SPOTIFY DATASET ANALYSIS

AGENDA

OVERVIEW

DATASET ANALYSIS

INSIGHTS AND OBSERVATIONS

CONCLUSION

OVERVIEW

OBJECTIVE:

To explore the Spotify Tracks dataset to uncover meaningful patterns and insights related to song characteristics, popularity trends, and artist performance

Since the goal is to understand what makes a track popular and how these features vary across **languages, genres, and time**.

WE NEED TO STUDY THE FOLLOWING FEATURES

- **DANCEABILITY** - How suitable a track is for dancing based on rhythm stability, beat strength
- **ENERGY** - Represents the intensity and activity level of a song
- **VALENCE** - Describes the musical positivity of a track - high valence sounds happy or cheerful, while low valence feels sad or moody
- **TEMPO** - Indicates the speed or pace of a song, measured in beats per minute (BPM)
- **ACOUSTICNESS** - Estimates how non-electronic a track is - higher values mean fewer synthesized sounds.

DATASET ANALYSIS

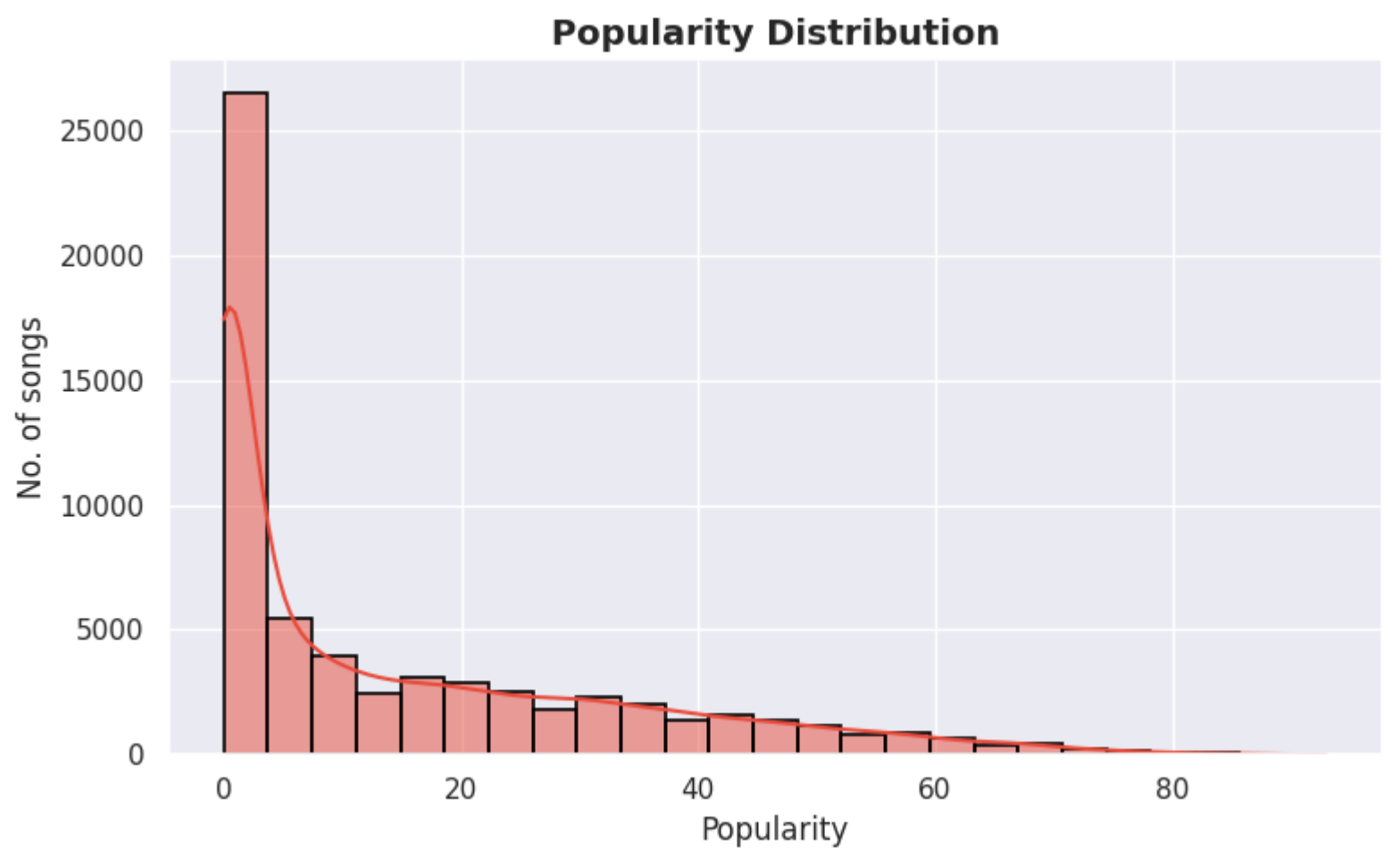
Each record represents a single track with details like:

- **Track Info :** track name, artist, album, year, language, track URL
- **Popularity Metric :** numerical score representing how well a song is performing
- **Audio Features:** danceability, energy, valence, acousticness, instrumentalness, speechiness, tempo, loudness, liveness, etc.
- **Musical Properties:** key, mode, time signature

