

Introduction

This project, part of the DSA210 course at Sabancı University, observe the relationship by analyzing one year of personal Spotify listening data alongside one year of sleep records. I aim to uncover meaningful insights into how these two aspects of my daily life interact using data analysis and machine learning techniques.

Sleep Data

- **Timeframe:** Nightly records across one year.
- **Key Attributes:**
 - Total sleep duration (in hours)
 - Sleep start and end times

Spotify Data

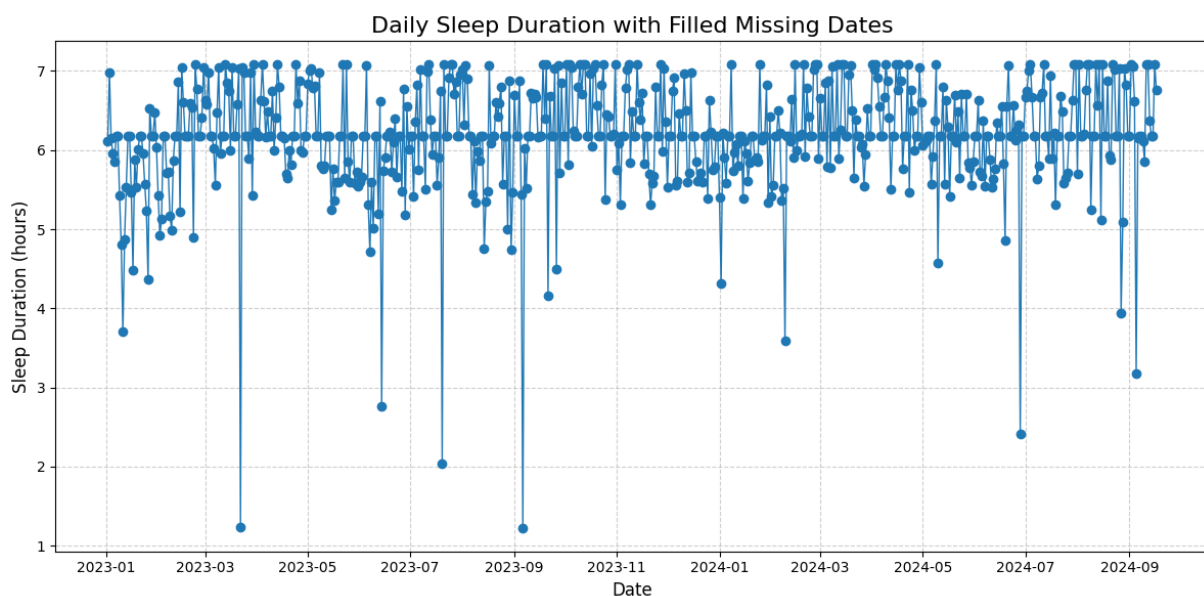
- **Timeframe:** A full year of data.
- **Key Attributes:**
 - Listening duration
 - Timestamps of playback

Sleep Data Analysis

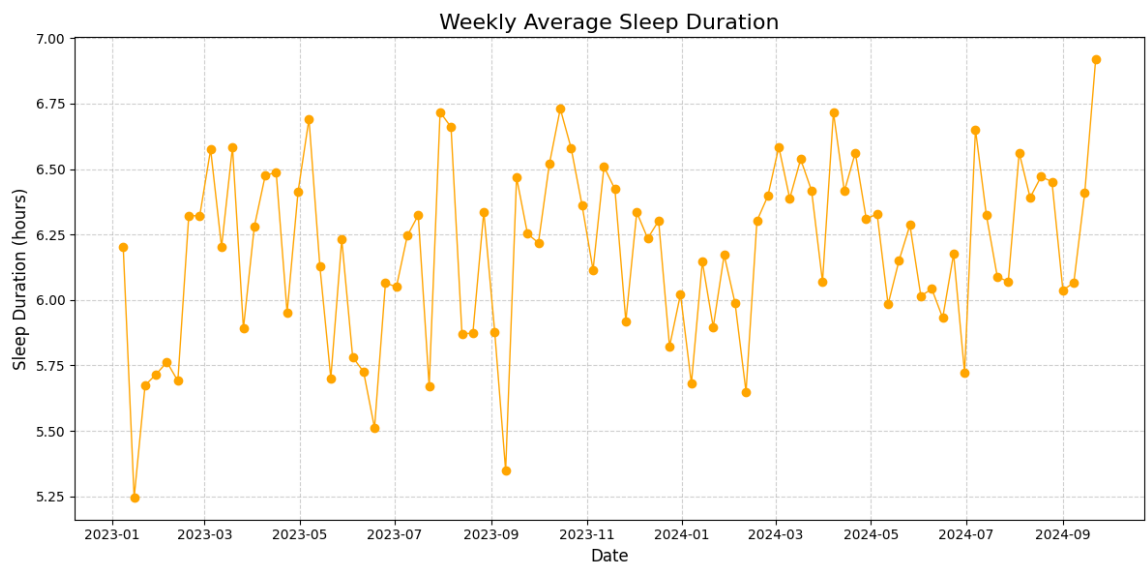
The sleep data represents my nightly records collected over the course of a year, providing detailed insights into my sleep patterns. It includes key attributes such as the total duration of sleep in hours, as well as the specific start and end times for each sleep session. This data offers a comprehensive view of my nightly sleep habits and trends over the year.

