
- I ranked genres by their average popularity.

Query:

```
SELECT track_genre, AVG(popularity) AS avg_popularity, COUNT(*) AS
track_count
FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_standardized`
GROUP BY track_genre
ORDER BY avg_popularity DESC;
```

Results:

The analysis reveals that **Pop-Film** is the most popular genre, with an average popularity score of 59.3, followed by **K-Pop** and **Chill**. These genres likely owe their popularity to mainstream appeal and listener preferences for specific moods or cultural content. On the other hand, niche genres like **Grunge** and **Indian** maintain steady popularity despite lower track representation.

Genre	Avg Popularity	Track Count
Pop-Film	59.28	999
K-Pop	56.91	999
Chill	53.70	999

Average Popularity of Genres
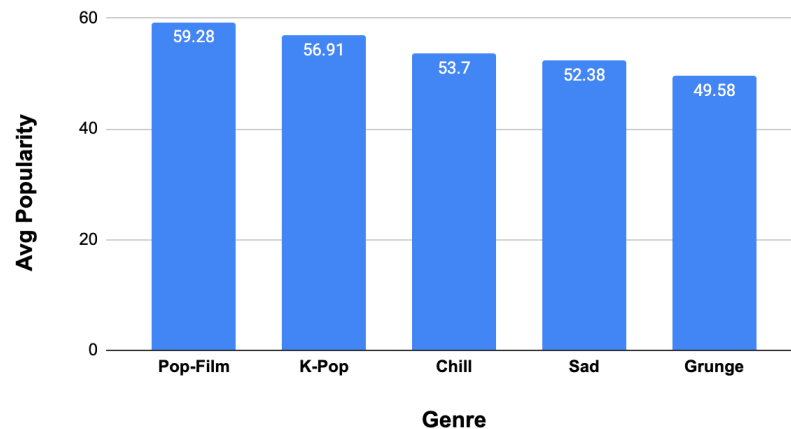
4.2   Audio Features by Genre

Goal: Explore how audio features (e.g., danceability, energy) vary across genres to uncover defining characteristics of popular genres.

Action: I grouped the data by t**rack_genre** and calculated the average values for features like **danceability, energy, tempo,** and **duration_ms.**
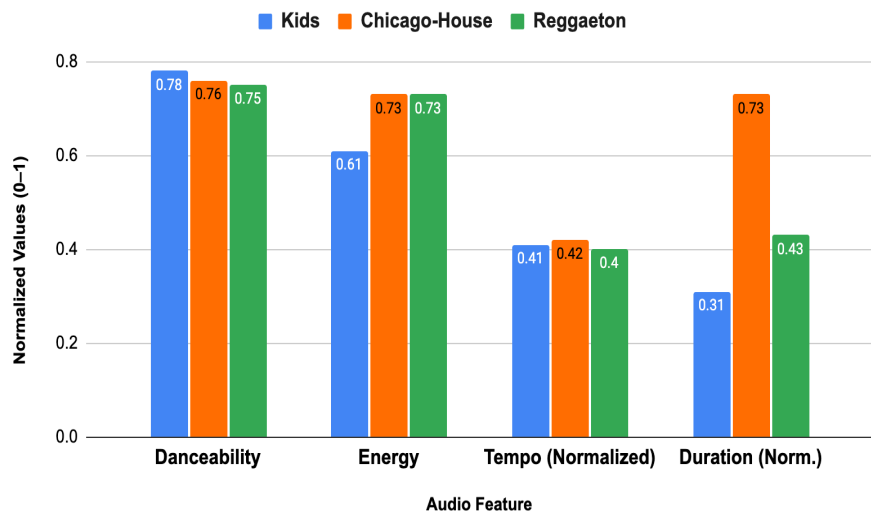
Query:

```
SELECT
track_genre,
AVG(danceability) AS avg_danceability, AVG(energy) AS avg_energy,
AVG(tempo) AS avg_tempo, AVG(duration_ms) AS avg_duration
FROM
`case-study-trends-in-music.spotify_tracks_dataset.Spotify_tracks_standardized`
GROUP BY track_genre
ORDER BY avg_danceability DESC;
```

Results:

Genres such as **Chicago-House** and **Reggaeton** exhibit high danceability and energy, making them ideal for danceable and high-intensity playlists. Meanwhile, **Kids** music stands out with high danceability but lower energy, reflecting its rhythmic simplicity. This analysis helps identify the defining characteristics of popular genres.

