Spotify Data Analysis Project Report

Spotify Data Analysis Project - Comprehensive Report

1. Project Background and Motivation

Music streaming services, especially Spotify, have revolutionized how we consume music. Understanding what makes certain tracks popular can be beneficial for artists, producers, and marketers. This project analyzes Spotify's dataset to explore how various audio features influence the popularity of a song.

2. Objectives

- Perform in-depth exploratory data analysis (EDA)
- Identify key audio features driving popularity
- Analyze trends over time
- Provide actionable insights for the music industry
- Lay the groundwork for future machine learning models

3. Dataset Description

The dataset consists of Spotify tracks with the following columns:

- Track Name
- Artist Name
- Popularity Score (0-100)
- Danceability
- Energy
- Key
- Loudness (dB)

- Mode (Major/Minor)
- Speechiness
- Acousticness
- Instrumentalness
- Liveness
- Valence
- Tempo (BPM)
- Duration (ms)
- Time Signature
- Release Date

Dataset Summary:

- Number of records: ~2000 tracks
- Time span: Varies by dataset (2010-2022)

4. Data Cleaning and Preparation

- Checked for and handled null values
- Converted duration_ms' to minutes
- Converted release_date' to datetime format
- Verified data types and corrected mismatches
- Removed duplicates to prevent skewed analysis

5. Exploratory Data Analysis (EDA)

5.1 Popularity Distribution

- Most tracks had moderate popularity between 40-70.

- Very few songs reached popularity scores above 85.

5.2 Correlation Analysis

Generated a heatmap to examine correlations:

- **Positive correlation:** Danceability, Energy, and Valence with Popularity

- **Negative correlation:** Loudness and Acousticness inversely impact popularity

5.3 Feature Impact Analysis

- **Danceability:** Highly danceable tracks generally gained higher popularity.

- **Energy:** Tracks with high energy levels were more popular.

- **Valence:** Songs with positive vibes (high valence) performed better.

- **Acousticness:** Tracks with high acoustic values were less popular.

6. Visualizations and Interpretations

6.1 Popularity Histogram

- Showed that most songs cluster around a moderate popularity score.

- Very few outliers reach the 90+ range.

6.2 Danceability vs Popularity Scatter Plot

- Displayed a positive trend indicating that danceable tracks perform better.

6.3 Energy vs Popularity

- Moderate correlation suggesting energetic tracks are better received.

6.4 Release Year Analysis

- Line plot showed that over the years, popular songs have generally become faster and louder.

6.5 Tempo Analysis

- Popular songs mostly clustered around 120-140 BPM, common for dance/pop genres.

7. Key Findings and Insights

- **Danceability and Energy are vital features** for popular tracks.

- **Valence (positivity)** adds to a track's mass appeal.

- **Acoustic songs** may perform less on average but still find niche popularity.

- **Release Date Trends:** Newer songs show increasing energy and tempo trends, reflecting changing listener preferences.

8. Recommendations

- **For Artists/Producers:** Focus on high-energy, danceable tracks with positive vibes.

- **For Marketers:** Promote tracks that score high in danceability and valence.

- **For Data Scientists:** Potential to build predictive models based on these features.

9. Future Work

- Genre-based analysis for more precise insights.

- Sentiment analysis on lyrics vs audio features.

- Apply machine learning models to predict popularity.

10. Conclusion

The Spotify dataset offers rich insights into music trends. Our analysis highlights how audio features like danceability, energy, and valence significantly impact a track's success. The findings serve as a

foundation for deeper studies, including machine learning models that predict a track's popularity based on these features.

Tools Used: Python, Pandas, Matplotlib, Seaborn, Numpy, Jupyter Notebook

Prepared By: Data Science Team