Daily Sleep Duration

The graph illustrates daily sleep duration from January 2023 to September 2024, showing an average sleep duration of 6–7 hours. However, notable fluctuations are present, with several dips below 5 hours and extreme outliers as low as 1–3 hours.



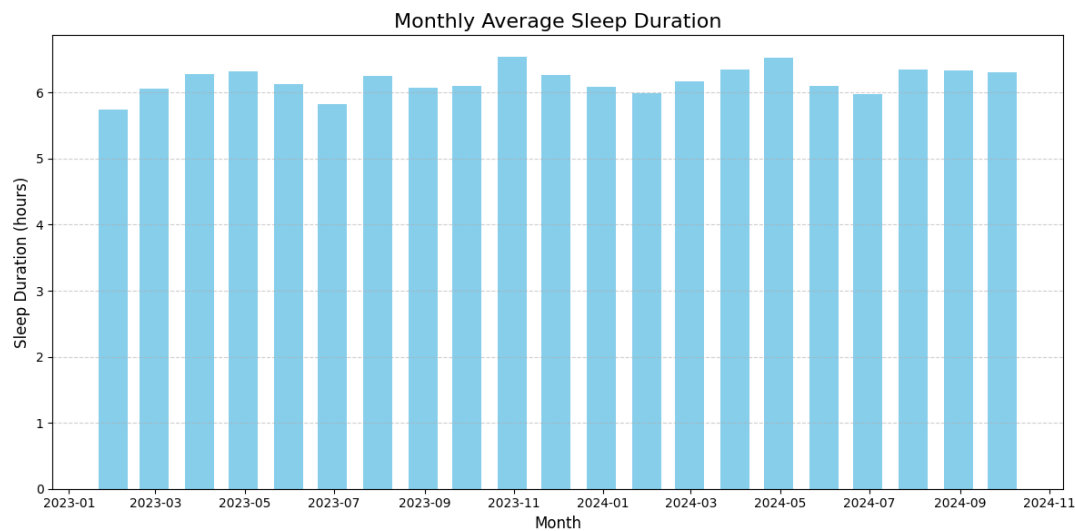
Weekly Average Sleep Duration



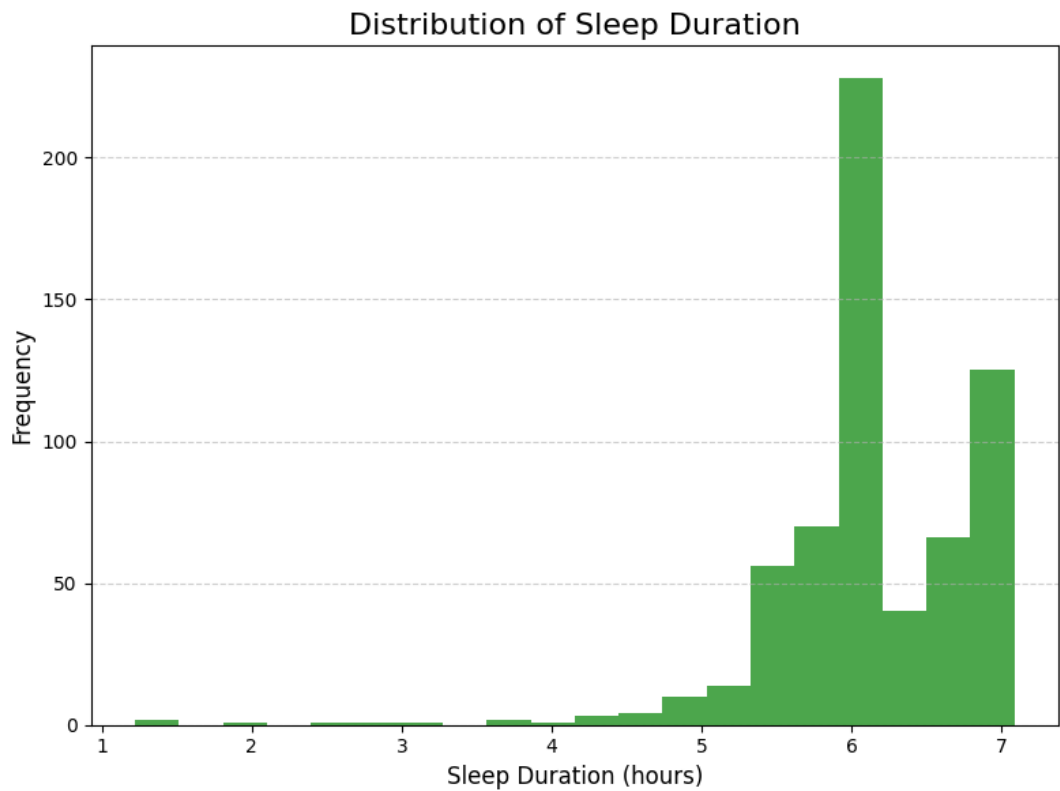
The graph illustrates changes in sleep patterns over time, showing weekly averages of sleep duration. It reveals noticeable ups and downs, with some weeks reflecting longer sleep while others show shorter durations. There are periods where sleep appears more stable, while other times experience fluctuations. Overall, the graph highlights the variability in sleep patterns over the analyzed period.