Audio Feature	Kids	Chicago-House	Reggaeton
Danceability	0.78	0.76	0.75
Energy	0.61	0.73	0.73
Tempo (Normalized)	0.41	0.42	0.4
Duration (Norm.)	0.31	0.73	0.43

Comparison of Audio Features Across Genres

4.3   <u>**Expanded Analysis: Low-Performing Genres and Audio Features**</u>

To validate the observed trends for high-performing genres, I conducted a mirrored analysis on low-performing genres to determine if lower energy and danceability correlate with lower popularity scores.

Query:

```
SELECT
    track_genre,
    AVG(popularity) AS avg_popularity,
    AVG(danceability) AS avg_danceability,
    AVG(energy) AS avg_energy,
    AVG(tempo) AS avg_tempo,
    AVG(duration_ms) AS avg_duration
FROM
    `case-study-trends-in-
music.spotify_tracks_dataset.Spotify_tracks_standardized`
GROUP BY track_genre
ORDER BY avg_popularity ASC
LIMIT 10;
```

Results:

- Genres with the lowest average popularity, such as *Iranian* (2.22) and *Romance* (3.55), exhibit significantly lower average danceability (0.30 and 0.43, respectively) and energy (0.54 and 0.29, respectively).
- By contrast, *Chicago-House* (a higher-performing genre) has a danceability score of 0.77 and energy level of 0.73, further reinforcing this contrast.

Audio Feature	Iranian	Romance
Danceability	0.30	0.43
Energy	0.54	0.29
Tempo (Normalized)	0.47	0.45
Avg. Popularity	2.22	3.54

Audio Features Comparison Across Low-Performing Genres

■ Iraninan   ■ Romance
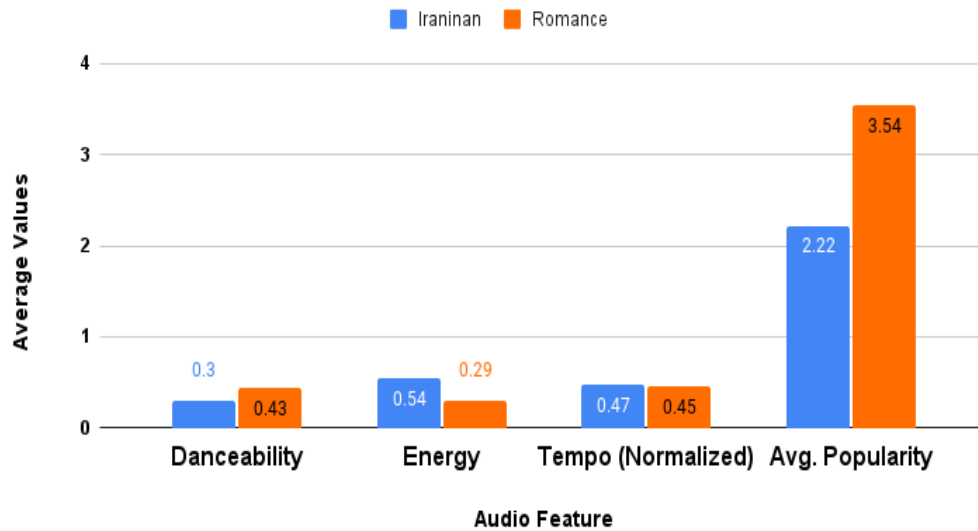
Key Observations:

- **Danceability Correlation:** Low-performing genres consistently have reduced danceability, indicating a weaker rhythmic appeal that may deter listener engagement.
- **Energy Correlation:** Energy levels in low-performing genres are markedly lower, highlighting the importance of energetic tracks in capturing audience attention.
- **Other Features:** Tempo and duration show less consistent patterns, suggesting they are secondary factors compared to energy and danceability.

Conclusion: My analysis confirms that the defining characteristics of high-performing genres (high energy and danceability) are inversely mirrored in low-performing genres. This strengthens the conclusion that energy and danceability are pivotal audio features driving Spotify's popularity.

5. **_Synthesizing Key Patterns_**

My analysis of genre popularity and audio features reveals meaningful patterns that explain the success and appeal of certain genres. By correlating insights from high-performing and low-performing genres, a clear relationship emerges between listener preferences and the characteristics of different genres.

- **Danceability and Energy Drive Popularity**
  - Popular genres such as Chicago-House and Reggaeton exhibit high danceability (0.76 and 0.75, respectively) and energy (0.73 for both). These attributes align with their frequent use in high-energy settings like parties and clubs.
  - Conversely, low-performing genres such as Iranian and Romance demonstrate significantly lower danceability (0.30 and 0.43) and energy (0.54 and 0.29). This contrast underscores the critical role of rhythmic and energetic qualities in driving popularity.

