Unveiling Music Trends: A Spotify Dataset Analysis

This presentation delves into a comprehensive time series analysis of popular music characteristics using a Spotify track dataset. Our goal is to identify evolutionary trends in various audio features and metadata, providing insights into how popular music has changed over the years.

Dataset Overview and Initial Exploration

The analysis is based on a dataset comprising 15,828 Spotify tracks, each featuring 22 distinct attributes. These attributes range from basic metadata like track ID, name, artist, and release year to intricate audio features such as acousticness, danceability, energy, instrumentalness, loudness, speechiness, tempo, and valence.

Initial data inspection reveals a rich tapestry of musical information, with varying data types (float64 for audio features, int64 for year/popularity, and object for textual identifiers and URLs). The dataset's size (15,828 entries, 22 columns) offers a robust foundation for identifying significant patterns and shifts in music characteristics over time.

Key observations from the initial data structure and content:

- Approximately 10,819 unique track names indicate a diverse song collection.

- Presence of 'artwork_url' and 'track_url' provides direct links for multimedia access, enriching potential applications.

- A significant portion of the data relates to audio features, which are crucial for our time series analysis.

- Most columns are well-populated, though a few (like 'album_name', some audio features, and 'language') show minor non-null count discrepancies, indicating potential for minor data cleaning or imputation if necessary.

The diverse `language` distribution, with 'Tamil', 'Unknown', and 'Hindi' being dominant, suggests the dataset might have a strong regional bias, which could influence global trend interpretations.

The rich metadata and audio features will allow us to track the evolution of music styles, production techniques, and listener preferences across different eras.

Distribution of Key Audio Features

Understanding the baseline distribution of critical audio features provides context for identifying trends in popular music. Here, we visualize the distributions of song popularity, duration, key, tempo, acousticness, loudness, and mode across the entire dataset.

Popularity Scores

The distribution of popularity scores indicates a long tail, with most songs having lower popularity and a few tracks achieving very high scores. This highlights the selective nature of "popular" music.

Song Duration

Most songs cluster around a typical duration, but there's a broad range, reflecting various song structures and genre conventions. The majority of tracks are between 3-4 minutes.

Musical Key

Certain musical keys (e.g., 0.0 and 7.0, corresponding to C and G major in a standard MIDI mapping) are more prevalent, suggesting preferences or common practices in music composition. Minor keys (e.g., 2.0, D minor) are also common.

Tempo (BPM)

Tempo distribution shows peaks around common dance tempos (100-140 BPM), indicating the energy and rhythm often associated with popular tracks.

Acousticness

The distribution of acousticness is skewed towards lower values, implying that most tracks in this dataset are not primarily acoustic, which is typical for modern popular music production.

Loudness (dB)

Loudness levels are generally concentrated in the mid-to-high range, reflecting the mastering practices in contemporary music, often associated with the "loudness war" phenomenon. Average loudness is -33.35 dB.

These foundational distributions set the stage for our temporal analysis, allowing us to pinpoint deviations and trends from these general patterns in popular songs over time.

Identifying "Popular" Tracks for Focused Analysis

To specifically analyze trends in popular music, we define "popular" as tracks falling within the top 75th percentile of the `popularity` score. This filtering ensures our analysis focuses on songs that have resonated most with listeners, providing a clearer signal for evolving trends.

By isolating these highly popular tracks, we mitigate the noise introduced by less listened-to songs and can observe more pronounced shifts in musical characteristics. This approach is crucial for understanding the dynamic landscape of mainstream preferences.

- **Filtering Criterion:** Songs with a `popularity` score at or above the 75th percentile of the entire dataset.

- **Dataset Size:** This filtering results in 4,053 tracks, representing a focused subset of the original data.

- **Impact:** Future analyses will leverage this `top_popularity_df` to explore how features like danceability, energy, acousticness, and instrumentalness have evolved specifically within the realm of highly successful music.

This segmentation is a standard practice in music data science to concentrate on the most impactful entries, enabling a more targeted and meaningful time series analysis.
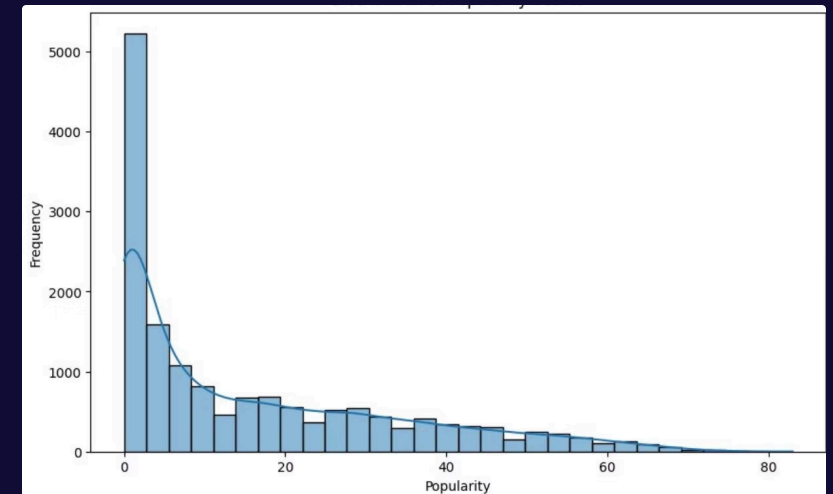
The histogram illustrates the right-skewed distribution of popularity, validating our approach to segment the top quartile for focused analysis.

