Spotify Dataset Analysis

AN EXPLORATORY DATA ANALYSIS (EDA) BASED STORY TO FIND SOME RELIABLE INSIGHTS

Presented by: PRANAB SARDAR

Data Description:

	Column Name	Description
0	track_id	A unique identifier for the track on Spotify.
1	track_name	The title of the song.
2	artist_name	The name of the artist(s) who performed the song.
3	year	The release year of the song.
4	popularity	A measure of how popular a track is, ranging f...
5	artwork_url	A URL pointing to the album artwork for the tr...
6	album_name	The name of the album the track belongs to.
7	acousticness	A confidence measure indicating whether the tr...
8	danceability	A measure of how suitable a track is for danci...
9	duration_ms	The duration of the track in milliseconds.
10	energy	A perceptual measure of intensity and activity...
11	instrumentalness	Predicts whether a track contains no vocal con...
12	key	The key the track is in, represented as an int...
13	liveness	Detects the presence of an audience in the rec...
14	loudness	The overall loudness of a track in decibels (dB).
15	mode	Indicates the modality (major or minor) of a t...
16	speechiness	A measure detecting the presence of spoken wor...
17	tempo	The overall estimated tempo of a track in beat...
18	time_signature	An estimated overall time signature of a track
19	valence	A measure from -1.0 to 1.0 describing the musi...
20	track_url	A URL to the Spotify track.
21	language	The detected language of the song's lyrics.

- **Acousticness:** Measure (0-1)of how acoustic the track is (higher=more acoustic)
- **Danceability:** Measure (0-1) describing how suitable a track is for dancing(based on tempo, rhythm and beat strength).
- **Popularity:** Popularity score of the track (0-100),based on Spotify user activity and plays.
- **Tempo:** Estimated tempo of track in beats per minute(BPM)
- **Energy:** Measure (0-1) of the track's intensity and (high energy = loud, fast, dynamic).
- **Duration:** The duration of the track (in milliseconds)