Monthly Average Sleep Duration

The graph is a bar chart representing the monthly average sleep duration in hours. Each bar corresponds to a specific month, starting from January 2023 and extending through late 2024. The height of each bar indicates the average sleep duration for that month. The chart uses light blue bars for visual clarity and is labeled with "Month" on the x-axis and "Sleep Duration (hours)" on the y-axis. The chart's title, "Monthly Average Sleep Duration," clearly conveys its purpose, showing a steady trend with slight variations in monthly sleep patterns.



Distribution of Sleep Duration

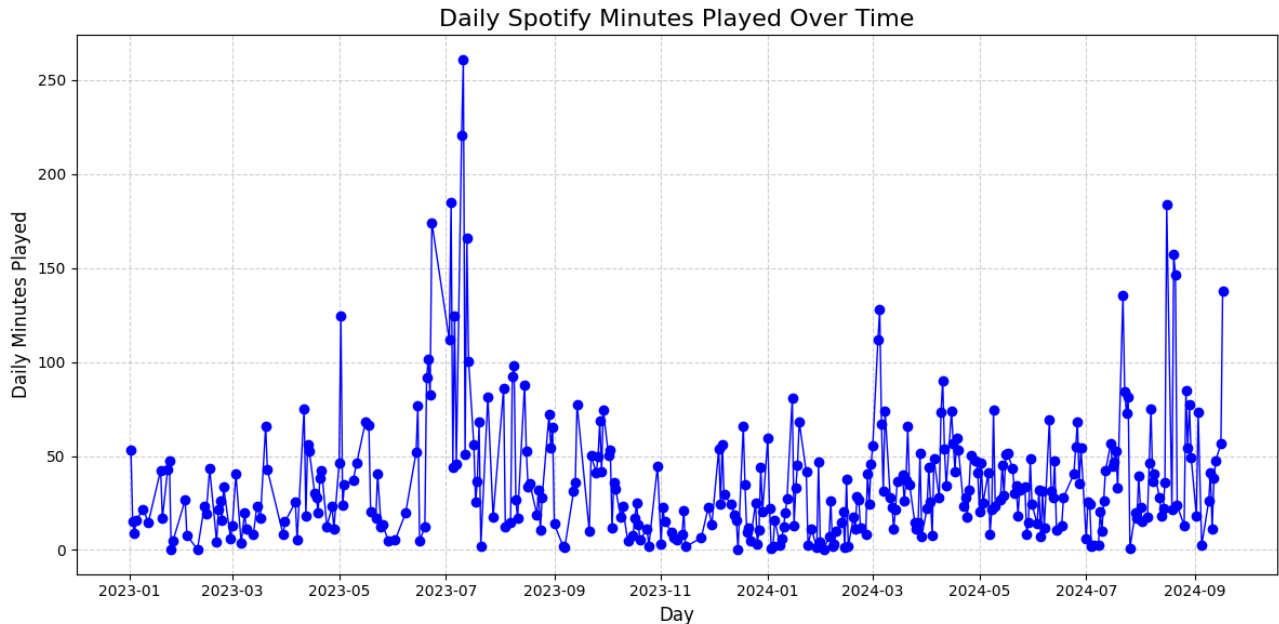


The graph is a histogram showing the distribution of sleep duration in hours. It represents how frequently certain sleep durations occur within the dataset. Most sleep durations are concentrated around higher values, indicating that longer sleep durations are more common. There is a increase in frequency as the sleep duration grows, with a noticeable peak at the higher end.