Total records : 62317 (with 22 attributes)

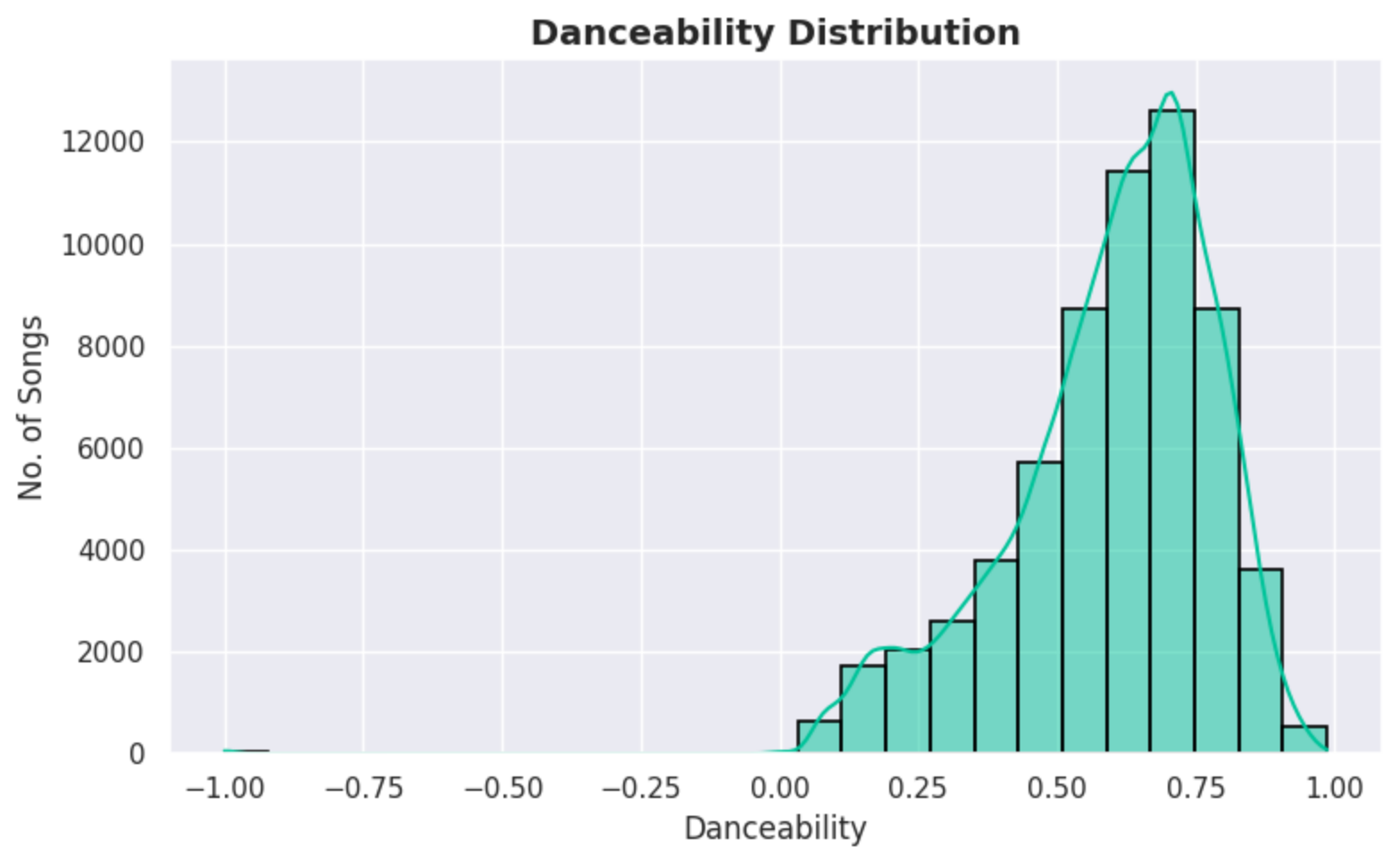
ANALYSIS AND OBSERVATIONS

Popularity Scores



Most tracks fall within 20-50 popularity; only a handful achieve global hit-level visibility (80+). The data is right-skewed, highlighting the rarity of viral songs