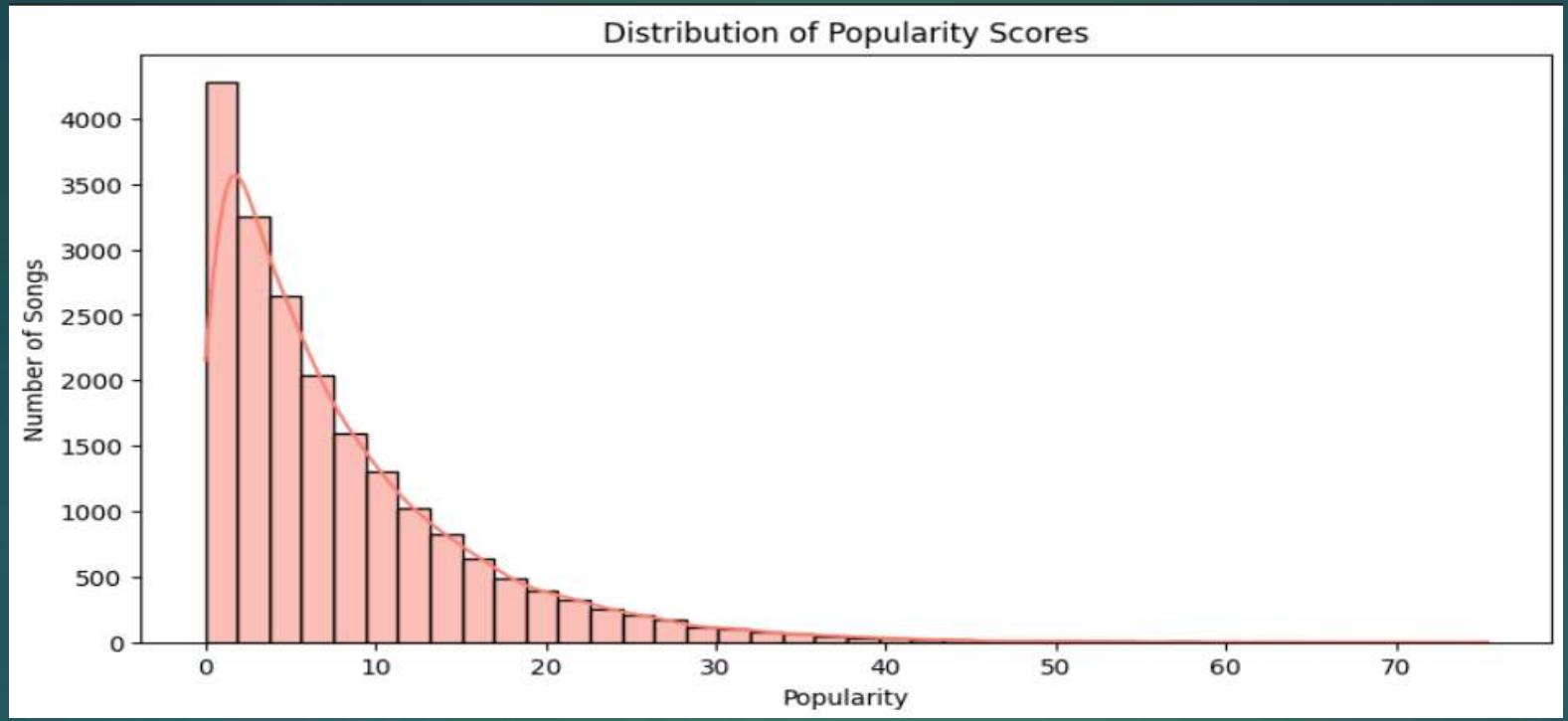
Univariate Analysis

1. Popularity: Popularity score of the track (0-100),based on Spotify user activity and plays.

2. Danceability: Measure (0-1) describing how suitable a track is for dancing(based on tempo, rhythm and beat strength).