Spotify Data

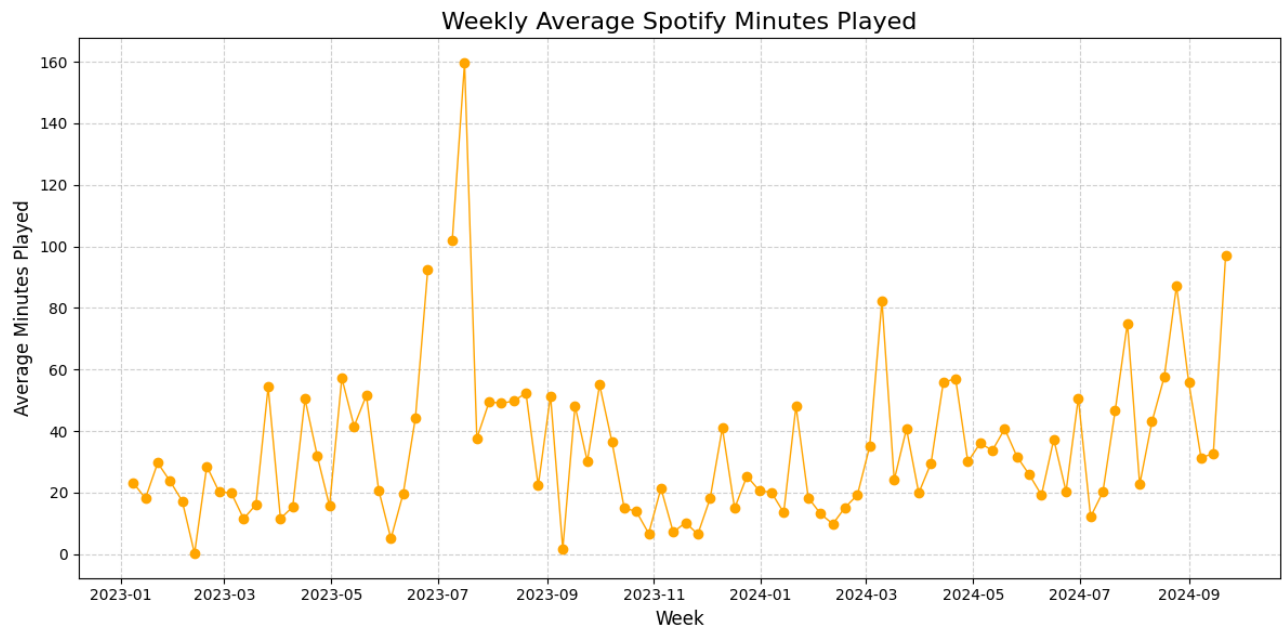
The Spotify data encompasses a full year of listening history the duration of each listening session, and precise timestamps of playback.

Daily Spotify Minutes Played Over Time



The graph is a line chart showing daily Spotify listening time in minutes over a year. The x-axis represents the days, while the y-axis shows the number of minutes listened each day. Data points are connected with a blue line, making fluctuations in listening time clearly visible. Peaks and dips are evident, with a prominent spike mid-year, indicating a significant increase in listening activity. The chart effectively visualizes the variability in daily listening habits over the analyzed timeframe.

