

Spotify Data Analytics Capstone – Full Solved PDF

Objective

The objective of this project is to analyze Spotify music data to identify trends, understand user preferences, and explore song characteristics using data analytics and machine learning techniques.

Dataset Description

The dataset consists of Spotify tracks with attributes such as danceability, energy, tempo, loudness, valence, and popularity. These features describe musical characteristics and listener engagement.

Data Loading & Cleaning

The dataset was loaded into a Jupyter Notebook using Python libraries such as Pandas and NumPy. Missing values were handled by removing incomplete records, and duplicate entries were eliminated to ensure data integrity.

Exploratory Data Analysis (EDA)

EDA was conducted to understand feature distributions and relationships. Danceability and energy showed right-skewed distributions, while tempo followed a near-normal pattern.

Trend Analysis

An analysis of popularity over time revealed that songs released in recent years generally have higher popularity, reflecting the growth of streaming platforms.

Correlation Analysis

Energy and loudness showed a strong positive correlation. Danceability and energy demonstrated moderate correlation with popularity.

User Preference Insights

Highly popular songs tend to be more energetic, louder, and more danceable than less popular tracks.

Predictive Modeling

Two predictive models were built: Linear Regression and Decision Tree Regressor. The Decision Tree model outperformed Linear Regression due to its ability to capture non-linear relationships.

Model Evaluation

Model performance was evaluated using Mean Squared Error (MSE) and R-squared metrics. The Decision Tree achieved better accuracy and explained more variance in popularity scores.

Conclusion

This project demonstrates how Spotify audio features can be used to understand music trends and predict song popularity. Data-driven insights can assist artists, producers, and streaming platforms.

Future Scope

Future work may include advanced machine learning models such as Random Forest or XGBoost, genre-based analysis, lyrics sentiment analysis, and user behavior data integration.