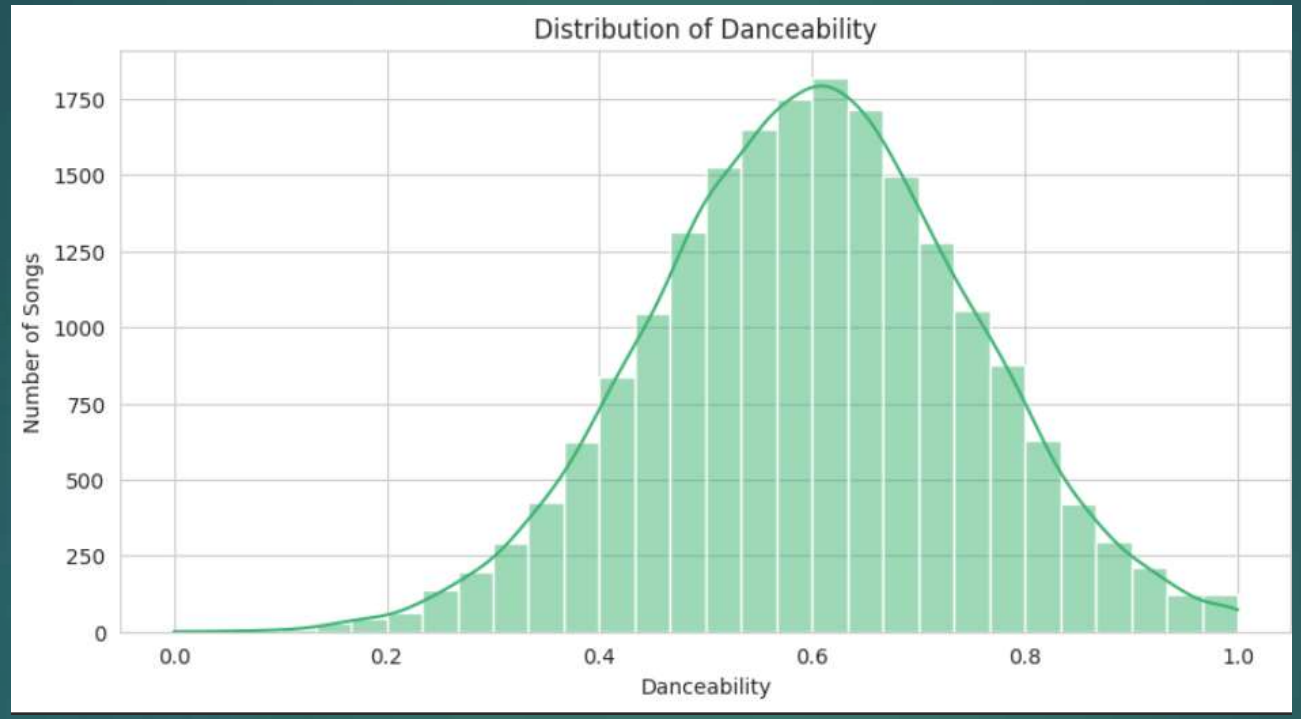
3. Tempo: Estimated tempo of track in beats per minute(BPM)

Distribution of Popularity Scores



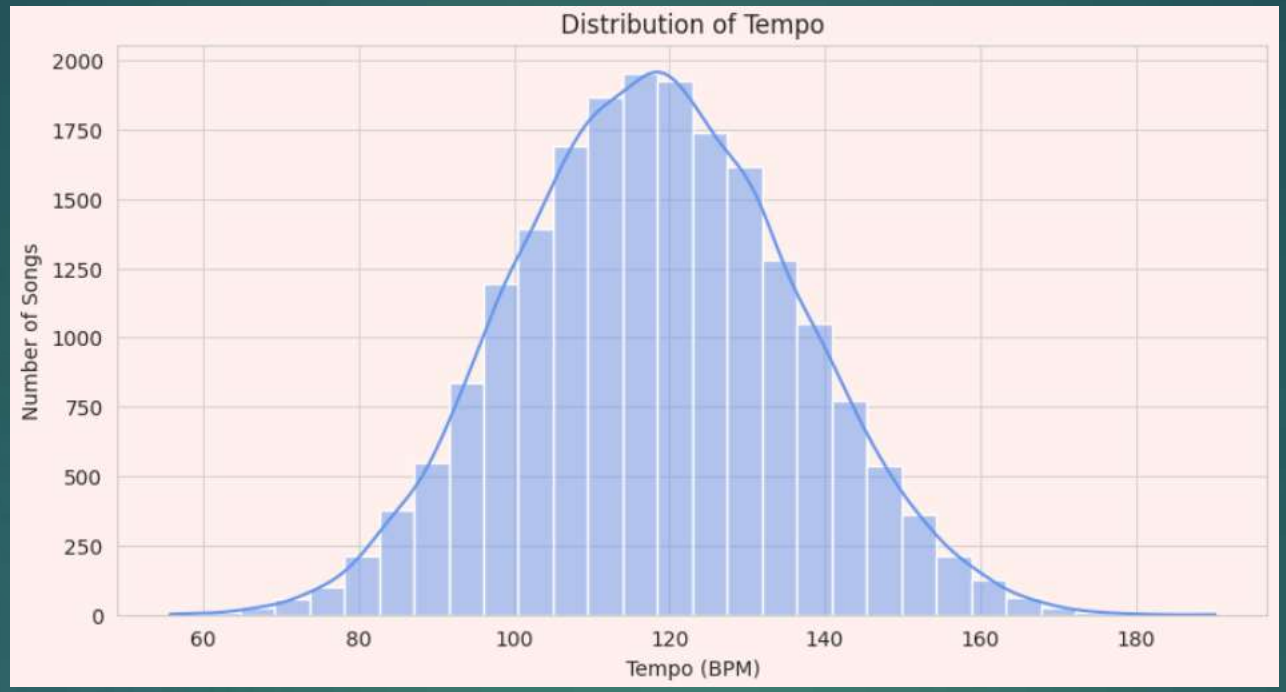
The data is heavily skewed to the left, indicating that a large number of songs have a low popularity score.