# Clustering Top Popularity Tracks by Key Audio Features

To uncover latent sub-genres or distinct sound profiles within highly popular music, we performed K-Means clustering on the top popularity tracks, using 'acousticness', 'instrumentalness', and 'speechiness' as features. This helps categorize popular songs into groups based on their inherent sonic characteristics.

We utilized `StandardScaler` to normalize the feature data, preventing any single feature from disproportionately influencing the clustering process. For demonstration, we chose 3 clusters, recognizing that optimal cluster numbers can be further refined through methods like the elbow method.

Key findings from the clustering:

- **Cluster 0 (Pop/Vocal-Centric):** Relatively low acousticness, low instrumentalness, and moderate speechiness. These tracks likely represent mainstream vocal-driven pop or similar genres.

- **Cluster 1 (Acoustic/Warm):** Higher acousticness, very low instrumentalness, and low speechiness. This cluster may contain more organic, unplugged, or ballad-like popular songs.

- **Cluster 2 (Instrumental/Cinematic):** Moderate acousticness, very high instrumentalness, and moderate speechiness. This suggests popular tracks that lean heavily on instrumentation, possibly including film scores, electronic instrumentals, or tracks with significant instrumental breaks.

These clusters demonstrate that even within "popular" music, there's significant diversity in sonic profiles, indicating that different styles can achieve widespread appeal.
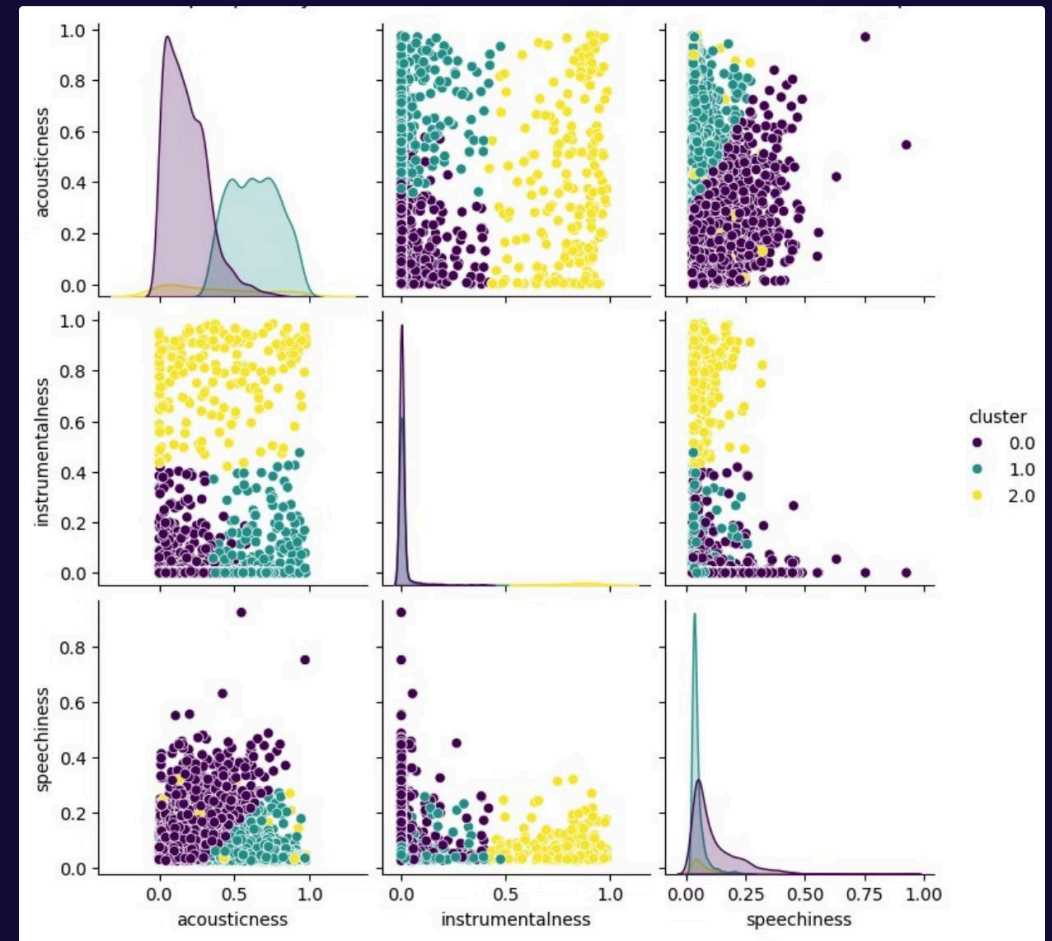
The pairplot visually distinguishes the clusters, showing how they separate along the axes of acousticness, instrumentalness, and speechiness.