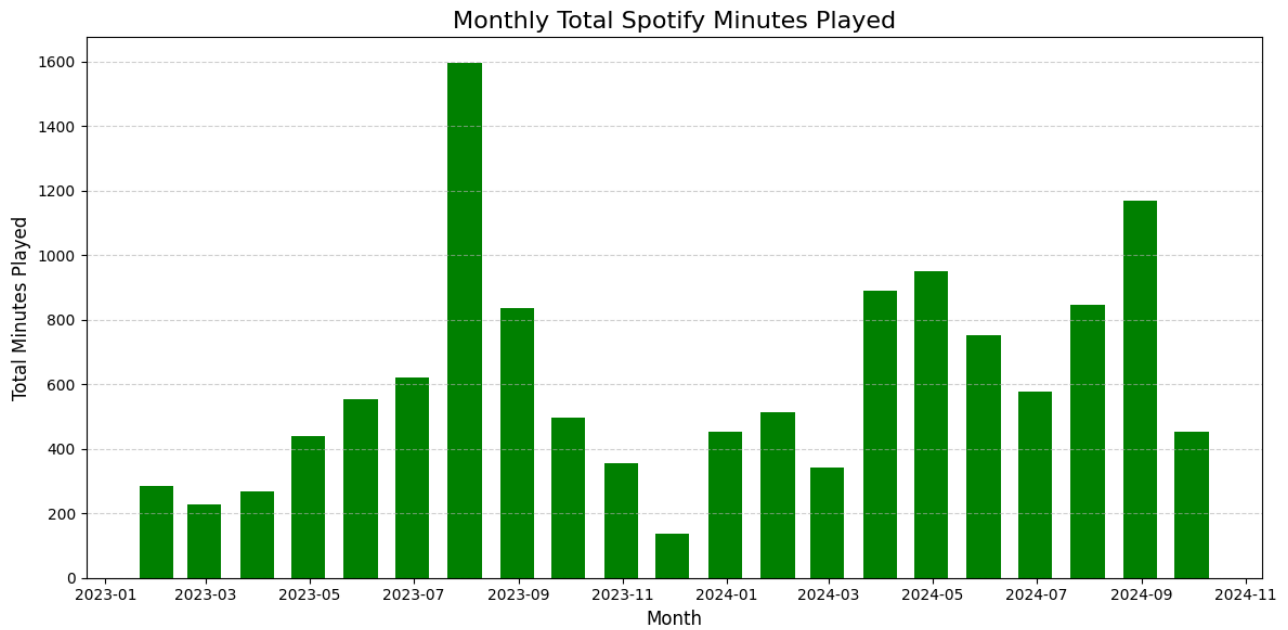
Weekly Average Spotify Minutes Played



The graph is a line chart showing the weekly average Spotify listening time, measured in minutes, over a specified period. The x-axis represents the weeks, while the y-axis displays the average listening time in minutes. The chart reveals fluctuations in listening with some weeks showing consistent activity and others experiencing spikes. A significant peak is noticeable in the middle of the timeframe, followed by a decline and more variable activity. Towards the end, there is a gradual increase in listening time. The chart effectively highlights changes and trends in weekly Spotify usage.

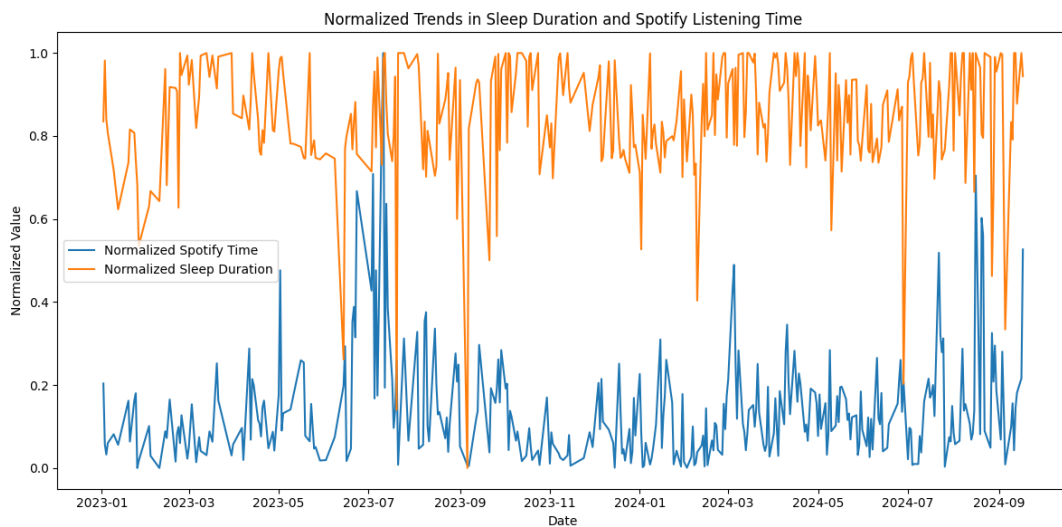
Monthly Total Spotify Minutes Played

The bar chart shows the total Spotify listening time per month over a given period. The x-axis represents months, while the y-axis shows total minutes played. A significant peak appears mid-year, followed by a decline and smaller fluctuations. This indicates variability in monthly listening habits, with some months showing much higher engagement than others.