Distribution of Danceability



The plot shows that the majority of songs have a danceability score between 0.50 and 0.75. Aiming for balanced danceability (0.6 - 0.8) might align with listener preferences

Distribution of Tempo



The distribution indicates that the majority of the songs in this dataset have a tempo between approximately 90 and 140 BPM. For commercial success, especially in pop or dance genres, targeting tempos in 100 – 130 BPM which aligns with natural human rhythm and dancefloor-friendly.

Insights and Implications:

- Most songs have low to moderate popularity (0-50). Very few tracks are highly popular.

- the majority of songs have a danceability score between 0.50 and 0.75. Aiming for balanced danceability (0.6 - 0.8) might align with listener preferences

- the majority of the songs in this dataset have a tempo between approximately 90 and 140 BPM. For commercial success, especially in pop or dance genres, targeting tempos in 100 – 130 BPM which aligns with natural human rhythm and dancefloor-friendly.

Bivariate Analysis

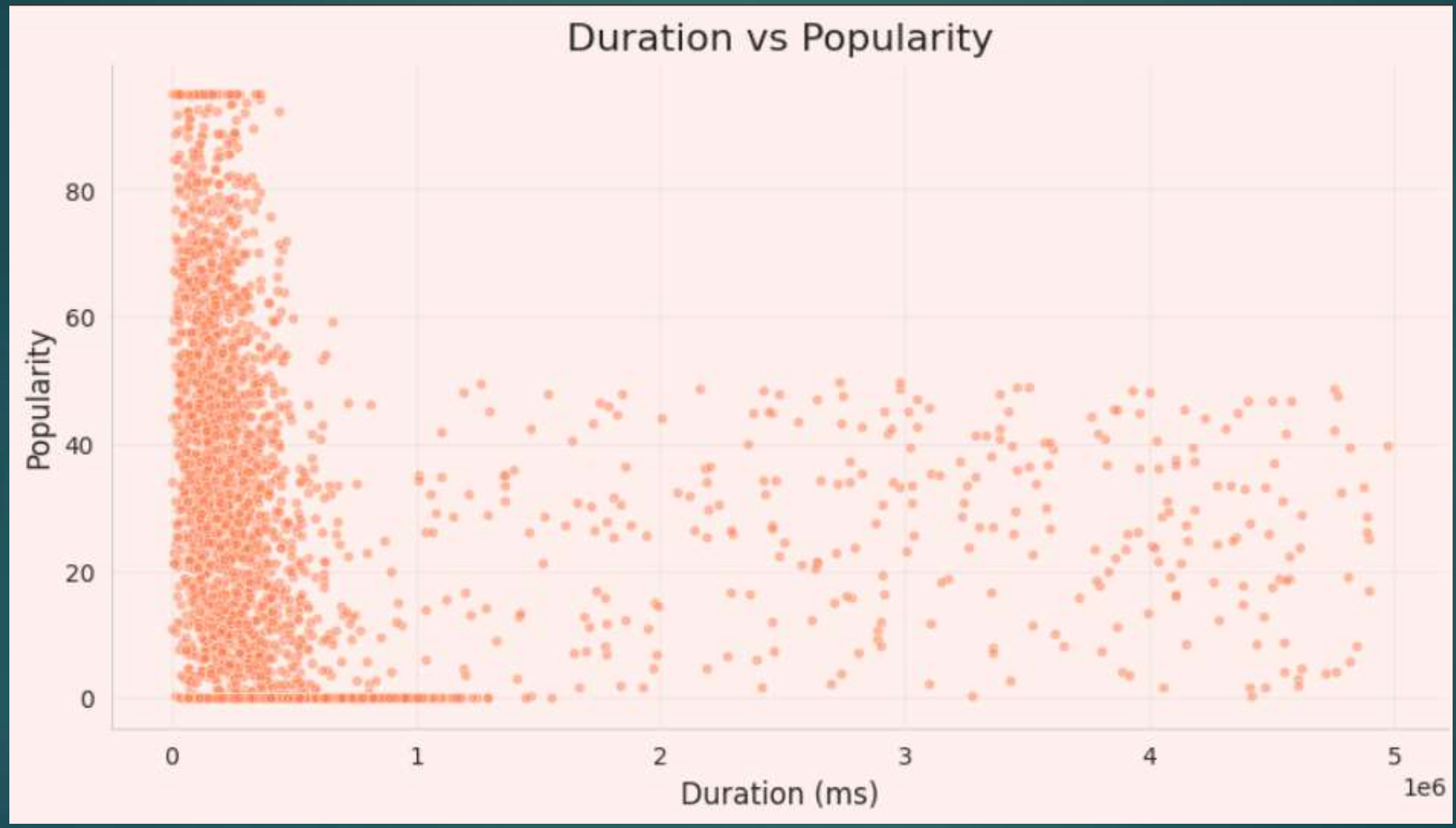
1. Duration vs Popularity: to explore if there is a correlation or relationship between how long a song is and how popular it becomes.