Mean feature values per cluster: acousticness instrumentalness speechiness

cluster

0.0 0.182683 0.011465 0.114960

1.0 0.630297 0.012353 0.047811

2.0 0.365318 0.787623 0.072611

Popularity Distribution Across Musical Attributes

Beyond general trends, understanding how specific musical attributes correlate with popularity is key. We examine the popularity distribution for top tracks across different keys, modes (major/minor), and time signatures. This helps identify if certain structural elements are more common in highly popular music.

Musical Key

While popular songs are found across all keys, some keys (e.g., C major (0), G major (7)) tend to have a higher median popularity, suggesting a subtle preference or ease of listenability. There's significant overlap, implying no single "most popular" key.

Popularity Distribution by Key for Top Popularity Tracks
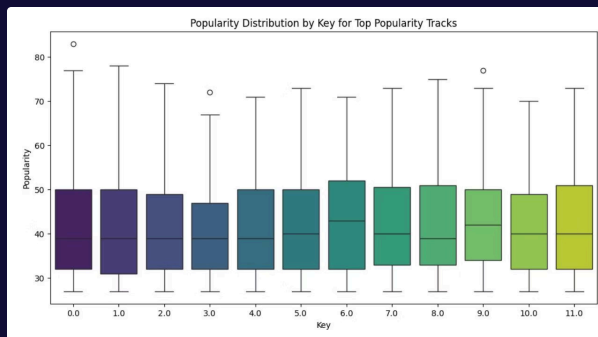
Mode (Major/Minor)

Major keys (mode 1.0) generally show a slightly higher median popularity compared to minor keys (mode 0.0), aligning with the perception that major keys often convey more upbeat or positive emotions, which can contribute to broader appeal.
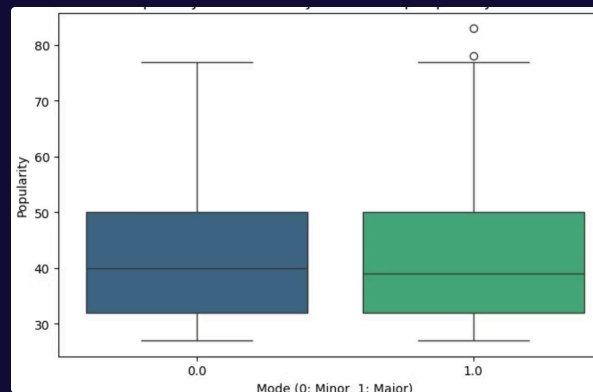
Time Signature

The vast majority of popular songs are in a 4/4 time signature (4.0), which is the most common in Western popular music. Other time signatures (e.g., 3.0 for waltz, 5.0, 1.0) appear less frequently and show wider popularity ranges, indicating niche appeal.
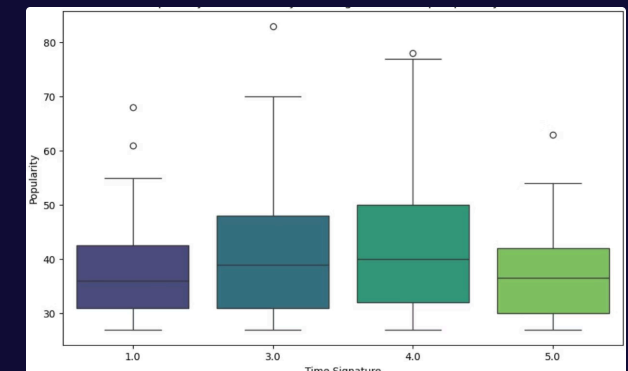
These distributions confirm that while popular music embraces diversity, certain foundational elements like major modes and 4/4 time signatures remain dominant, providing a rhythmic and harmonic comfort zone for listeners.

Made with GAMMA

Cross-Lingual Popularity Trends

Language plays a significant role in music consumption and popularity. This analysis explores how average danceability, energy, and valence for popular songs vary across different language categories, revealing potential cultural preferences embedded in musical attributes.

We group popular songs by their detected language and calculate the mean for 'danceability', 'energy', and 'valence'. This provides insights into the emotional and physical characteristics of popular music in different linguistic contexts.

- **English Tracks:** Exhibit high average danceability (0.66), energy (0.86), and valence (0.70), suggesting a leaning towards energetic and positive sounds, typical of global pop.

- **Tamil Tracks:** Show moderate danceability (0.56), energy (0.65), and valence (0.48), potentially indicating a blend of energetic and more contemplative styles within popular Tamil music.

- **Hindi Tracks:** Display high danceability (0.65), moderate energy (0.61), and moderate valence (0.55), reflecting a vibrant, dance-oriented, but perhaps less overtly energetic popular sound than English tracks.

- **Other Languages (Telugu, Malayalam, Korean):** Provide unique profiles, though with fewer samples, requiring cautious interpretation. Malayalam and Korean popular tracks show particularly high energy and danceability.