DANCEABILITY

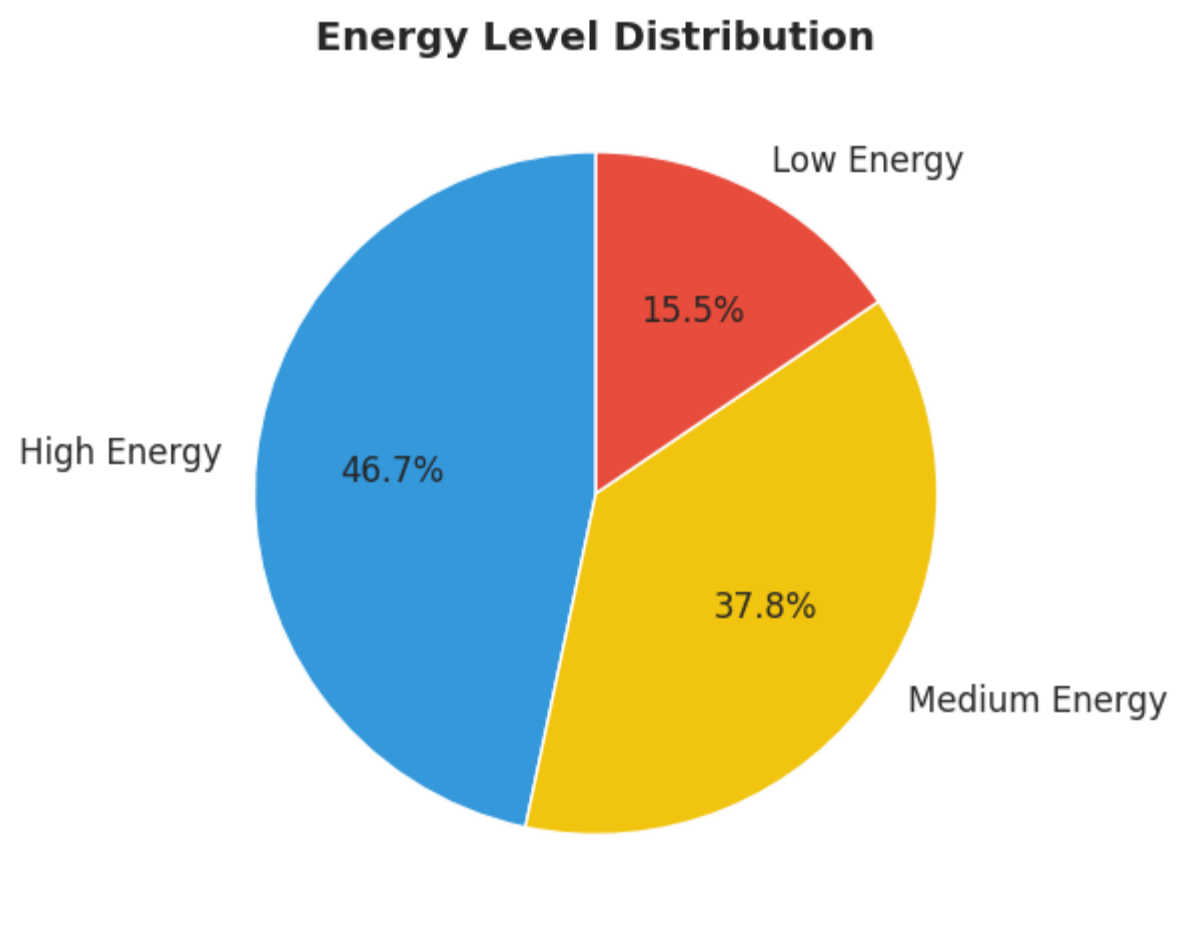


Most songs have **moderate-to-high** danceability **(0.5–0.8)**, showing Spotify's library favors rhythm-driven and upbeat tracks

Very **few tracks** have extremely low danceability **(< 0.3),** meaning most songs possess some rhythmic flow or beat alignment.

The peak density around **0.6–0.7** suggests Spotify's catalog leans toward tracks suitable for casual or upbeat listening rather than purely acoustic or spoken content.

