DSA210 Project

Music Listening Habits and Productivity Analysis Report

Introduction

This report examines the relationship between music listening habits and productivity during work or study sessions. By analyzing music preferences, task performance, and working hours, the project uncovers actionable insights to optimize productivity through personalized music recommendations. Utilizing data-driven techniques, such as exploratory data analysis (EDA), machine learning models, and visualizations, this report delivers findings supported by both statistical analysis and visual results.

Data Preprocessing

Key Steps:

1. **Outlier Removal:**
    - Restricted tempo to 60-200 BPM and loudness to a normalized range of -5 to 5.
    - Excluded rows with productivity scores of 0 to reduce noise.
2. **Missing Value Handling:**
    - Imputed missing values using genre-specific medians, defaulting to overall medians if unavailable.
3. **Feature Engineering:**
    - Created interaction features like Tempo_Loudness_Interaction.
    - Encoded genres numerically using target encoding.
4. **Data Aggregation:**
    - Aggregated session-level metrics for tempo, loudness, and genre count.

Findings and Analysis

Clustering Results

Using K-Means clustering, three distinct groups were identified based on tempo, loudness, and productivity levels. The clustering quality was validated using Silhouette Scores.

Cluster	Avg Productivity	Avg Tempo (BPM)	Avg Loudness (Normalized)	Genre Preference Score
0	77.93	126.48	0.91	77.51
1	63.99	165.23	0.94	65.37
2	62.23	126.78	0.81	62.63

Cluster Interpretations:

- Cluster 0: High productivity is associated with moderate tempo and balanced loudness.
- Cluster 1: Moderate productivity aligns with higher tempo and loudness levels.
- Cluster 2: Low productivity corresponds to moderate tempo and lower loudness.

Correlation Insights:

- The correlation matrix shows weak negative correlations between tempo and productivity (-0.068), and between loudness and tempo (-0.18), suggesting tempo and loudness have limited direct impact individually but are more influential in combination.
- A slight positive correlation (0.049) exists between loudness and productivity, indicating higher loudness may improve productivity to a certain extent.

Genre-Based Observations:

- Genres like classic soul, funk, and motown have the highest productivity scores.
- Hip-hop and trap genres are associated with moderate productivity.
- Dance-pop and melodic house genres align with less productive tasks, offering a more relaxed experience.

Tempo and Loudness Ranges:

- High productivity: Tracks with moderate tempo (120-130 BPM) and normalized loudness around 0.90.

- Moderate productivity: High tempo tracks (>160 BPM) with normalized loudness around 0.94.
- Relaxed sessions: Tracks with moderate tempo (125-130 BPM) and lower loudness (0.75-0.85).

Visual Findings:

- Scatterplots: Clear separations between clusters based on tempo and loudness.
- Bar Charts: Genres like classic soul dominate high-productivity clusters.
- Feature Importance: Genre encoding had the highest influence on productivity predictions, followed by tempo-loudness interaction.

Recommendations by Task Type

For High Productivity:

- Genres: Classic soul, funk, motown.
- Tempo: 120-130 BPM.
- Loudness: 0.90 or higher.
- Example Songs: Tracks with consistent rhythms and minimal vocal interruptions.

For Moderate Productivity:

- Genres: Hip-hop, trap, and upbeat pop.
- Tempo: Above 160 BPM.
- Loudness: Around 0.94.
- Example Songs: Energizing beats with dynamic instrumentation.

For Relaxed or Chill Sessions:

- Genres: Dance-pop, melodic house, indie poptimism.
- Tempo: 125-130 BPM.
- Loudness: 0.75-0.85.
- Example Songs: Ambient tracks with mellow beats.

Model Performance

Random Forest Regression:

- Best Parameters: max_depth=3, n_estimators=50.
- Performance: Explained ~54.7% of variance, RMSE = 9.76.
- Demonstrated improved accuracy after removing noisy rows.

Linear Regression:

- Provided moderate predictive power but less accurate than Random Forest.

Key Observations:

- Higher tempo correlates with mid-level productivity (Cluster 1).
- High productivity (Cluster 0) is associated with balanced tempo and loudness values.
- Cluster 2 has the lowest loudness preference and productivity.

Progress and Improvements

Initial Challenges:

1. Dataset preprocessing required normalization of attributes like loudness.
2. Feature selection was iterative, ensuring attributes contributed to meaningful clusters.
3. Initial clustering resulted in overlapping groups.

Resolution Steps:

- Preprocessing: Implemented robust normalization and outlier handling.
- Clustering: Adjusted hyperparameters to refine cluster separations.
- Visualization: Enhanced clarity with scatterplots and pairplots to highlight trends.

Iterative Improvements:

1. Fine-tuned the clustering algorithm for sharper group delineations.
2. Added genre encoding to incorporate user preferences comprehensively.
3. Developed a recommendation system tailored to each cluster.

Discussion

Insights:

1. Productivity patterns are influenced by tempo and loudness preferences.
2. Specific genres and their associated tempo/loudness ranges align with productivity goals.
3. Clustering highlights how individual traits can guide personalized recommendations.

Applications:

- The methodology can extend to real-time music recommendation systems.
- Insights may support industries like fitness, education, and mental health by optimizing productivity environments.

Conclusion

This project successfully implemented a clustering approach to derive actionable insights on productivity optimization through music preferences. The iterative development ensured robust analysis, with clear visualizations and meaningful recommendations for each user segment.