These variations highlight distinct musical landscapes associated with different languages, influencing global music trends and local popular appeal.

Average Danceability of Popular Songs by Language
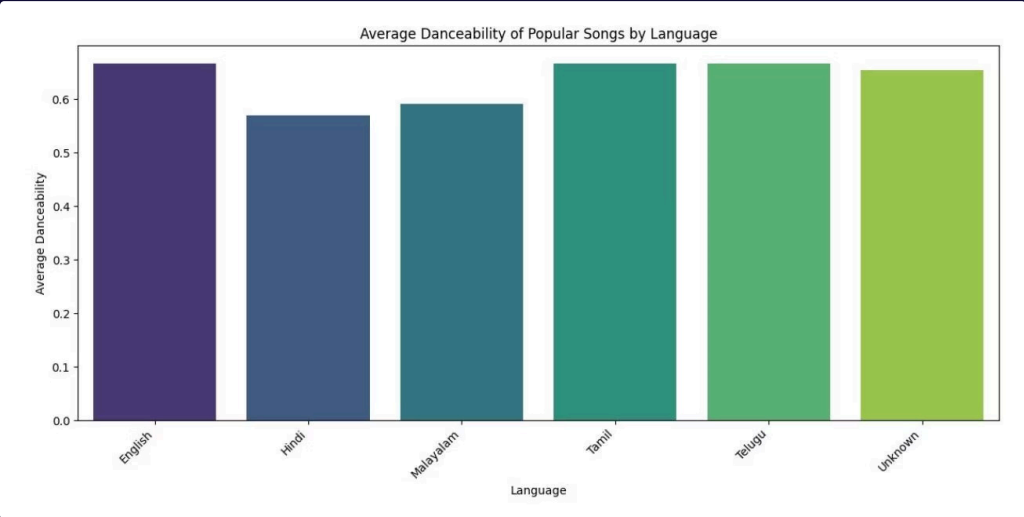
The bar chart clearly shows the varying average danceability across languages, with English, Korean, and Hindi tracks leading in this metric.

| English	0.66	0.86	0.70
Hindi	0.65	0.61	0.55
Korean	0.63	0.78	0.62
Malayalam	0.65	0.82	0.62
Tamil	0.56	0.65	0.48
Telugu	0.60	0.73	0.54

Evolution of Song Characteristics Over Decades

Music production and consumption habits evolve significantly over time. We analyze how song duration and liveness (a measure of recorded performance vs. studio production) have changed for popular tracks across different decades.

Average Duration of Popular Songs

The average duration of popular songs has shown fluctuations, but a general trend might indicate a slight decrease or stabilization around a preferred length. Shorter attention spans or optimized streaming formats could influence this.

Distribution of Duration (ms) for Top Popularity Tracks Across Decades
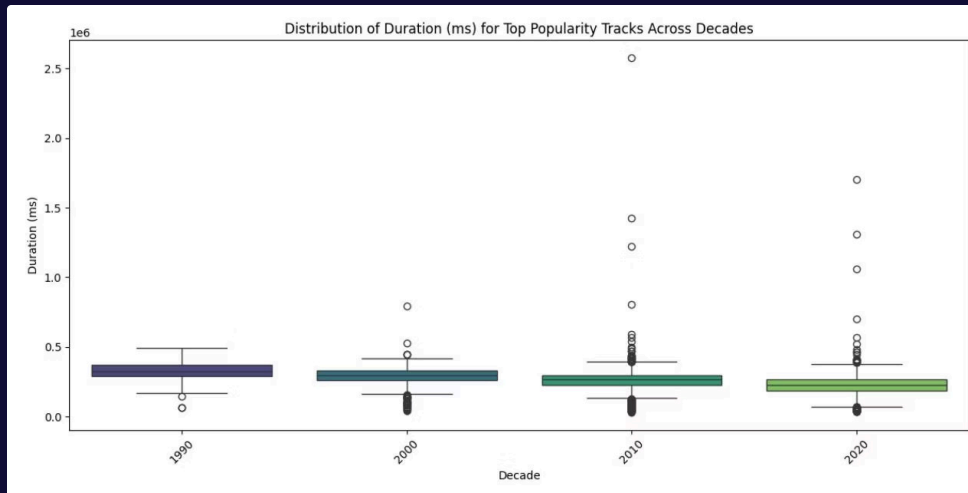
Average Liveness of Popular Songs

Liveness scores, which indicate the presence of a live audience or performance, also change. A lower liveness might suggest more studio-produced tracks, while higher scores could point to a resurgence of live-recorded elements in popular music.