Correlation Analysis

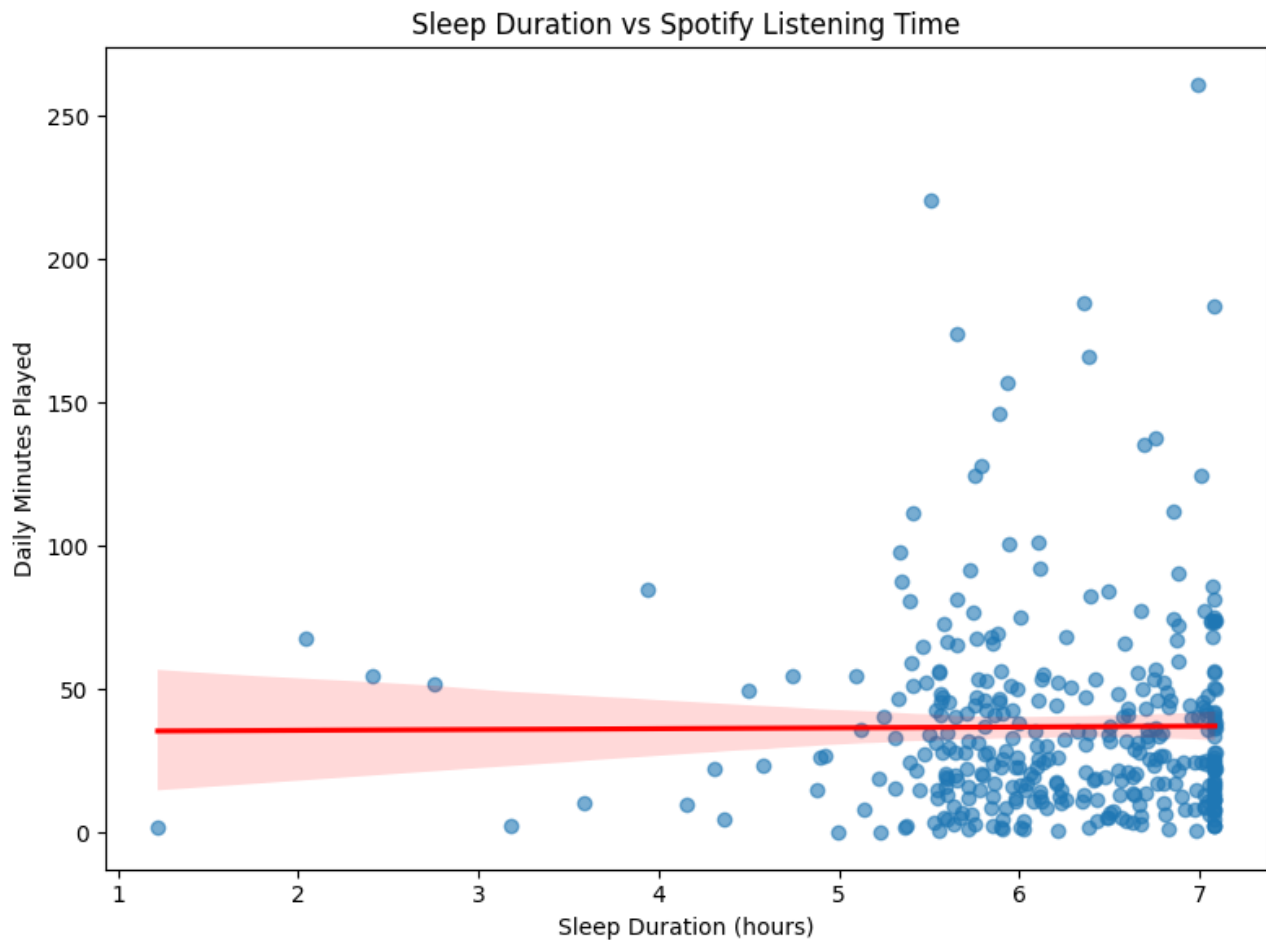
Normalized Trends in Sleep Duration and Spotify Listening Time



The graph shows normalized Spotify listening time (blue) and sleep duration (orange) over time. The x-axis is the date, and the y-axis shows normalized values from 0 to 1. The orange line is steadier, while the blue line has more spikes and drops.

Sleep duration vs Spotify Listening Time

The scatterplot shows daily sleep duration on the x-axis and Spotify listening time on the y-axis, with data points represented as blue dots.



Hypothesis Testing

Null Hypothesis (H_0): There is no significant relationship between sleep duration and Spotify listening time.

Alternative Hypothesis (H_1): There is a significant relationship between sleep duration and Spotify listening time.

Pearson Correlation

- Correlation Coefficient: 0.0068
- This indicates an extremely weak positive relationship between Sleep Duration (hours) and Daily Minutes Played.
- A value close to zero suggests little to no linear relationship between the two variables.
- P-Value: 0.8993
- Since the p-value is much greater than 0.05, the correlation is not statistically significant.

- Conclusion: Sleep duration and Spotify listening time do not have a meaningful linear relationship.

