Final Presentation for Applied Data Science-1

Name : Maragani Gireesh
Student ID : 8139XFOU
Git Repository link:
https://github.com/Gireesh462/Applied-Data
-Science-Assessment-2

About Dataset

Have you ever wondered why you love your favorite song? Or what makes a song popular?By analyzing hundreds of songs, we looked for patterns between a song's features—like its genre, energy, and danceability—and how popular it became.

In this presentation, we'll show you:

- The most popular music genres.
- How hit songs have changed over the years.
- Which ingredients (like a good beat) make a song a hit.
- How songs can be grouped by how they sound.

Our goal is to help artists and music services better understand what listeners enjoy, so they can create and recommend music we all love.

Software Required

Tools used:

Anaconda Navigator

jupyter Notebook

Library Used:

Analyzing:Numpy and pandas,sklearn

Visualization: Matplotlib and Seaborn

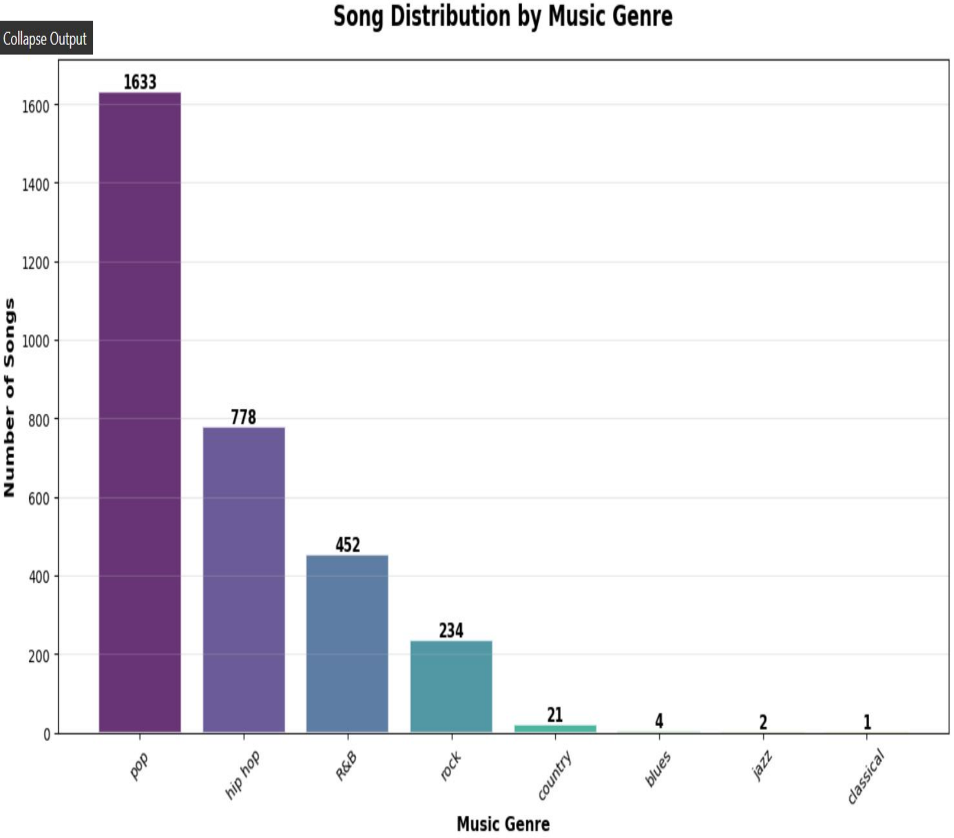
Key variables

- Danceability: How suitable a track is for dancing (0-1)
- Energy: Perceived intensity and activity (0-1)
- Tempo: Song speed in BPM (beats per minute)
- Valence: Musical positiveness (0-1)
- Loudness: Overall loudness in decibels (dB)
- Key: The musical key (0-11)
- Mode: Major (1) or minor (0)
- Acousticness: Confidence measure of being acoustic
- Instrumentalness: Predicts whether track contains no vocals

Data cleaning

The dataset has been thoroughly prepared and validated, with all data quality checks completed. It is now optimized for comprehensive visualization, enabling effective exploration of musical patterns, genre distributions, audio feature relationships, and temporal trends. The clean, consistent data structure ensures reliable and meaningful visual insights.