Distribution of Liveness for Top Popularity Tracks Across Decades
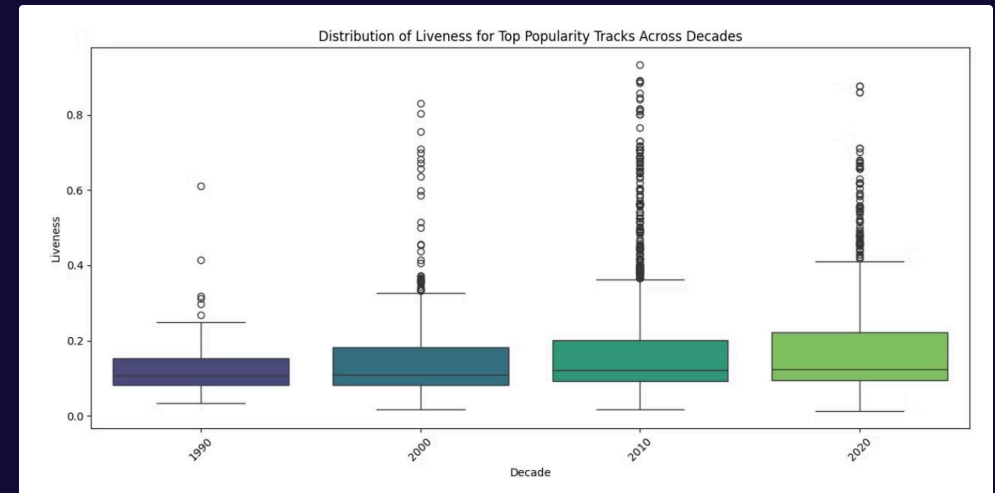
These temporal shifts provide valuable clues about evolving production techniques, listener preferences, and the commercial pressures shaping popular music. For instance, the rise of electronic music might correlate with lower acousticness and higher energy, while a preference for raw, authentic sounds could boost liveness.

Time Series Analysis of Key Features

This section presents a detailed time series analysis of various audio features for popular songs. By tracking changes year-over-year, we can uncover significant trends in musical characteristics.

01

Average Song Popularity

The average popularity of songs fluctuates yearly, potentially reflecting external factors like new artists, genre shifts, or platform algorithms. Visualizing this trend helps understand the general vitality of the music industry.
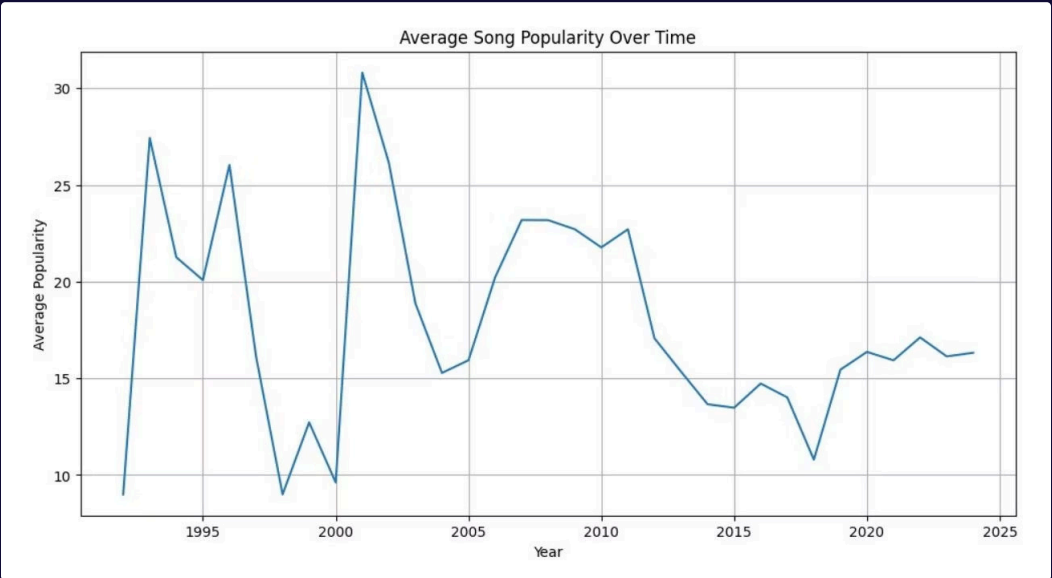