Energy Levels



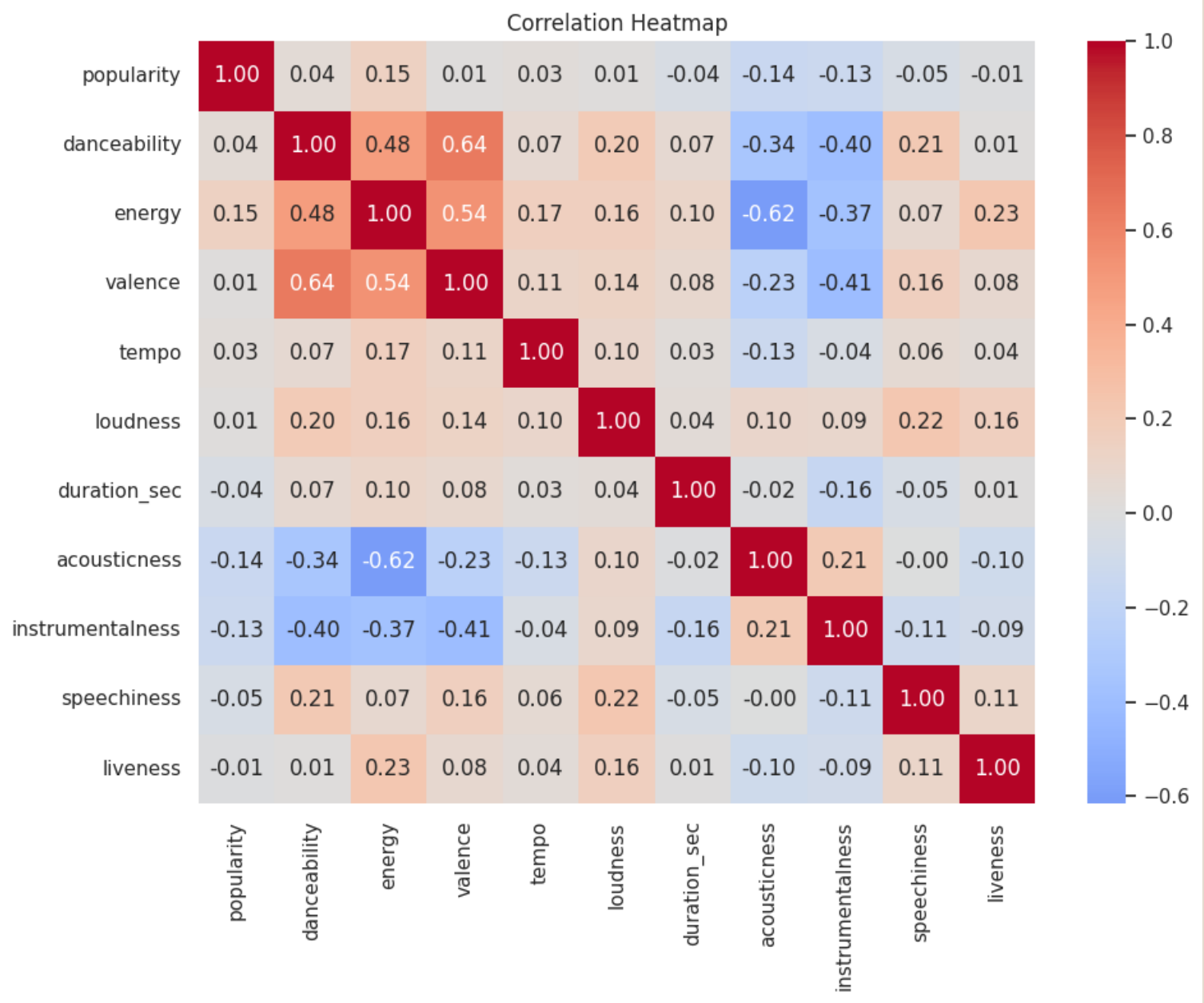
The energy level ranges are as follows :
Low Energy : 0 – 0.33
Medium Energy : 0.34 – 0.66
High Energy : 0.67 – 1.00

High Energy (0.67–1.00) dominates, comprising the majority of tracks, suggesting Spotify has a strong focus on balanced, moderately intense songs suitable for casual listening.

Low Energy (0.00–0.33) is the smallest slice, indicating that very few tracks are calm, acoustic, or mellow.

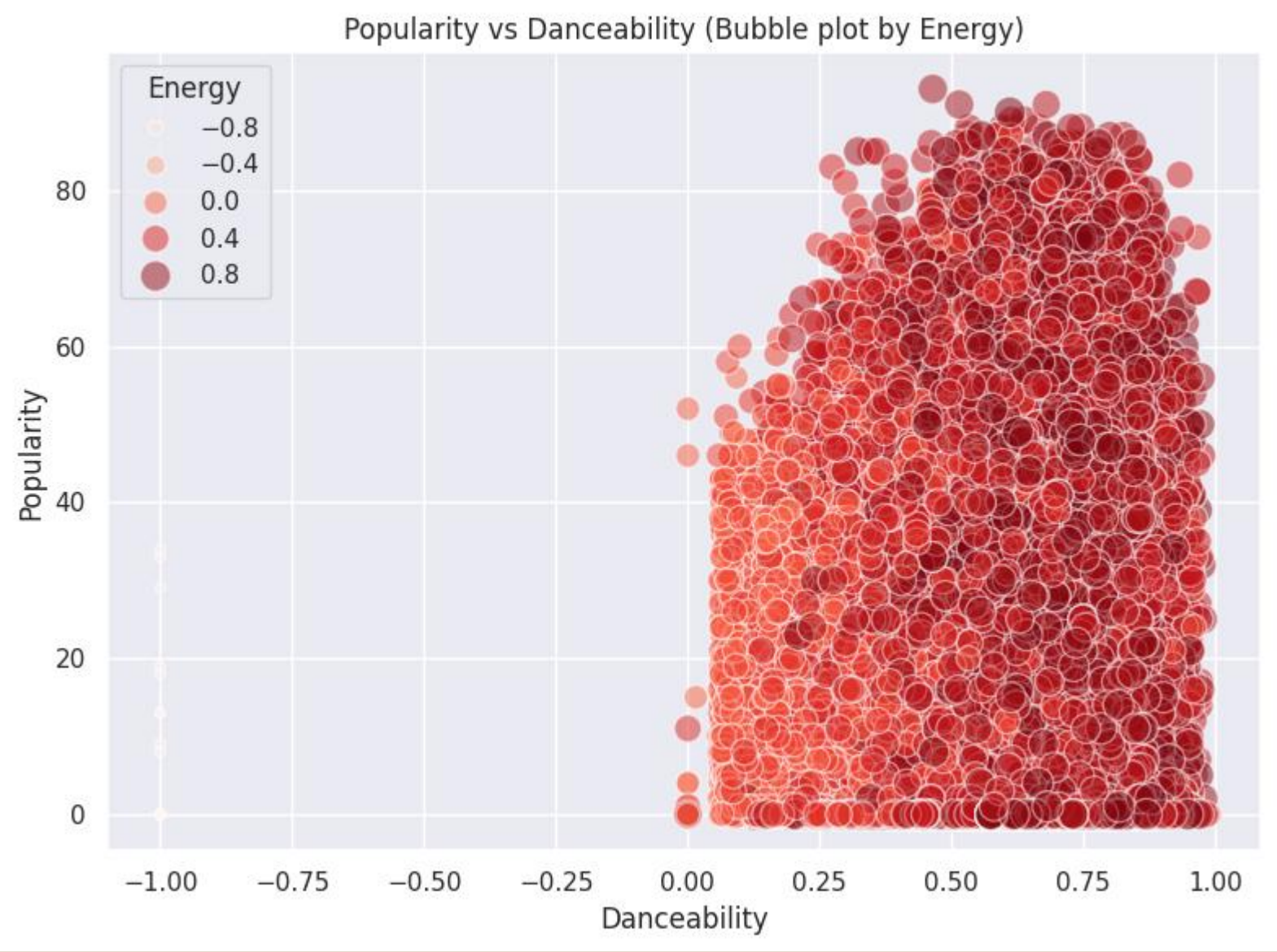
- Most Spotify tracks have high energy, with fewer high-energy songs and very few low-energy tracks, reflecting a bias toward engaging and upbeat music."