2. Popularity-Danceability : generally shows a weak to moderate positive correlation

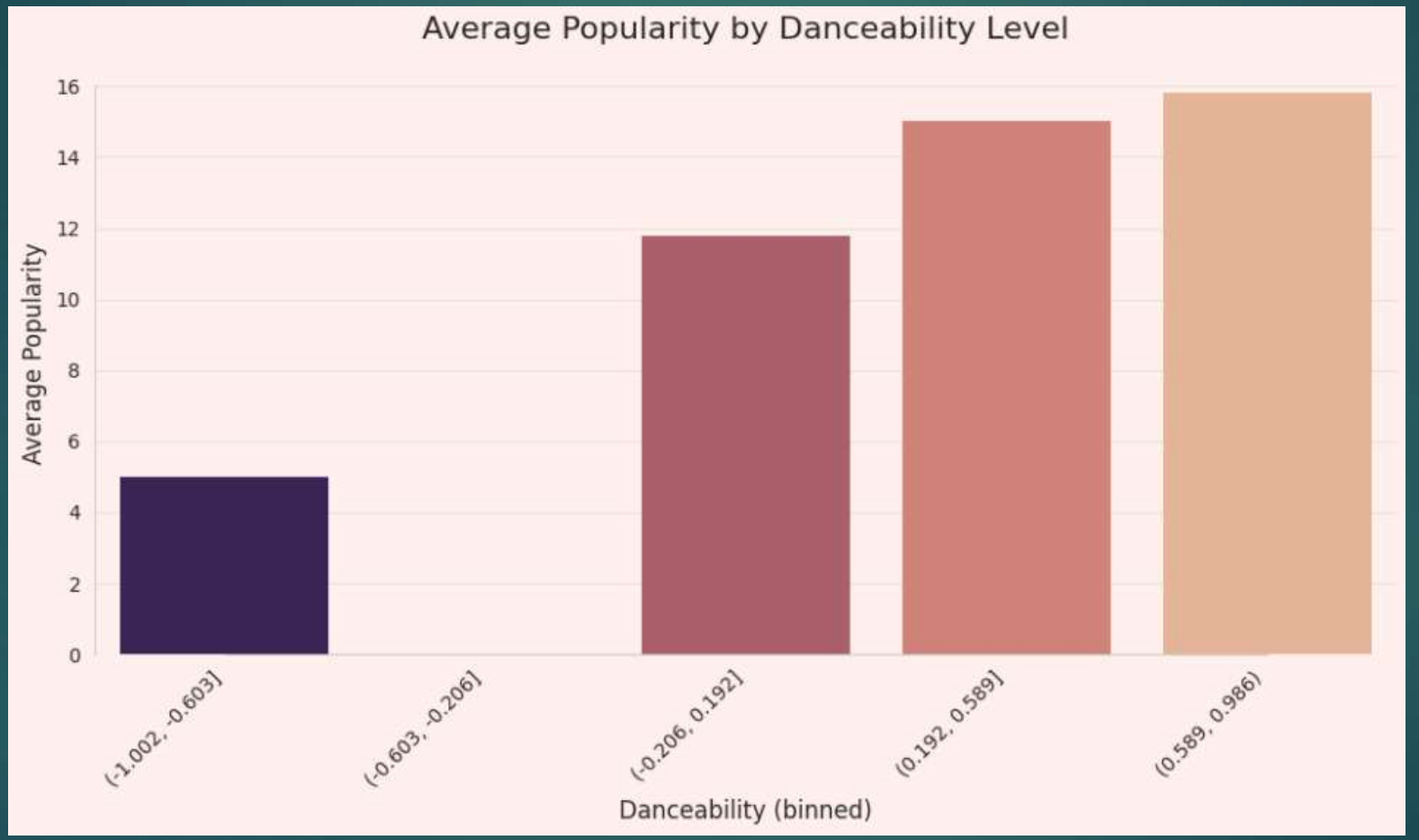
3. Popularity-Acousticness : Determine the most critical strengths, weaknesses, opportunities, and threats to address.

4. Popularity-Energy: to understand the specific empirical relationship, direction, and strength of the association between song "popularity" and "energy"