02

Optimal Danceability and Energy

Analyzing average danceability and energy levels for popular songs reveals shifts in what makes music engaging. A consistent rise might indicate a global preference for more upbeat, movement-inducing tracks.
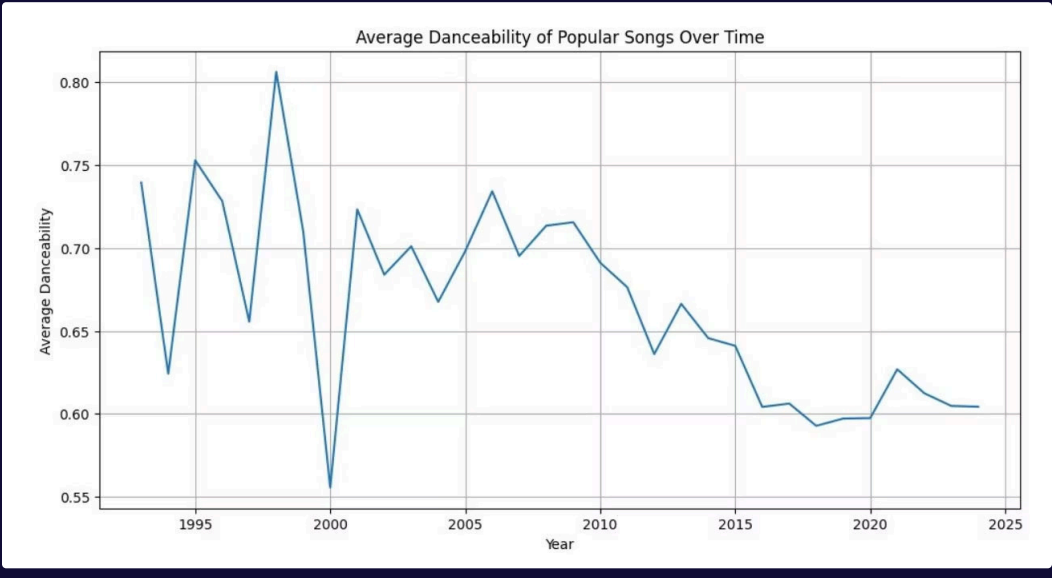
03

Acousticness and Instrumentalness

Trends in acousticness and instrumentalness hint at the balance between natural sounds and synthesized elements. A decline in acousticness could point towards increased electronic production, while instrumentalness might reflect genre evolution.
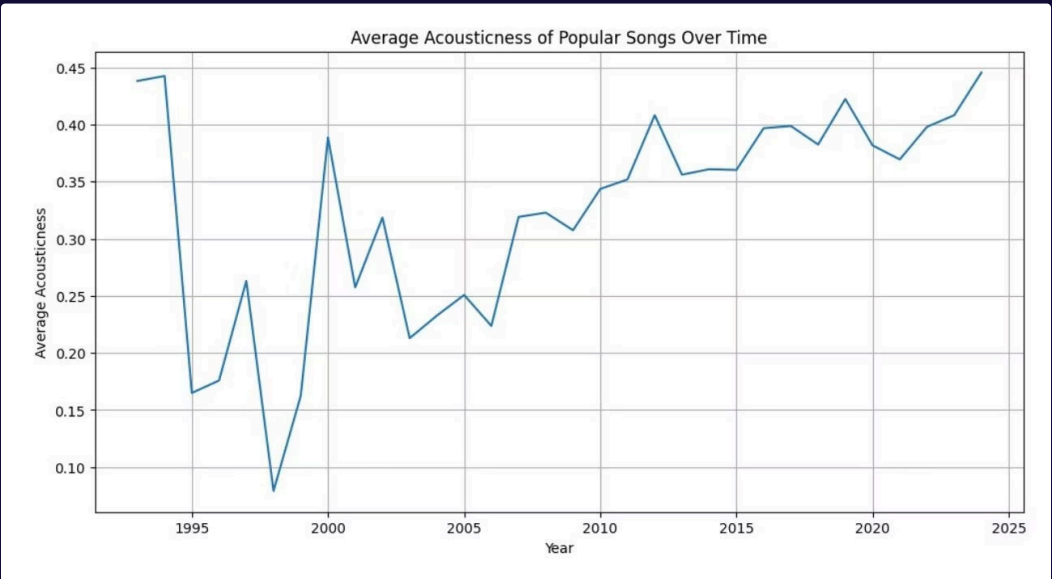
04