Corelation Heatmap



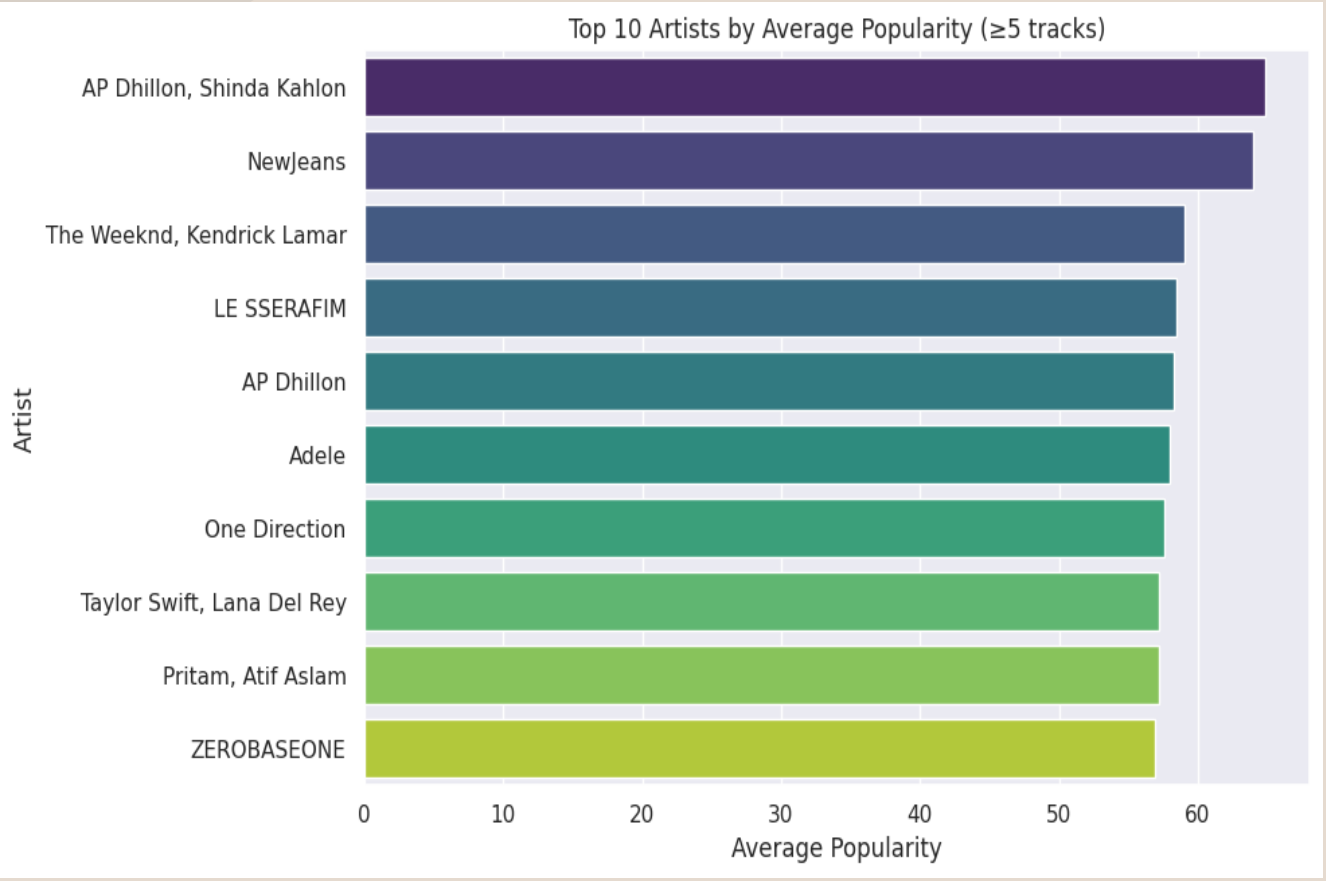
The correlation heatmap shows that energy, loudness, and danceability are positively related, while acousticness and instrumentalness tend to be negatively correlated with these features. Popularity shows only moderate links with audio attributes, indicating that hit songs are influenced by multiple factors rather than a single feature.

