Analysis of 30000 Spotify Songs Dataset

Introduction

Background

Spotify, as a leading music streaming service, offers a vast collection of songs across various genres. This project aims to analyze the '30000 Spotify Songs' dataset to extract meaningful insights from the audio features and distribution of songs.

Objective

The primary goal of this analysis is to explore the distribution of songs across different genres and subgenres, understand the temporal distribution of song releases, and examine various audio features such as danceability, energy, and valence.

Methodology

Data Collection

The dataset was sourced from Spotify through the `spotifyr` package, which provides extensive data on song characteristics.

Data Cleaning and Preparation

The dataset underwent a cleaning process where missing values were handled, and necessary data type conversions were made, particularly for the album release dates.

Exploratory Data Analysis

The analysis consisted of examining the distribution of songs across genres and over time, as well as a detailed investigation of various audio features. Tools used in this process included Python with libraries such as Pandas, Matplotlib, and Seaborn.

Findings

Genre and Subgenre Distribution

- The dataset displayed a wide range of musical genres, with pop, rap, and rock being the most common.
- Visualizations showed the quantitative distribution of songs across these genres and the top 20 subgenres.

Temporal Distribution of Songs

- Analysis of the album release dates revealed trends in the volume of music released over the years.
- A notable increase in the number of songs over recent years was observed, possibly reflecting the platform's expanding collection.

Audio Features Analysis

- Songs exhibited a broad spectrum of audio features.
- Key observations included:
    - A general tendency towards tracks with higher danceability.
    - Energy levels were also relatively high, indicating more dynamic and intense tracks.
    - Loudness, tempo, and duration varied, reflecting diverse music production styles.

Feature Variation by Genre

- Different genres showed distinct audio profiles. For instance:
    - Pop and EDM genres tended to have higher danceability.
    - Rock and rap were characterized by higher energy levels.
    - Emotional content, as indicated by valence, varied significantly across genres.

Conclusion

The analysis provided valuable insights into the nature of music available on Spotify. It highlighted the diversity in genres and the unique characteristics of songs within each genre. These findings can be instrumental in understanding current music trends, artist styles, and listener preferences.

Future Work

Further research could include:

- A deeper temporal analysis to understand changing trends in music over the years.
- Cluster analysis to identify groups of similar songs.
- Predictive modeling to forecast song popularity or success.

Appendices

Include any additional data, code snippets, or detailed methodology here.