- **Niche Genres Cater to Specific Audiences**
  - Genres like Kids music focus on a distinct audience by emphasizing high danceability but maintaining lower energy levels. This combination appeals to younger listeners through simplicity and accessibility.
  - Similarly, low-performing genres often cater to niche audiences or rely on unique cultural contexts. For example, Iranian and Romance may resonate with specific demographics but lack the broad appeal needed to achieve higher popularity scores.

- **Popularity is Not Always Feature-Driven**
  - While audio features like danceability and energy significantly influence popularity, external factors such as cultural trends, branding, or associations with visual media also play a vital role. For example, Pop-Film and K-Pop achieve high popularity despite moderate scores in audio features.
  - The low-performing analysis confirms this perspective: Iranian and Romance genres may have cultural or emotional resonance for specific listeners, even though their audio features do not align with mainstream preferences.

6. Key Takeaways

From this synthesis, three overarching trends emerge:

- **Energetic and Danceable Tracks Perform Well Globally**:
  Genres like Chicago-House and Reggaeton thrive due to their upbeat and engaging nature, reflecting listener preferences for rhythm and energy.
- **Specialized Genres Provide Consistent Appeal:**
  Tailored genres like Kids demonstrate the importance of designing music for specific audiences, even if they do not dominate mainstream charts.
- **Cultural Influence Shapes Popularity**:
  Some genres achieve popularity through external factors, underscoring the importance of context and branding in the music industry.

7. Recommendations

My analysis reveals distinct strategies for mainstream and niche genres, both requiring tailored approaches to optimize performance:

7.1 Mainstream Genres: Crafting the "Sweet Spot"

To maximize the success of mainstream genres, focus on refining the characteristics of top-performing songs. Analysis of top-performing tracks (those in the 90th percentile of popularity) highlights the following sweet spots:

- **Danceability**
  The average danceability of top-performing songs across mainstream genres like Reggaeton and Chicago-House is between **0.72 and 0.78**. While high danceability generally correlates with popularity, exceeding this range could result in diminishing returns, potentially due to oversaturation or loss of rhythm balance.
- **Energy**
  Energy levels for top-performing tracks peak at around **0.70 to 0.75**, beyond which listener fatigue may set in. For instance, excessively high-energy tracks might be perceived as overwhelming outside of specific contexts (e.g., workout playlists).
- **Optimal Tempo and Duration**
  Tempo is most effective in the range of **110–130 BPM**, with durations averaging

between **3 to 4 minutes**. Tracks outside these ranges are less likely to retain engagement.

Actionable Strategies:

1. **Data-Informed Music Production:** Focus on producing tracks with danceability and energy levels within the identified sweet spots to maximize audience engagement.
2. **Playlist Integration:** Highlight tracks that meet these characteristics in prime positions on mainstream playlists such as "Top 50 Global" or "Party Hits."

7.2   Niche Genres: Building Audience Loyalty

Niche genres, such as Kids, Romance, and Iranian, require a more nuanced approach to leverage their unique appeal:

- **Emotional and Cultural Resonance:**
    - Genres like Romance thrive on emotional connection rather than broad popularity. Tracks with moderate energy (**0.30–0.50**) and slower tempos (**80–100 BPM**) resonate with audiences seeking relaxation or sentimentality.
    - Similarly, culturally specific genres like Iranian can target listeners through curated playlists that emphasize cultural pride and familiarity.
- **Tailored Playlist Strategies:**
  Develop audience-specific playlists that highlight the unique appeal of niche genres (e.g., "Kids' Dance Party," "Romantic Nights," or "Persian Vibes"). Focus on building a loyal listener base rather than competing directly with mainstream genres.
- **Cross-Media Partnerships:**
  Collaborate with children's TV shows, cultural influencers, or online communities to amplify visibility for niche genres.

Actionable Strategies:

1. **Segmented Marketing Campaigns:** Use social media and streaming platforms to run highly targeted campaigns that align with audience preferences for niche genres.
2. **Community Engagement:** Foster communities around niche genres by collaborating with creators, influencers, or events that celebrate these genres

7.3   <u>General Strategic Insights</u>

- **Monitor Trends:** Use predictive modeling to continuously analyze shifting listener preferences in both mainstream and niche categories.
- **Experimentation with Audio Features:** Test variations in danceability, energy, and tempo to identify potential new trends, particularly for emerging genres.