Valence and Loudness

Valence (musical positivity) tracks emotional tone, while loudness (the "loudness war") signifies mastering trends. A rising average valence could suggest happier music, while sustained high loudness would confirm ongoing industry practices.
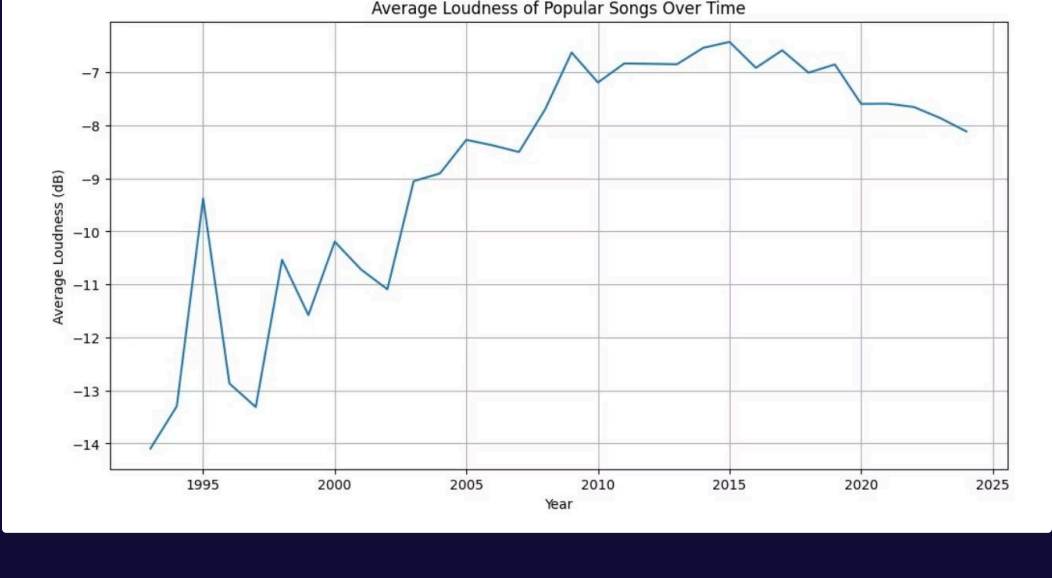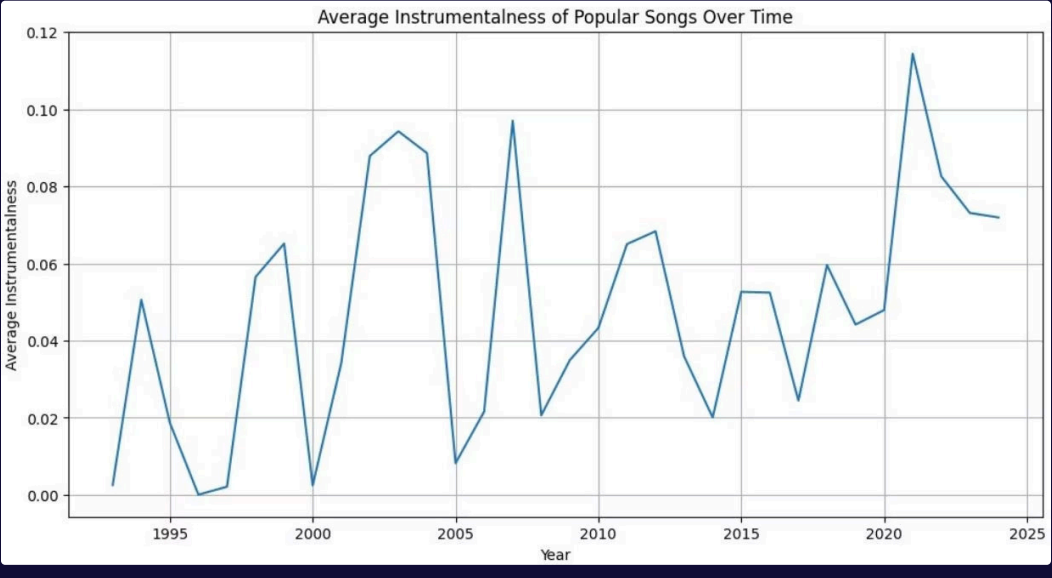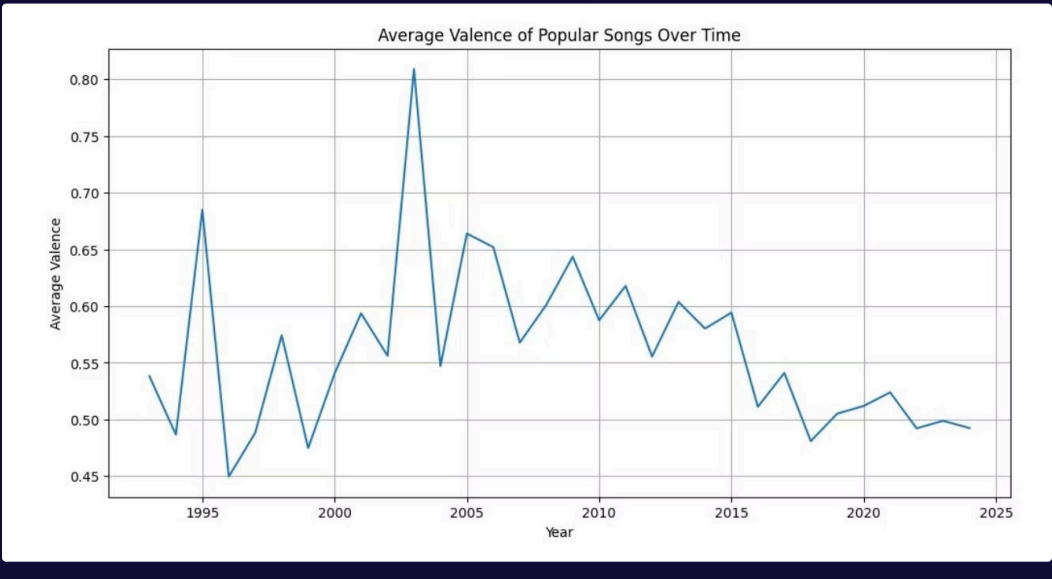
Emerging Trends and Future Directions