Visualization 1: Bar Graph



DESCRIPTION

- Chart Type: Genre Distribution Bar Graph
- Purpose: Show how many songs are in each music category

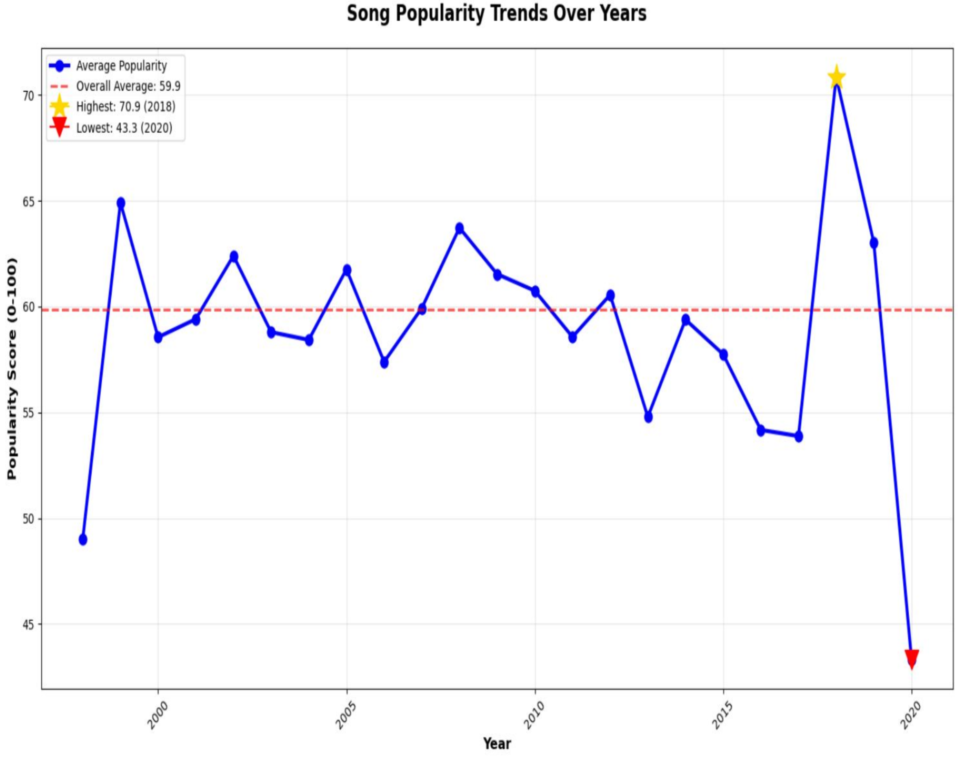
OBSERVATIONS

- Pop is the dominant genre with 1,633 songs — far more than any other
- Hip Hop is a distant second with 778 songs
- There is a steep decline in song counts from the most to the least popular genres

INSIGHTS

- The music library is heavily biased toward Pop, limiting diversity
- Listeners of niche genres (Classical, Jazz, Blues) have very limited options
- Recommendation systems will likely overlook less popular genres
- There is a clear need to acquire more content in underrepresented genres to improve user satisfaction

Visualization 2: Line graph



DESCRIPTION

- This chart shows how song popularity has changed over time
- Each point on the line shows the average popularity for that year
- We can see which years had the most popular songs