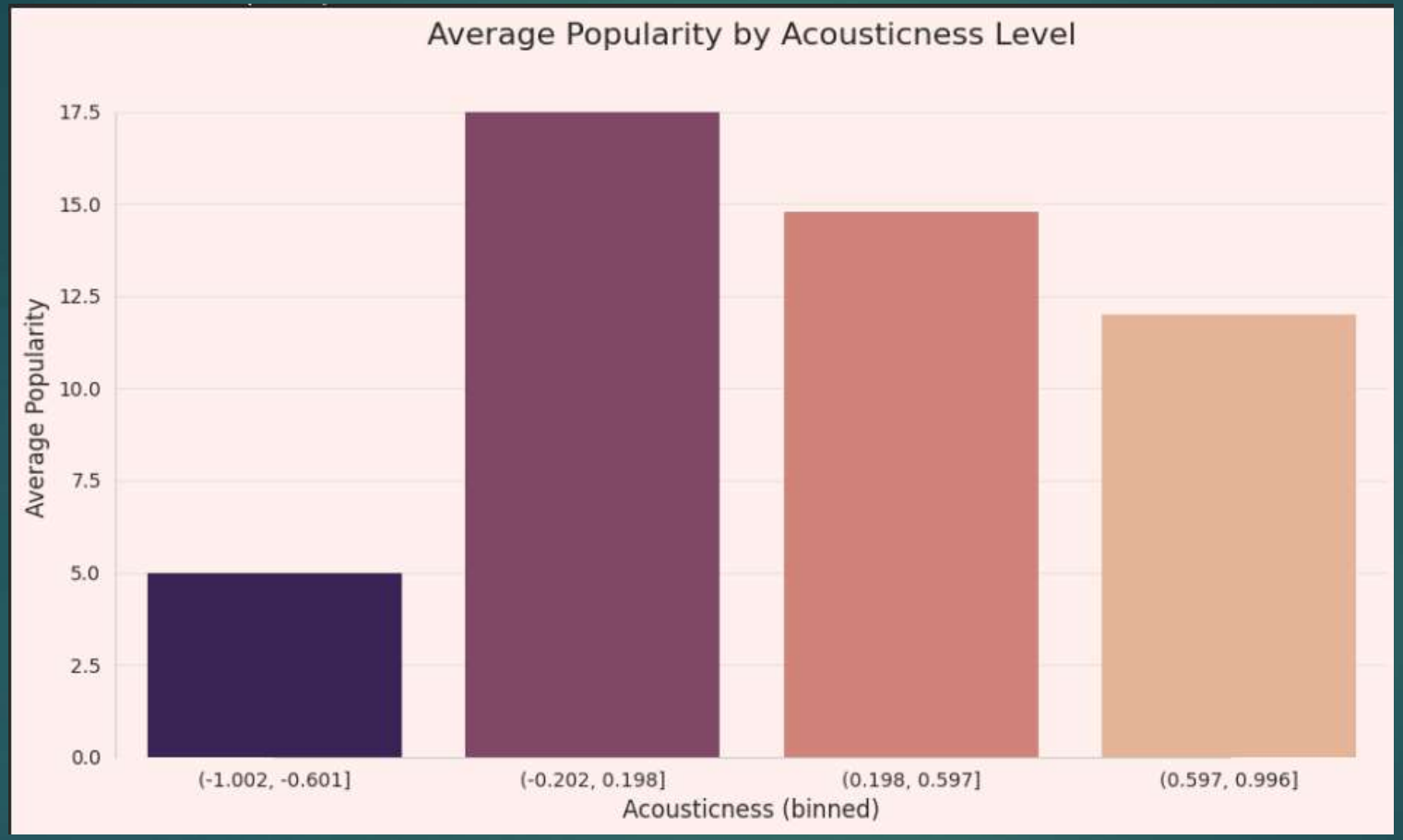
Duration vs Popularity



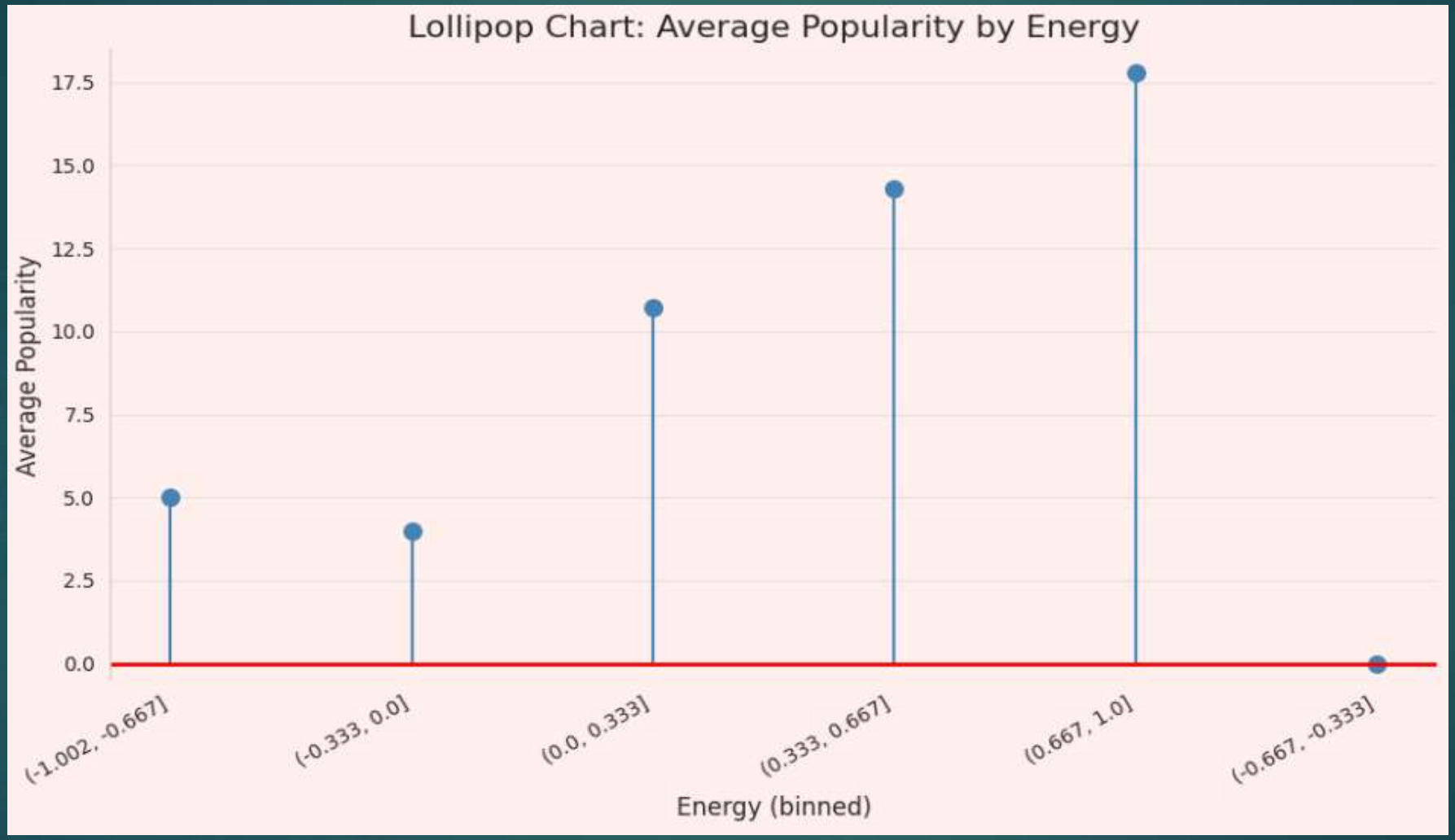
Popularity vs Danceability



Popularity vs Acousticness



Popularity vs Energy



***Insights and Implications*:-**

- Most popular songs tend to have shorter durations — the highest concentration of popularity is seen for songs under 1 million milliseconds (around 16 minutes). Longer tracks show much lower popularity. Implication: Listeners generally prefer concise songs; artists and producers should aim for shorter, engaging tracks to maximize reach and streaming success.