Our comprehensive time series analysis reveals dynamic shifts in popular music characteristics, offering critical insights for data analysts and music enthusiasts alike. Understanding these evolving trends can inform music production, marketing, and predictive modeling.
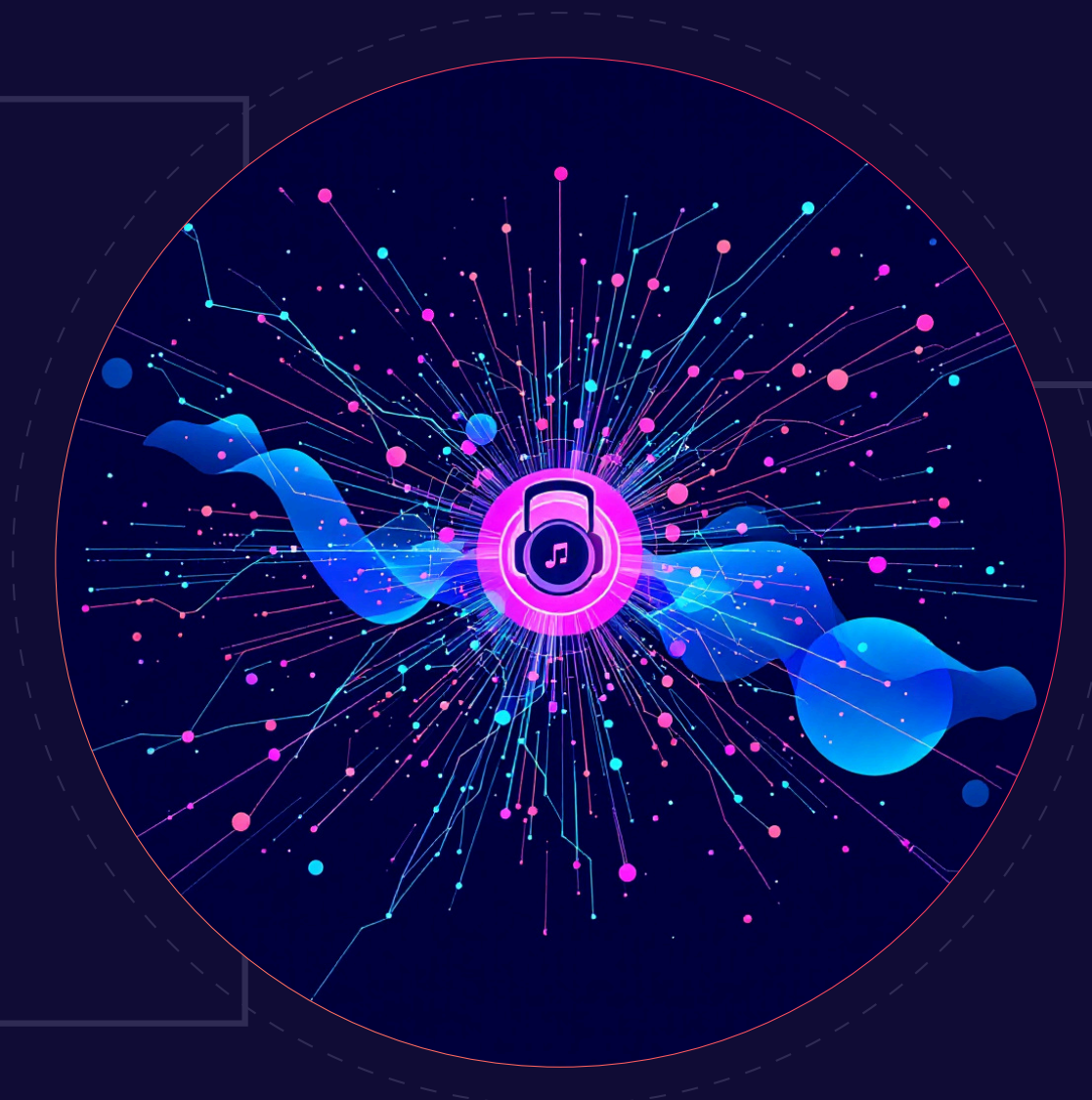
Genre Evolution
How styles change over time?

Predictive Modeling
Which features forecast hits?

Cross-Cultural Impact
How music travels globally?

Key takeaways from the temporal analysis include:

- **Danceability & Energy:** Consistent or increasing trends in these features suggest a persistent preference for engaging, high-energy tracks.
- **Acousticness & Instrumentalness:** The interplay between these features can signal shifts between organic and synthetic soundscapes, potentially marking the rise or decline of electronic music dominance.
- **Loudness & Valence:** Observing the "loudness war" and the emotional tone (valence) provides context on production practices and the overall mood of popular music.
- **Language & Key/Tempo:** Shifts in linguistic prevalence and musical keys/tempos highlight cultural impacts and evolving compositional preferences.

> "The data doesn't just tell us what music was popular; it tells us how the very fabric of sound design and emotional resonance has changed with each passing year."

Future research could delve deeper into causal factors behind these trends, such as technological advancements in music production, socio-cultural influences, or the impact of streaming platform algorithms on what becomes popular. Further analysis could also incorporate external data, such as economic indicators or major cultural events, to provide a more holistic understanding of music's evolution.

Made with **GAMMA**