T-Test

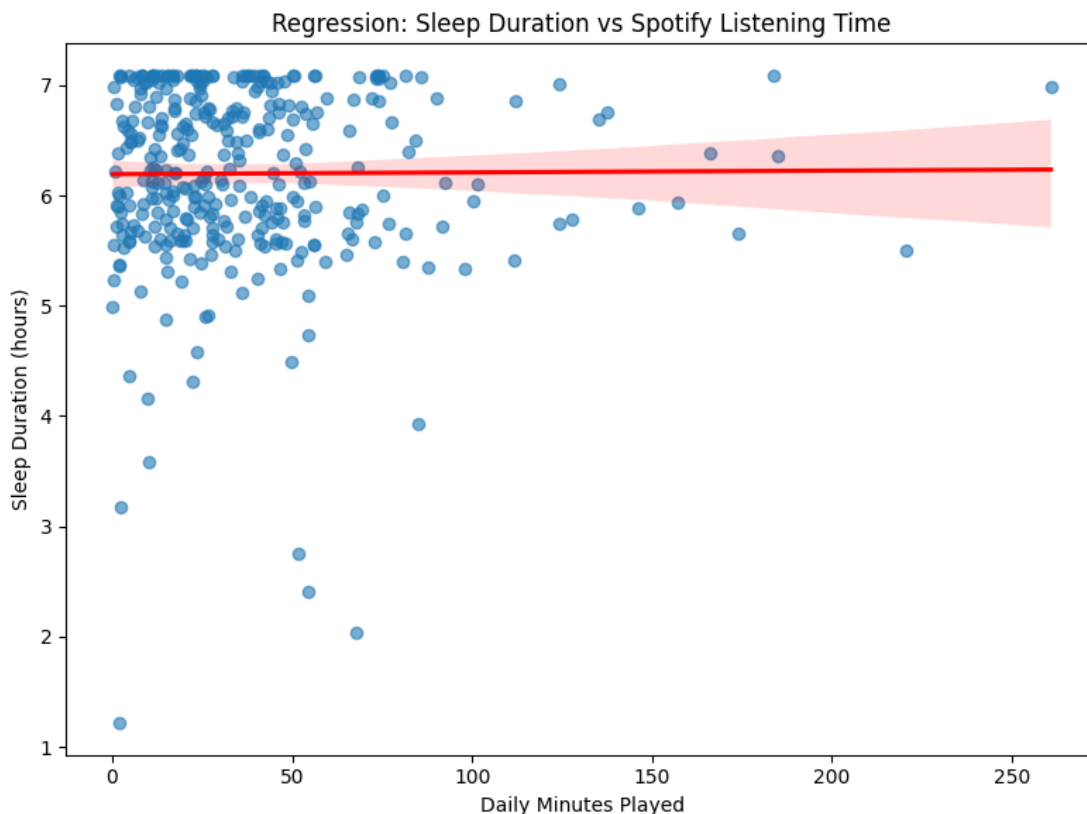
- T-Statistic: 0.2015
- A small t-statistic indicates minimal difference between the means of Spotify listening time for below-average and above-average sleep groups.
- P-Value: 0.8404
- Since the p-value is much greater than 0.05, the difference in Spotify listening time between the two sleep categories is not statistically significant.
- Conclusion: The amount of Spotify listening time does not significantly differ based on whether a person had below-average or above-average sleep.

Chi-Square Test

- High Spotify listening is almost equally distributed across sleep categories:
 - Above Avg Sleep: 88 (High), 93 (Low)
 - Below Avg Sleep: 85 (High), 80 (Low)
- Chi-Square Statistic: 0.1854
- This small value indicates that the observed counts are very close to the expected counts under the null hypothesis.
- P-Value: 0.6668
- Since the p-value is greater than 0.05, there is no statistically significant association between sleep categories and Spotify listening categories.
- Conclusion: Sleep categories (below/above average) and Spotify listening categories (high/low) are not significantly associated.