Population VS Danceability



- Higher danceability slightly boosts popularity, but the effect is modest.
- Popularity varies widely across danceability and energy levels.
- Outliers show some very popular songs at different danceability/energy, highlighting influence of other factors.

Top Artists by Average Popularity

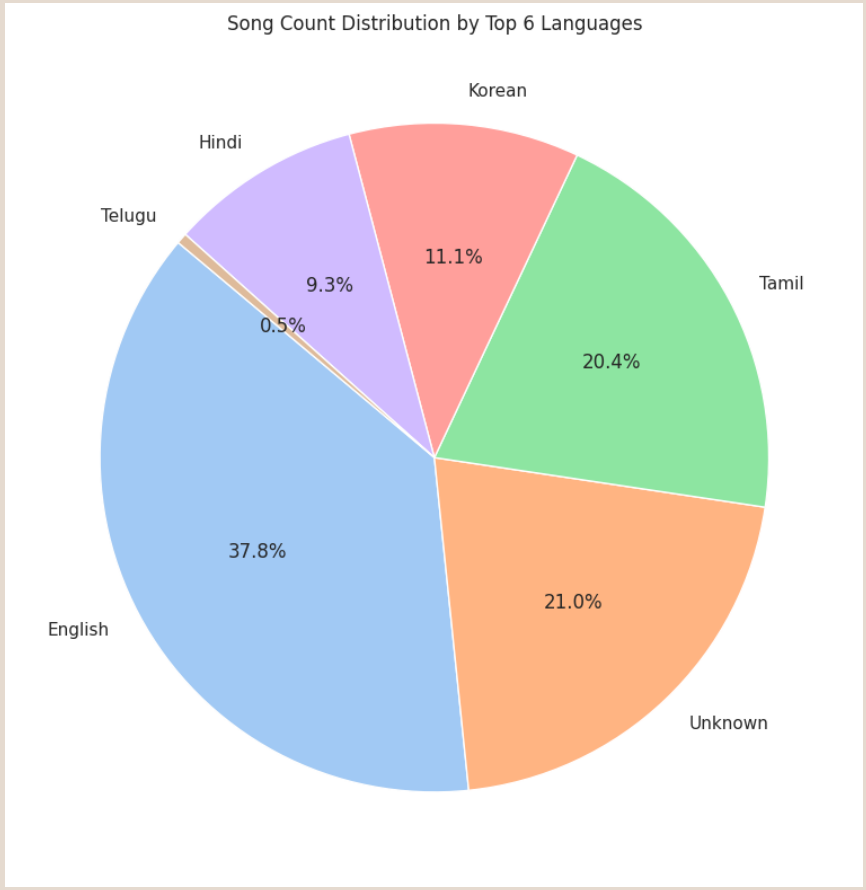


Consistent high performers: These artists maintain high average popularity across multiple tracks, indicating a strong and engaged fanbase.

Filter ensures reliability: By including only artists with 5 or more tracks, the ranking reflects sustained success rather than a single hit.

Variation in popularity: There is a visible gap between the top artist and others, suggesting some artists consistently outperform peers in listener reach and engagement.

