

Report on the Correlation Between Music Listening and Physical Activity

**Simay Reel
31190**

1. Introduction

This report investigates the correlation between music listening habits (Spotify listening time) and physical activity levels (daily steps and calories burned). Using exploratory data analysis (EDA) and machine learning techniques, the study examines the hypothesis:

- **H₀:** The more music I listen to in a day, the more I walk and burn calories.
- **H_a:** The amount of music listened to in a day does not have a significant positive relationship with the number of steps taken and calories burned.

The report discusses the methodology, data preparation, analysis, and findings, followed by conclusions and recommendations for future work.

2. Data Sources and Preparation

2.1 Spotify Listening Data

- Extracted from JSON files containing timestamps, track names, and durations.
- Preprocessed into daily aggregated listening times (in minutes).

2.2 Health Data

- Extracted from XML files, including metrics for daily steps and calories burned.
- Converted to CSV format with consistent date formatting and cleaned for analysis.

2.3 Merged Dataset

- Datasets were merged using the date as a common key, ensuring alignment and consistency across all features.

3. Methodology

3.1 Exploratory Data Analysis (EDA)

- Analyzed daily trends for listening time, steps, and calories burned.
- Computed correlation coefficients and visualized relationships using scatter plots and heatmaps.
- Identified and removed outliers to improve data quality.
-

3.2 Machine Learning Analysis

- Built linear regression models and decision tree regressors to predict physical activity metrics based on listening time.
- Applied log transformation to normalize data and re-evaluated correlations.

4. Findings

4.1 Exploratory Data Analysis

- **Daily Trends:**
Peaks in music listening often occurred in the evenings, but no consistent relationship was observed with steps or calories burned.
- **Correlation Analysis:**
Listening Time vs. Steps: ~ 0.00 (no linear correlation).
Listening Time vs. Calories Burned: ~ 0.04 (very weak positive correlation).
Steps vs. Calories Burned: ~ 0.05 (weak correlation).
- **Outliers:**
Listening Time: Stable with no significant outliers.
Steps: High variability with outliers above 800 steps.
Calories Burned: Minimal variability, with stable trends.

4.2 Machine Learning Models

- Linear regression and decision tree models demonstrated weak predictive power ($R^2 \sim 0.05$).
- Log transformations did not significantly improve results.
- Scatter plots reinforced the lack of strong relationships.

5. Discussion

5.1 Key Observations

- No statistically significant correlation was found between Spotify listening time and physical activity metrics.
- High variability in daily physical activity likely diluted potential correlations.
- External factors such as mood, weather, and routines might independently influence metrics.

5.2 Limitations

- The dataset did not account for factors like music genre, mood, or activity types beyond walking.

- The study's timeframe and sample size may not adequately represent broader trends.
- Linear regression assumes a specific relationship that may not exist in this context.

6. Visualizations

- Daily trends plotted for listening time, steps, and calories burned.
- Heatmaps and scatter plots visualizing correlations.
- Weekly averages to highlight broader trends.

7. Conclusion

- The hypothesis that music listening time positively correlates with physical activity levels was not supported. Consequently:
 - Reject the null hypothesis (H_0).
 - Fail to reject the alternative hypothesis (H_a).
- The weak correlations observed suggest that music listening time does not significantly influence physical activity metrics like steps or calories burned.

8. Recommendations for Future Work

- Incorporate Additional Variables: Explore the role of music genre, mood, or concurrent activities.
- Expand the Dataset: Collect data over a longer period and from a larger, more diverse sample.
- Explore Non-linear Relationships: Use advanced models like neural networks to uncover complex patterns.
- Investigate Seasonal Trends: Examine how time of year affects music listening and activity habits.