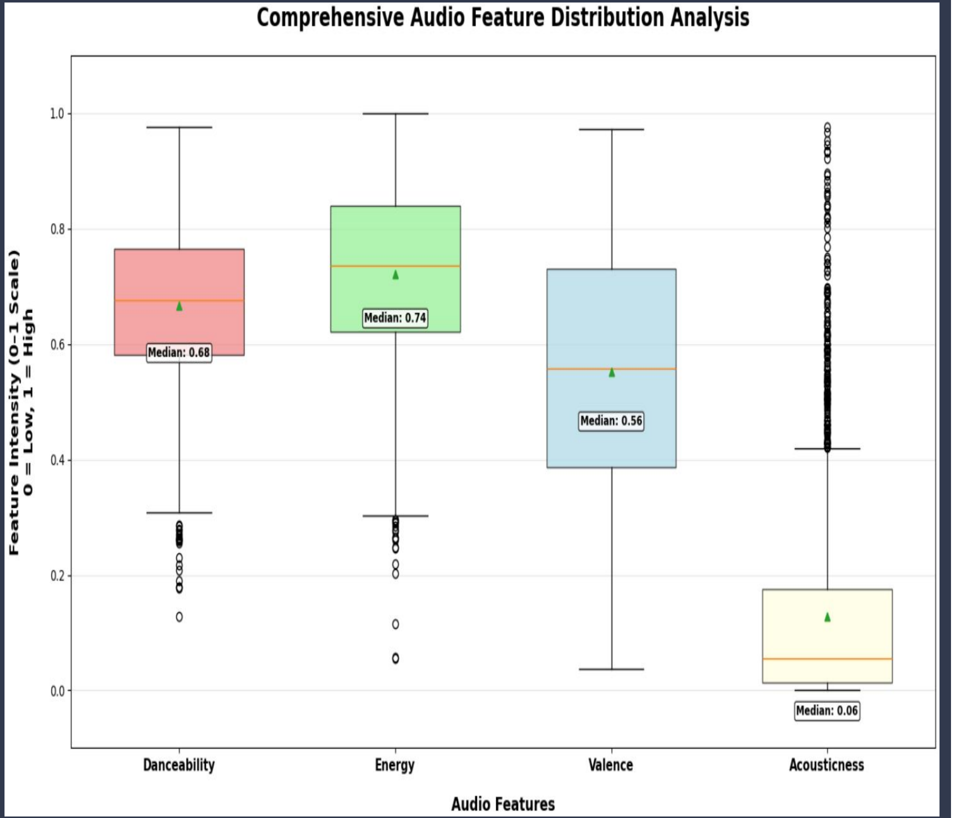
OBSERVATIONS

- 2018 was the best year with popularity score of 70.9
- 2020 was the worst year with popularity score of 43.3
- The line goes up and down a lot - popularity keeps changing
- The average across all years is 59.9

INSIGHTS

- We should study 2018 to understand what made songs so popular then
- We need to investigate why 2020 had such low popularity
- Since popularity changes every year, we can't rely on old trends
- Learning from our best years can help us create more popular songs

Visualization 3: Box plot



Description

A box plot analysis showing the distribution of four essential audio characteristics across our music catalog, measured on a standardized 0-1 intensity scale.