Popularity by Language

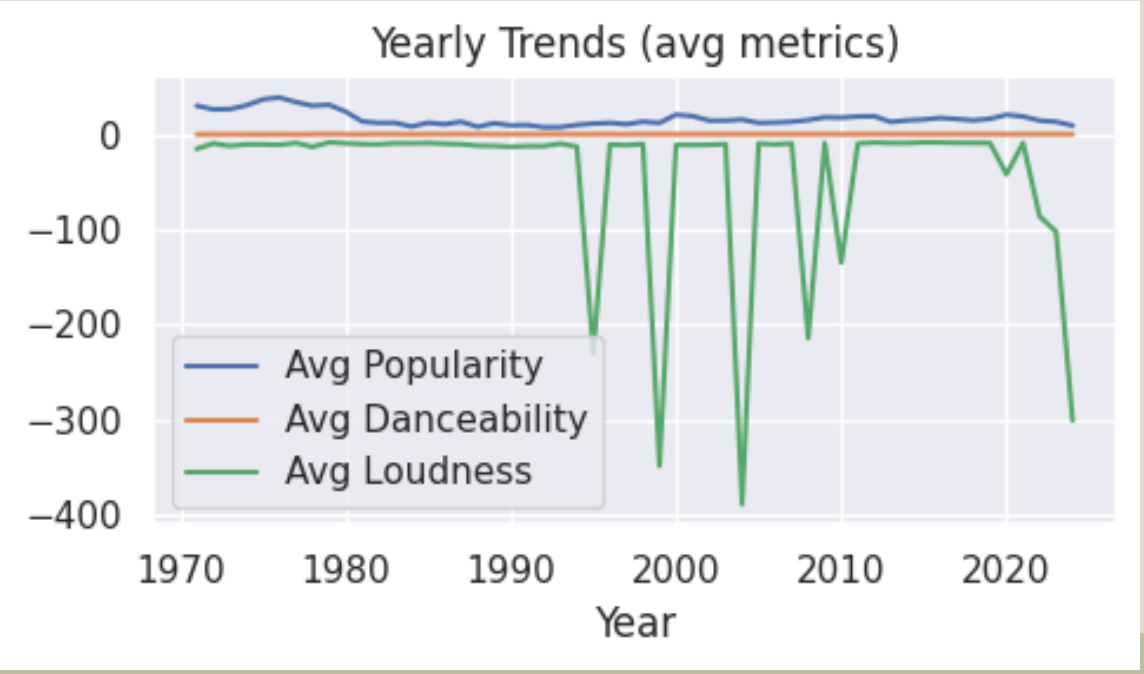


English dominance: English tracks often show higher median popularity compared to other languages, reflecting their wider global audience.

Wide variability: Popularity varies significantly within each language, indicating that hits can appear in any language but with different likelihoods.

Outliers: Some songs achieve exceptionally high popularity regardless of language, highlighting that individual hits can transcend general trends.

Yearly Trends

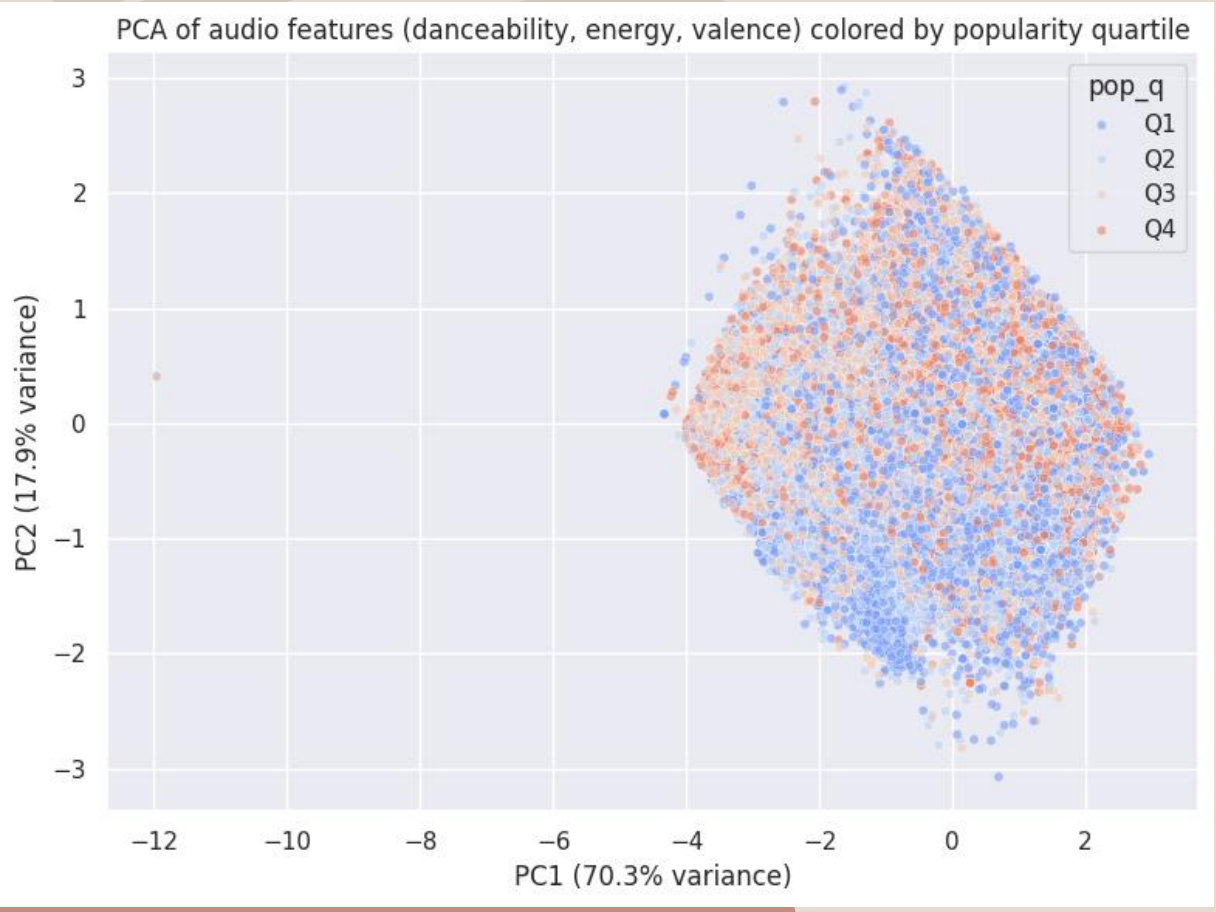


Popularity trends: Average popularity shows fluctuations over the years, reflecting shifts in listener preferences and mainstream music trends.

Danceability evolution: There is a gradual change in average danceability, suggesting that songs are becoming slightly more or less danceable over time.

