Spotify Playback Trends Analysis During Exam Periods

Prepared by: Selen Özkalkan
Date: January 2025
Course: DSA210 Term Project

Purpose of the Project

Music plays a significant role in regulating emotions, enhancing focus, and relieving stress. Exam periods often induce heightened stress levels, potentially altering listening habits. This project investigates how Spotify listening patterns change during exam periods compared to non-exam periods.

The goal was to analyze personal Spotify playback data from 2024, assess daily playback trends, and determine whether there is a significant increase in listening activity during exam periods.

Objectives

The primary objectives of the project were:

1. **Pattern Identification:** Analyze and compare daily playback trends for exam and non-exam periods.
2. **Hypothesis Testing:** Statistically test whether daily playback increases during exam periods.
3. **Behavioral Insights:** Provide insights into how external stress-inducing events (e.g., exams) influence music-listening habits.

Hypothesis

The study was guided by the following hypotheses:

- **Null Hypothesis (H_0):** Daily playback during exam periods is less than or equal to playback during non-exam periods.
- **Alternative Hypothesis (H_1):** Daily playback during exam periods is greater than playback during non-exam periods.

A one-sided statistical analysis was performed to test these hypotheses.

Data Collection

1. Spotify Playback Data

- Personal playback history was extracted in JSON format using Spotify's data export feature.
- The dataset contained metadata such as:
    - Playback timestamps (`ts`).
    - Playback duration in milliseconds (`ms_played`).
    - Track and album information.
- The dataset covered playback history for both 2023 and 2024. However, only 2024 data was analyzed to align with the available exam schedule.

2. Exam Period Data

- Exam periods were provided in a structured table format, including:
    - Start and end dates for each exam period in 2024.
    - Descriptive labels for each period (e.g., "Exam Prep 1").

Data Processing and Methodology

The analysis was conducted in Python using libraries such as `pandas`, `plotly`, and `scipy`. The workflow consisted of the following steps:

1. Data Preprocessing

- **Playback Data Filtering:**
    - Playback data was converted to daily aggregated trends by summing playback duration (in minutes) for each day.
    - 2023 data was excluded to focus solely on 2024 playback trends.
- **Missing Dates:**
    - Missing dates in the playback data were filled with zeros to maintain temporal continuity.
- **Exam Period Categorization:**
    - Exam periods were defined using their start and end dates.
    - Dates within these ranges were labeled as "Exam Period" days, while all other dates were labeled as "Non-Exam Period."

2. Visualization

A time-series chart was created to display daily playback trends for 2024. Exam periods were highlighted with red-shaded regions for visual clarity. This provided an intuitive representation of playback trends over time.

3. Statistical Analysis

- **Data Segmentation:**
  - Playback data was divided into two groups:
    - Daily playback during exam periods.
    - Daily playback during non-exam periods.
- **Hypothesis Testing:**
  - A **Welch's t-test** was performed to compare the two groups, as it accommodates unequal variances.
  - A **Mann-Whitney U test** was also conducted for robustness.

Findings

Visualization

The time-series chart revealed clear spikes in playback activity during exam periods. Exam periods consistently showed higher daily playback durations compared to non-exam periods.

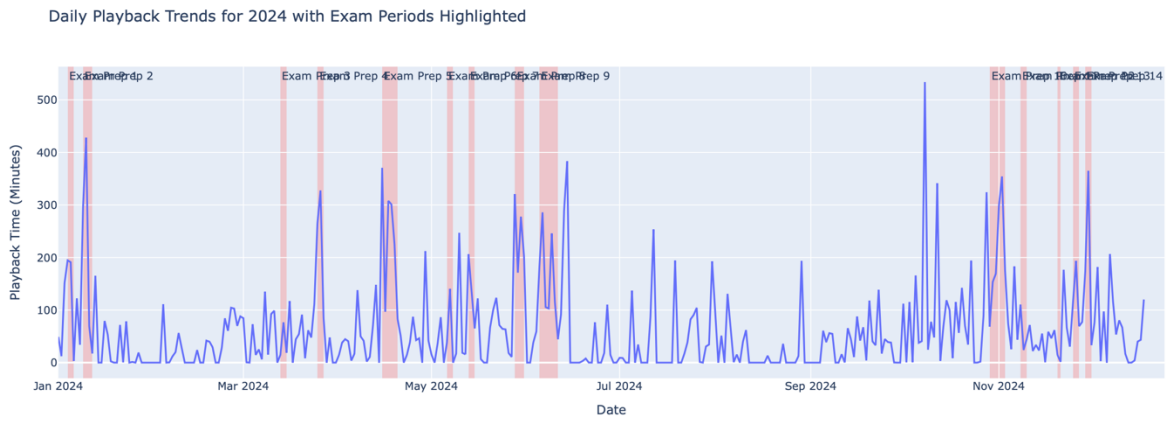

Table 1

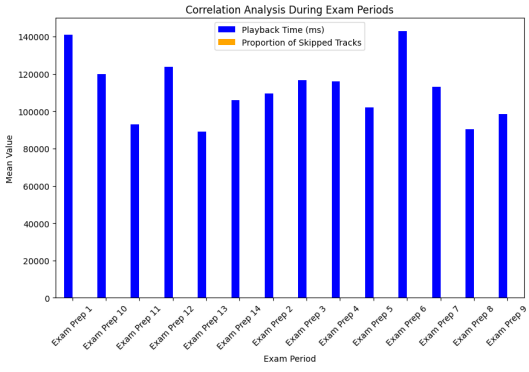

Table 2

Statistical Results

1. **Descriptive Statistics:**
    - **Exam Periods:**
        - Mean playback: **158.12 minutes/day**.
        - Standard deviation: **115.42 minutes**.
    - **Non-Exam Periods:**
        - Mean playback: **47.42 minutes/day**.
        - Standard deviation: **67.59 minutes**.
2. **Hypothesis Test Results:**
    - **Welch's t-test:**
        - t-statistic: **6.89**
        - p-value: **< 0.0001**
        - Result: The null hypothesis was rejected, indicating significantly higher playback during exam periods.
    - **Mann-Whitney U Test:**
        - p-value: **< 0.0001**
        - Confirmed the robustness of the t-test result.

Conclusions

The analysis strongly supports the hypothesis that daily playback increases significantly during exam periods. This aligns with the assumption that music is often used for focus, stress relief, or mood regulation during high-pressure times in my personal life.

Limitations

1. **Scope of Data:**
    - The analysis was limited to personal playback history, which may not generalize to broader populations.
2. **External Factors:**
    - The study did not account for additional variables such as study habits, mood, or changes in daily schedules.

Future Work

1. **Expand the Dataset:**
    - Analyze playback data from multiple users to validate findings on a larger scale.
2. **Genre and Mood Analysis:**
    - Investigate whether specific genres or types of music are more prevalent during exam periods.

3. **Sentiment Correlation:**

- Use sentiment analysis to explore links between playback patterns and emotional states during exams.

Acknowledgments

This project was completed as part of the DSA210 Term Project. The data and insights derived are solely based on personal playback history, and no external datasets were used.