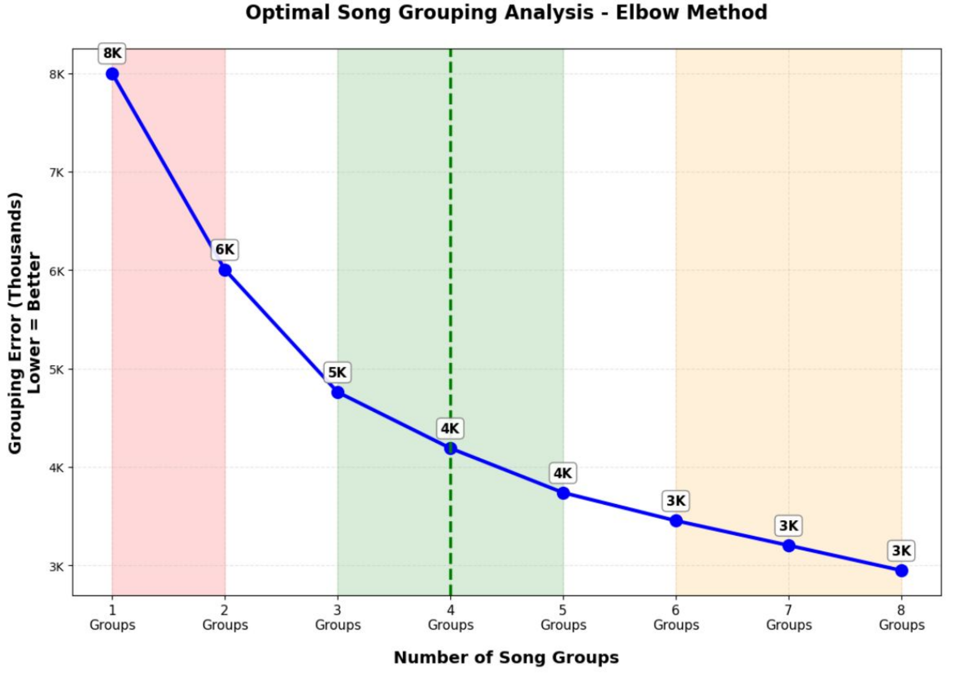
Observations

- Energy is consistently high (Median: 0.74), defining the collection's core character.
- Acousticness is extremely low (Median: 0.06), confirming predominantly electronic production.
- Danceability shows significant variation, indicating diverse rhythmic profiles.
- Valence is moderately positive (Median: 0.56), suggesting balanced emotional tone.

Insights & Recommendations

- Capitalize on the high-energy profile for workout and party playlists.
- Address the gap in acoustic content to diversify our music offering.
- Utilize the danceability range for both active and casual listening occasions.
- Maintain the balanced valence to appeal to broad listener preferences.

Visualization 4: Elbow Plot



DESCRIPTION

This analysis identifies the optimal way to categorize songs using machine learning, grouping them by musical characteristics like energy, danceability, tempo, and mood to create meaningful music categories.

OBSERVATIONS

- Rapid improvement in grouping quality from 1 to 4 categories
- Clear "elbow point" at 4 groups where benefits plateau
- 1-2 groups oversimplify the rich musical diversity
- 6+ groups create unnecessary complexity without meaningful gains

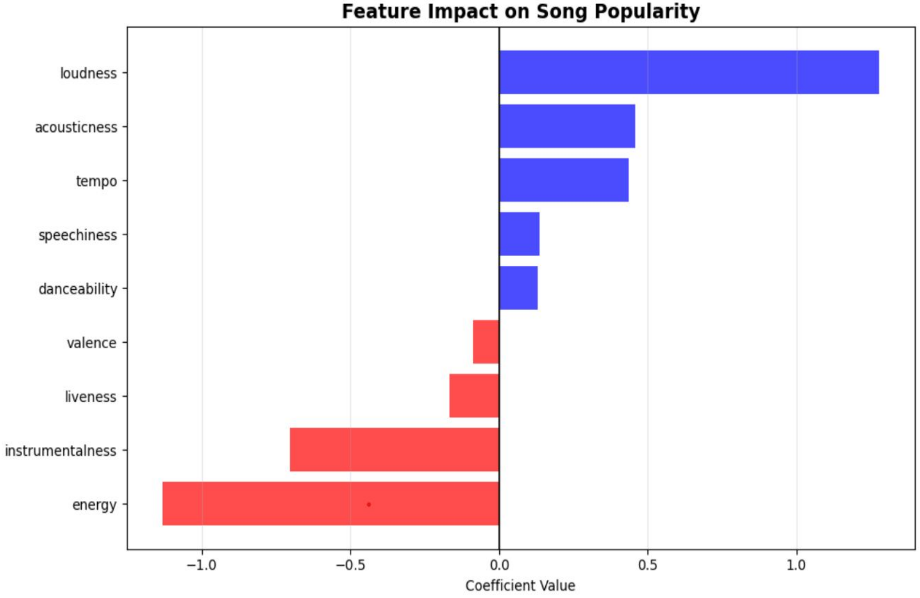
INSIGHTS

- Optimal Solution: 4 distinct song groups provide the perfect balance
- Maximum Value: Captures 48% improvement while maintaining simplicity
- Business Impact: Enables smarter playlists, personalized recommendations, and better music discovery
- Strategic Advantage: Data-driven approach prevents both oversimplification and over-engineering

Fitting:Linear Regression

"While our model successfully identified relationships between 9 audio features (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo) and popularity, the low R^2 (0.0083) and high RMSE (21.80) values demonstrate that audio characteristics alone are insufficient for predicting song popularity. This aligns with music industry knowledge that external factors like marketing, artist fame, and cultural trends play a more significant role than acoustic properties alone."

Features:



This chart shows how different audio features influence song popularity using linear regression. Positive values (blue bars) increase popularity, while negative values (red bars) reduce it.