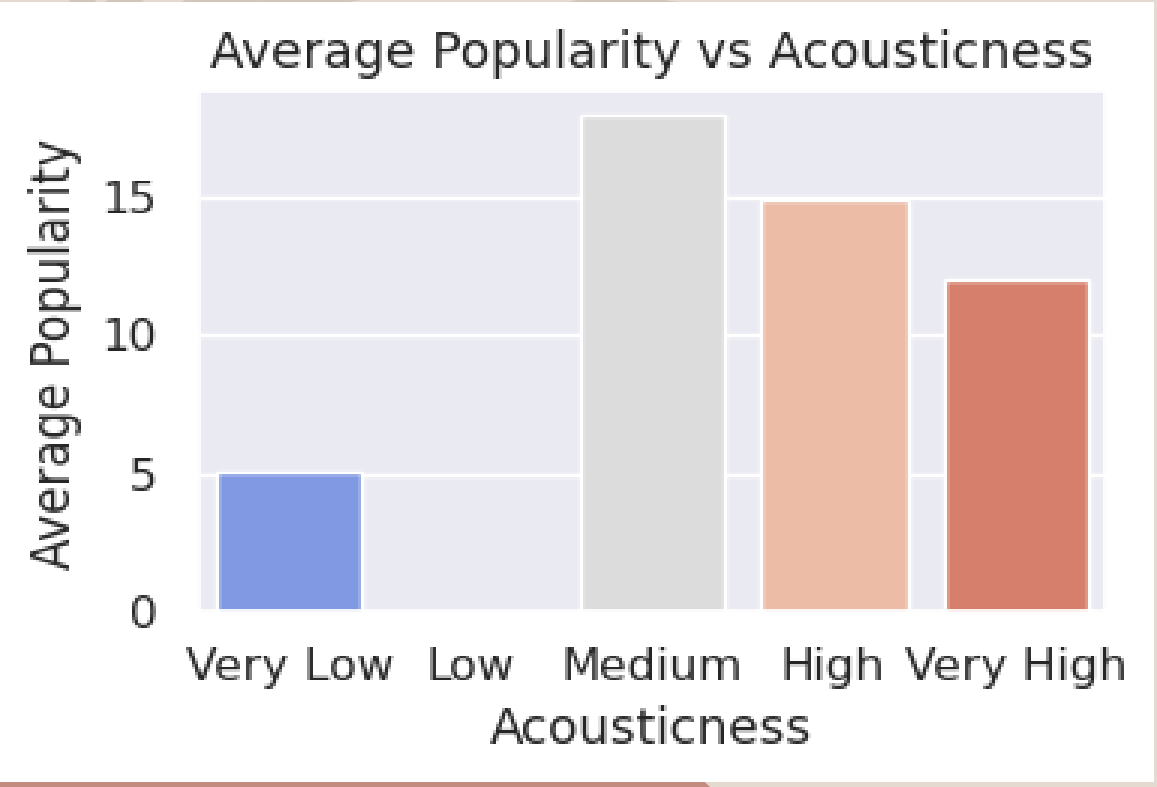
Loudness trends: Average loudness varies across years, possibly reflecting changes in production practices, such as mastering styles or the "loudness war."

PCA Scatter



- Songs form distinct clusters along PC1 and PC2, reflecting natural groupings of danceability, energy, and valence.
- Popularity varies across the PCA space, showing audio features influence but don't fully determine hits.
- PC1 explains most variance (energy & danceability),
- PC2 adds variation (valence/other features).

Popularity VS Acousticness



- **Lower Acousticness = Less Popularity**
  More produced/electronic tracks tend to be less popular.

- **High Acousticness = Moderate Popularity**
  Unplugged or acoustic-heavy songs are moderately popular.

- **Moderate Acousticness = Highest Popularity**
  Medium acousticness yields highest popularity — listeners prefer a mix of acoustic and produced sounds.

Conclusion:

- **Popularity Distribution**: Most tracks have moderate popularity (scores between 20–50).
  Only a few songs reach global hit status (80+).

- **Danceability:** Majority of songs are moderately to highly danceable (0.5–0.8).
  Indicates Spotify's catalog leans toward upbeat and rhythm-driven music.

- **Energy Levels:** High-energy songs dominate the dataset.
  Very few calm or low-energy tracks exist, showing preference for lively music.

- **Feature Correlations**: Energy, loudness, and danceability are positively correlated.
  Acousticness and instrumentalness are negatively correlated with these.
  Popularity has only moderate correlation with audio features — multiple factors influence a hit.

- **Language Impact**: English tracks generally have higher popularity.
  However, exceptional hits exist in all languages.

Contd...

- **Artist Performance**: Top artists consistently maintain higher average popularity.
  Sustained success depends on consistent audience engagement, not one-time hits.

- **Yearly Trends**: Popularity, danceability, and loudness vary over the years.
  Reflects evolving listener preferences and production styles.

- **Acousticness Insight**: Moderate acousticness (a mix of acoustic and electronic) yields the highest popularity. Purely acoustic or overly electronic songs are less favored.

- **Overall Observation**: Popularity depends on a balance of musical features rather than extremes.
  Spotify listeners generally prefer energetic, danceable, and emotionally positive tracks.

THANK YOU!

PRESENTED BY:

TANVIR