Regression: Sleep Duration vs Spotify Listening Time

Scatter Plot with Regression Line



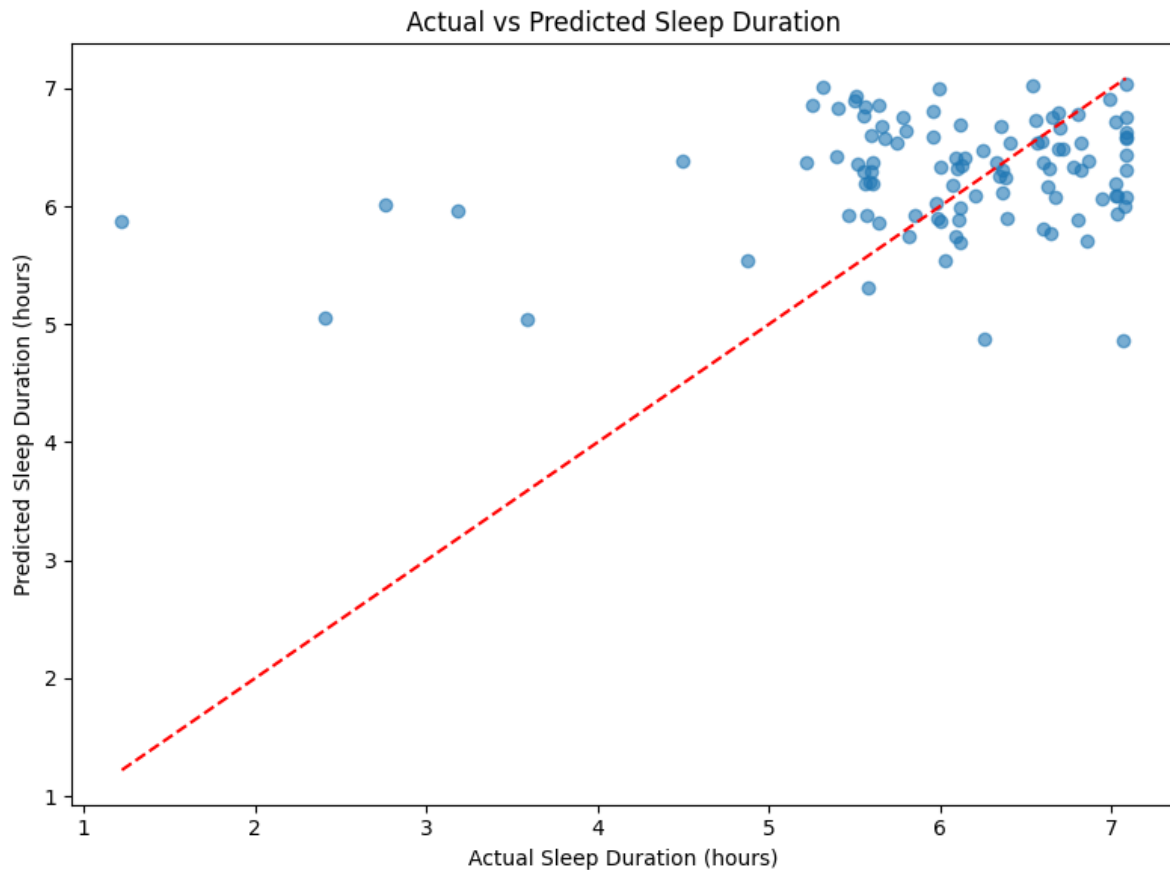
- The scatter plot shows a random spread of points with no clear trend or pattern.
- The regression line is almost flat, confirming the weak and insignificant relationship between Spotify listening time and sleep duration.
- The confidence interval (shaded area) around the regression line is wide, indicating a lack of confidence in the fit.

Conclusions

- **Relationship:**
There is no meaningful relationship between Spotify listening time and sleep duration in this dataset.
- **Predictive Power:**
Spotify listening time is not a useful predictor of sleep duration, as evidenced by the near-zero R-squared value and insignificant p-value.
- **Practical Insight:**
Music listening time does not appear to influence sleep behavior in a measurable way. Other factors might be more important in determining sleep duration.

Actual vs Predicted Sleep Duration

- The scatter plot shows:
 - Points are widely scattered around the diagonal red line, indicating poor agreement between the actual and predicted values.



- Many predicted values deviate significantly from the actual sleep durations, further confirming poor model performance.
- The random forest model fails to capture meaningful patterns between Spotify listening time and sleep duration.

Conclusion

- Relationship:
 - The Random Forest model confirms the results of the regression and statistical tests: there is no meaningful relationship between Daily Minutes Played and Sleep Duration.
- Predictive Power:
 - The feature Daily Minutes Played is not useful for predicting sleep duration, as evidenced by the poor MSE and negative R-squared.

Conclusion

The analysis indicates that there is no meaningful relationship between Spotify listening time and sleep duration. The Pearson correlation shows an extremely weak positive relationship,

with a near-zero correlation coefficient and a p-value far above the threshold for statistical significance. T-tests also reveal no significant difference in Spotify listening time between individuals with below-average and above-average sleep durations. Similarly, the chi-square test confirms that sleep categories (below or above average) and Spotify listening categories (high or low) are not significantly associated.

Regression analysis further supports these findings, showing a flat regression line and a near-zero R-squared value, indicating no predictive power of Spotify listening time for sleep duration. The Random Forest model also fails to find meaningful patterns, as shown by poor performance metrics such as a negative R-squared value and high MSE. Overall, the results suggest that Spotify listening time does not influence or predict sleep behavior, and other factors may be more relevant in determining sleep duration.