- Songs with higher danceability scores have noticeably higher average popularity. Implication: Danceable tracks are more appealing to listeners, possibly due to their suitability for parties, playlists, and social media trends. Artists can focus on rhythm and groove elements to boost engagement.

- Moderate levels of acousticness correspond to higher popularity, while very high or very low acoustic songs are less popular. Implication: Listeners tend to prefer a blend of acoustic and electronic elements — too much acoustic texture may limit mainstream appeal, while a balanced production style could enhance popularity.

- Popularity increases with energy — energetic tracks tend to score higher on popularity metrics. Implication: Energetic, upbeat songs perform better, likely due to their association with mood-lifting and active listening contexts. Artists can focus on lively tempos and powerful beats to attract more streams.

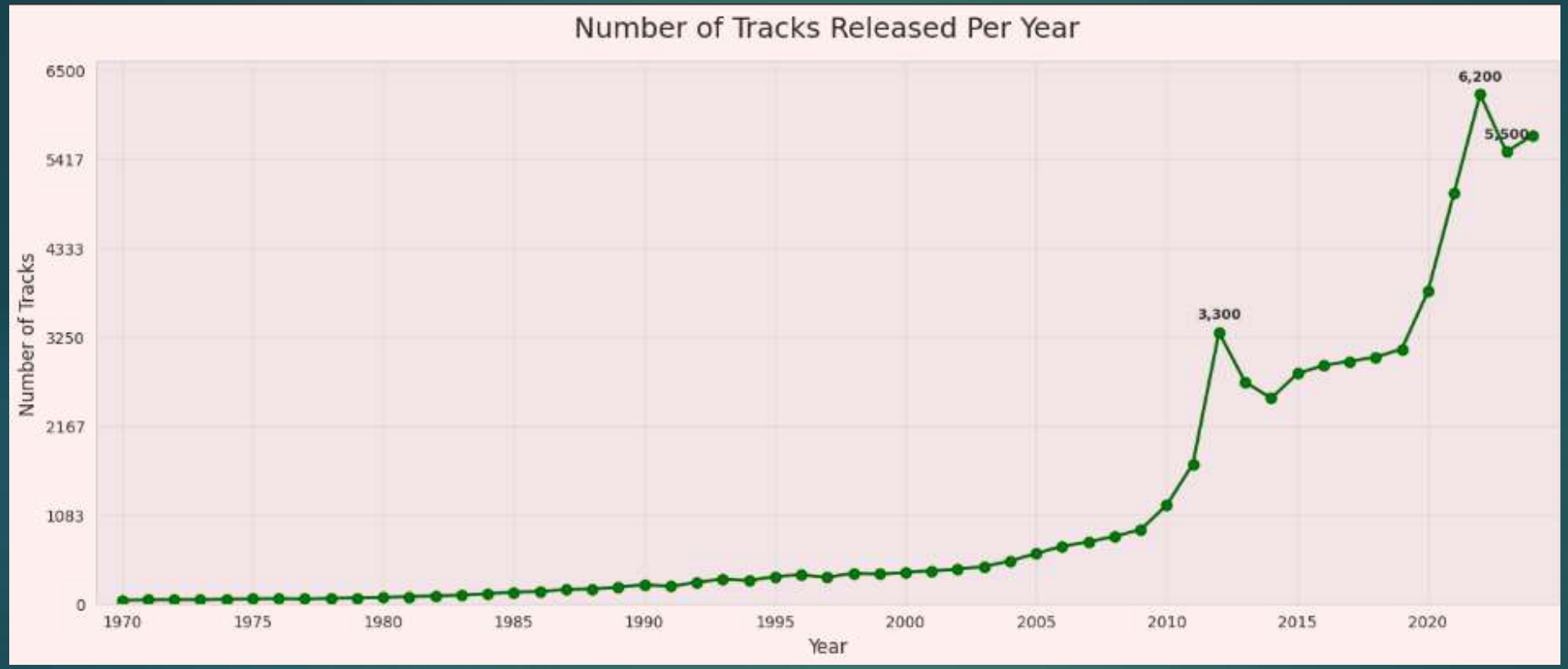
Time Series Analysis

1. Number of Tracks Released per year: To check an increase or decrease in no. of songs released

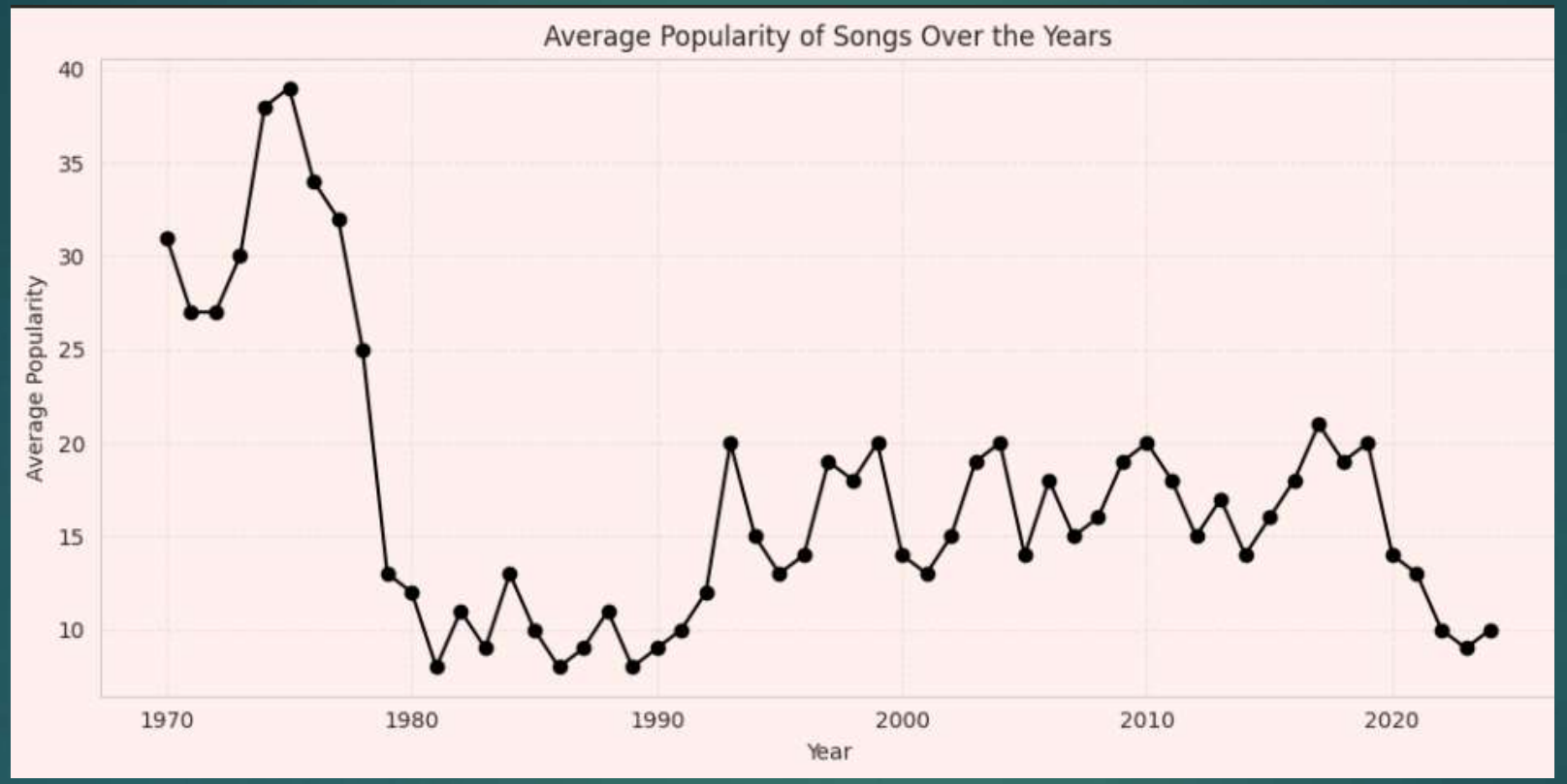
2.Average Popularity of Songs over the Years: To check whether the songs remain popular with time