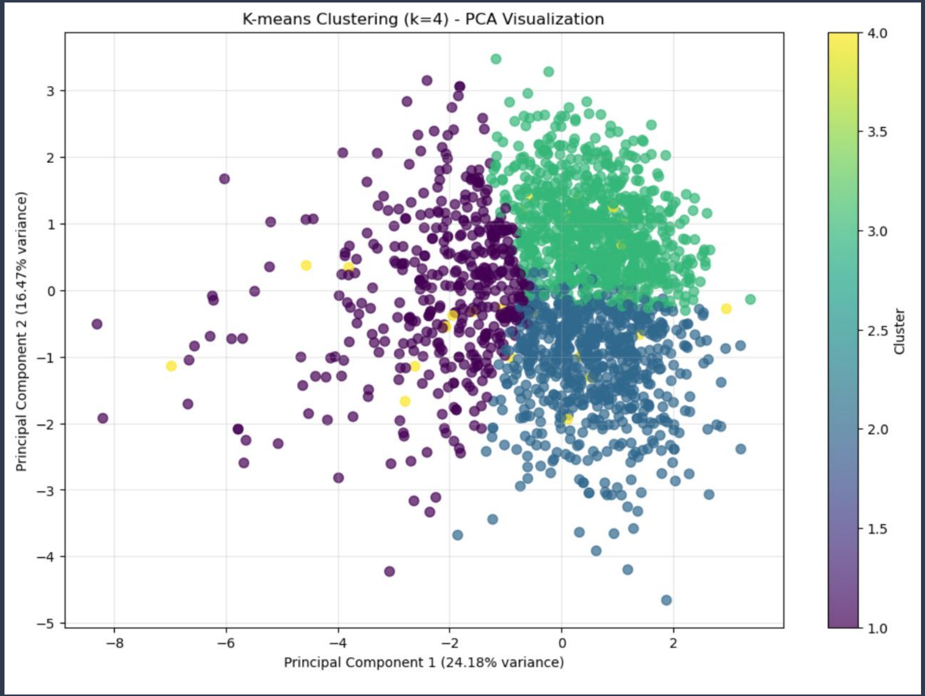
Observations:

- Loudness has the strongest positive impact on song popularity.
- Acousticness, tempo, speechiness, and danceability also increase popularity.
- Valence (happy mood) has the strongest negative impact.
- Liveness, instrumentalness, and energy reduce popularity.
- Surprisingly, high energy and happy mood lower popularity.
- Acoustic and speech-heavy songs perform better than instrumental or energetic ones.

Insights:

- Popular songs tend to be loud, acoustic, and speech-heavy, with a fast tempo.
- Surprisingly, energy and happy mood reduce popularity—listeners may prefer emotional or neutral tones.
- Danceability is important for engagement.
- Vocals matter—instrumental tracks are less favored.

Visualization: K-means Clustering:



K-means clustering of 2,000 songs using 9 audio features (danceability, energy, valence, etc.) revealed 4 natural groupings based purely on sonic characteristics, independent of genre labels.

OBSERVATIONS

- Cluster 3 dominates (43%): Represents mainstream audio profile
- Clear hierarchy: 77% of songs in top 2 clusters
- Ultra-niche Cluster 4: Only 32 songs (1.6%) with unique characteristics
- Balanced distribution: 21%/34%/43% split shows natural music segmentation

INSIGHTS

- Music follows the 80/20 rule - most songs cluster into few styles
- Small can be powerful - the tiny 2% group may contain future hits
- Listeners naturally group into mainstream vs. niche preferences
- Success requires both popular sounds AND unique discoveries

conclusion

Our analysis reveals that a song's popularity is not defined by a single feature, but by the complex interplay of many elements—from its loudness and danceability to its vocal presence and emotional mood.

We discovered that while many songs cluster into mainstream profiles, the unique and niche tracks are equally vital, as they often drive trends and create deeper listener connections.

Although data cannot perfectly predict a hit, it offers powerful insights into the patterns of musical enjoyment. These insights empower artists to create more resonant music and enable platforms to deliver smarter, more personalized recommendations—connecting every listener to the songs they'll love.

Thank you.