3. Danceability-Energy Trend: To check the change in the trend over time

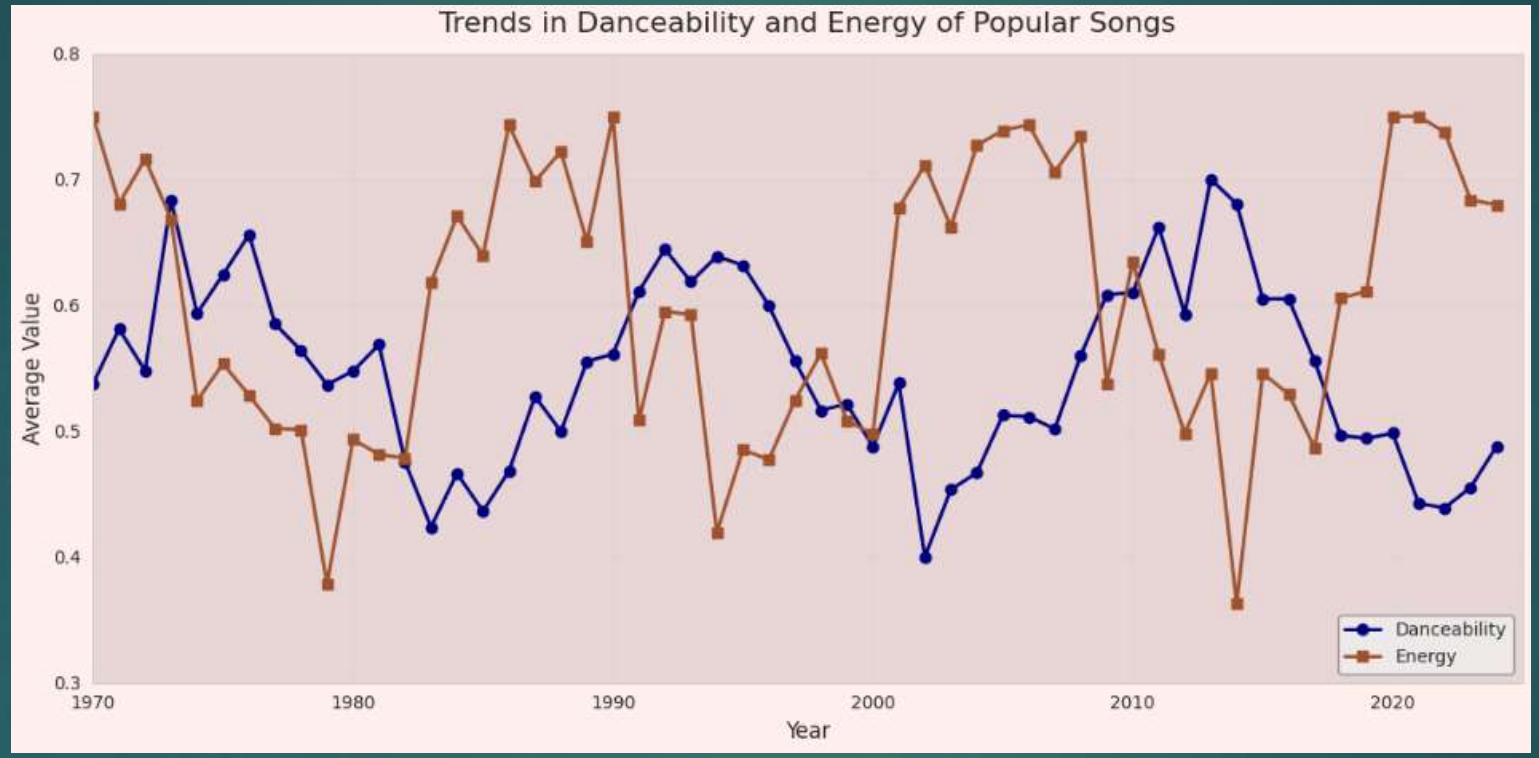
Track Released per Year



Average Popularity per Year



Danceability-Energy Trend



***Insights and Implications*:-**

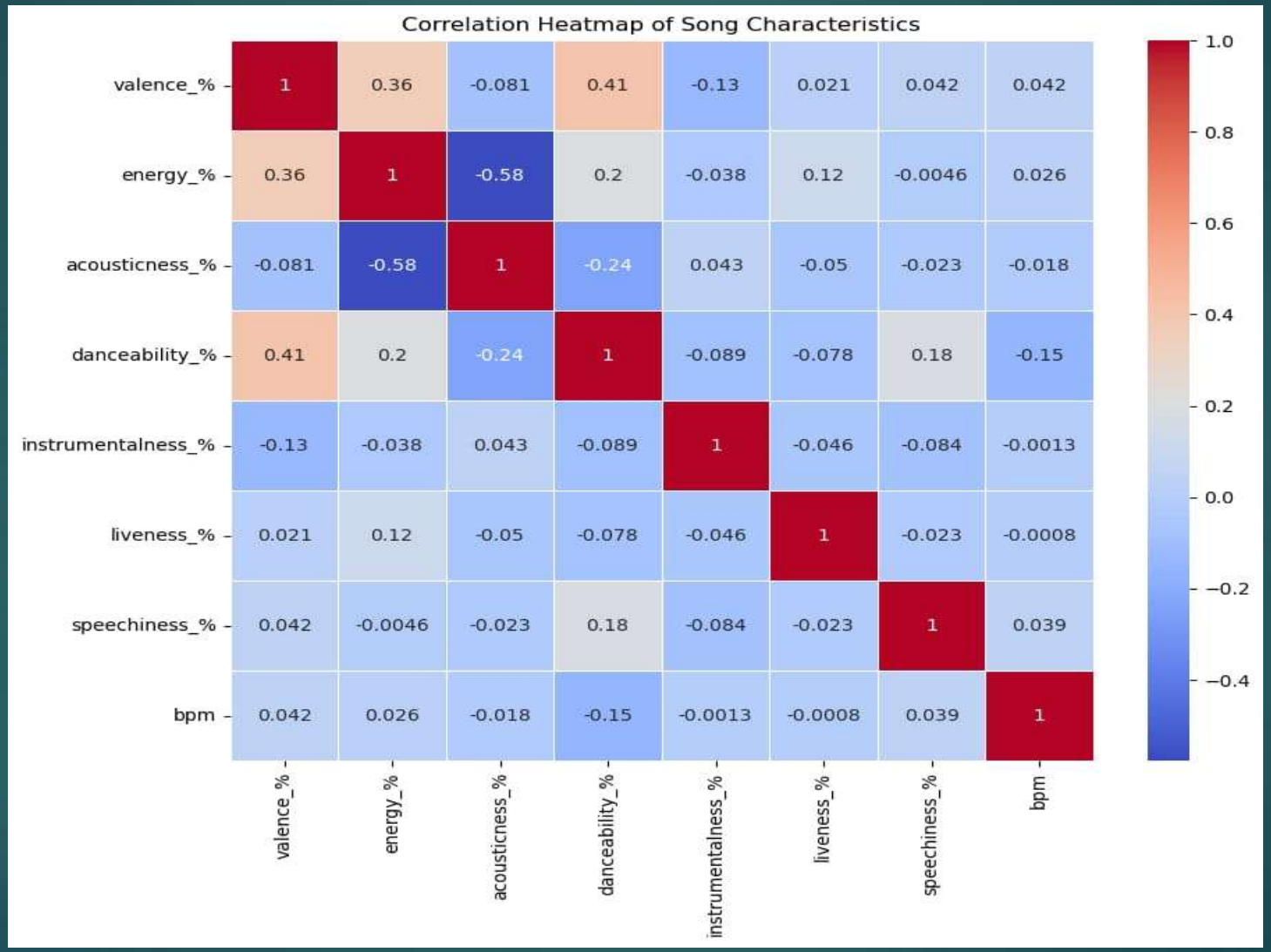
- Song Releases Trend: The number of tracks released increased steadily from 1970 to 2020, peaking around 2020 before slightly declining. Implication: The 2020 peak indicates a highly competitive period for artists; however, the slight decline afterward suggests that releasing music now might face less competition, giving quality songs a better chance to stand out.

- Popularity Over the Years: Average popularity was highest during the 1970s and then dropped sharply in the 1980s–1990s, remaining relatively low and stable afterward. Implication: With more songs being released in recent years, competition is tougher, and gaining high popularity requires strong marketing and originality.

- Danceability and Energy Trends: Danceability and energy have stayed relatively consistent over time, with energy levels generally being slightly higher. Implication: Listeners' preference for energetic and danceable tracks has remained steady, so producers should maintain this balance to align with audience expectations.

- Overall Market Observation:Despite fluctuations in song releases and popularity, the core musical characteristics preferred by audiences have not drastically changed. Implication: Success now depends less on changing trends and more on creativity, production quality, and audience engagement.

Correlation Heatmap



Insights and Implications

- Loud, energetic, and positive songs tend to go together. Implication: Boost loudness and energy for high-impact, feel-good tracks.

- Popular songs cluster around moderate-to-high energy and danceability. Implication: Balance groove and energy; tempo matters less than feel.

Thank You