Project 1: Inferring Sleep Patterns from Browser History Data

Author: Bianca Linares

Date: February 24, 2025

Course: CS 4501/6501

GitHub Repository: https://github.com/blinares-cs/Project_1_Sleep_Analysis/tree/main

Introduction

Sleep plays a critical role in maintaining overall well-being, but maintaining a consistent sleep schedule can be challenging, particularly during academically demanding times. Like many students, I have experienced fluctuations in my sleep patterns throughout the semester, which prompted me to determine whether browser history data can be an indicator of sleep behavior. The primary objective of this project is to understand and estimate sleep start and wake times using NirSoft's BrowserHistoryView tool, which extracts browsing activity from Chrome, Edge, Firefox, and other browsers. By analyzing these patterns, I aim to determine how accurately sleep duration can be inferred from browser history alone. Additionally, I examine how sleep patterns differentiate before and during the semester and whether certain days of the week exhibit significant deviations in sleep quality. This analysis covers a timeframe from December 1, 2024, to February 23, 2025, that includes both the academic break and the current semester. Using Python for data processing and visualization, browsing timestamps were analyzed to infer sleep trends. Various statistical techniques, such as calculating average sleep duration, most common sleep times, variability in sleep habits, and weekly trends, were applied to assess sleep behaviors. This project evaluates the possibility of using browsing data for sleep tracking and understanding the impact of an academic semester on sleep consistency.

Data Collection and Processing

The dataset used in this analysis was extracted using NirSoft's BrowserHistoryView tool, which compiles browsing history from multiple web browsers, including Google Chrome, Microsoft Edge, and Mozilla Firefox. The dataset includes timestamps of website visits and visit durations, allowing for an approximation of daily activity windows. The extracted dataset required several preprocessing steps before the analysis. The "Visit Time" column, which recorded the timestamp of each website visit, was converted into a structured datetime format to allow for time-based calculations. Similarly, the "Visit Duration" column, which represents how long a webpage remained open, was converted into a timedelta format. Any missing values in the "Visit Duration" column were filled with zero seconds to prevent errors in later calculations. After ensuring that the timestamps were properly formatted, the dataset was sorted chronologically to ensure that all visit records followed a consistent timeline.

Once the data was structured correctly, a new data frame was created, extracting only the "Visit Time", which included the date, and the "Visit Duration" column and computing the end time of browsing sessions by adding the visit duration to the visit timestamp. This process allowed for more accurate tracking of when browsing activity stopped each day. From this dataset, the first visit of each day and the last visit before midnight was collected. Based on these values, a function was created to determine sleep duration more precisely. If the last recorded visit of the day occurred before midnight, it was considered the approximate bedtime. The wake-up time was determined based on the first visit recorded the following morning. Since some days I have worked past midnight, the last visit, once the duration was added, of the previous day that extended to the next day was also considered when determining sleep. This approach helped to reduce errors in cases where work extended into the early hours of the following day. Additional filtering was also applied to remove extreme outliers that could distort the results. Specifically, sleep durations exceeding 16 hours were removed, as it is highly unlikely that I would consistently sleep for such long periods. However, unlike many sleep studies that impose a minimum sleep duration threshold, I chose not to filter out very short sleep durations. I made this decision because I have occasionally not slept at all before my classes or for my job, and I wanted to ensure that nights with little to no sleep were accurately reflected in the analysis. Following the processing of the data, 42 sleep durations were included in the analysis.

During analysis, the dataset was also divided into data before the semester (December 1, 2024 - January 12, 2025), which included 19 sleep durations, and during the semester (January 13 - February 23, 2025), which included 23 sleep durations. I did this because I hypothesized that the days before the semester represented times when academic stress was likely lower, allowing for a more regular sleep schedule. The semester time has other factors involved such as deadlines, exams, and unpredictable workloads that may contribute to shorter or more inconsistent sleep durations. By splitting the data into these two time frames, the analysis could compare how academics influenced my sleep schedule and whether a measurable change in sleep duration occurred between these two time frames.

Analysis Methodology

Once the dataset was structured and the durations of sleep were calculated, several other calculations and visualizations were done. These methods evaluated overall sleep duration, variations across different days, and differences between the times before the semester and times during the semester.

The first step was creating a line graph of sleep duration over time to visualize sleep trends. A dashed horizontal line at seven hours was included to represent the amount of sleep that I desire to have each day, making it easy to see when sleep fell below or exceeded this benchmark. This graph provided a clear view of irregular sleep durations and possible patterns of sleep deprivation.

To understand how sleep durations were distributed, a bar graph categorizing sleep durations into different ranges was generated. Durations were grouped into five categories: less than four hours, four to six hours, six to eight hours, eight to ten hours, and more than ten hours. The number of durations that fell into each category was counted and displayed in the bar chart. This visualization helped assess whether shorter or longer sleep durations were more common than the desired amount of sleep.

Another visualization created was a bar graph showing the average sleep duration per day of the week. This involved extracting the day of the week for each sleep entry, computing the mean sleep duration for each day, and mapping it to its corresponding weekday name. This graph highlighted whether sleep durations were shorter on weekdays and longer on weekends.

To determine how the semester affected sleep, the average sleep duration for the semester period from January 13 to February 23, 2025, was calculated. The same process was applied to compute the average sleep duration for the dates before the semester from December 1, 2024, to January 12, 2025, allowing for a direct comparison. The difference between the two averages helped determine whether sleep duration decreased once the semester started.

Another part of the analysis involved calculating the average sleep start and wake-up times. This was done by extracting the hour and minute components from the last recorded visit, which was considered the sleep start, and the first recorded visit, which was considered the wake-up time. These values were averaged to estimate when sleep typically started and ended each night.

To measure how consistent sleep schedules were, the standard deviation of sleep start and wake-up times was computed. A higher standard deviation meant that sleep schedules were more irregular, while a lower value indicated more stability. This analysis helped determine whether bedtime or wake-up times were more unpredictable throughout the semester.

These calculations and visualizations provided a detailed analysis of how sleep duration fluctuates, how sleep patterns differ across weekdays and weekends, and how the semester affects sleep consistency. The insights gained from this analysis helped evaluate whether browser history could be a reliable way to approximate sleep behaviors and track trends over time.

Results

The analysis of sleep patterns from browsing history uncovered notable fluctuations in sleep duration. Comparing trends before and during the semester, along with variations across different days, provided a deeper look into both expected and unexpected sleep behaviors.

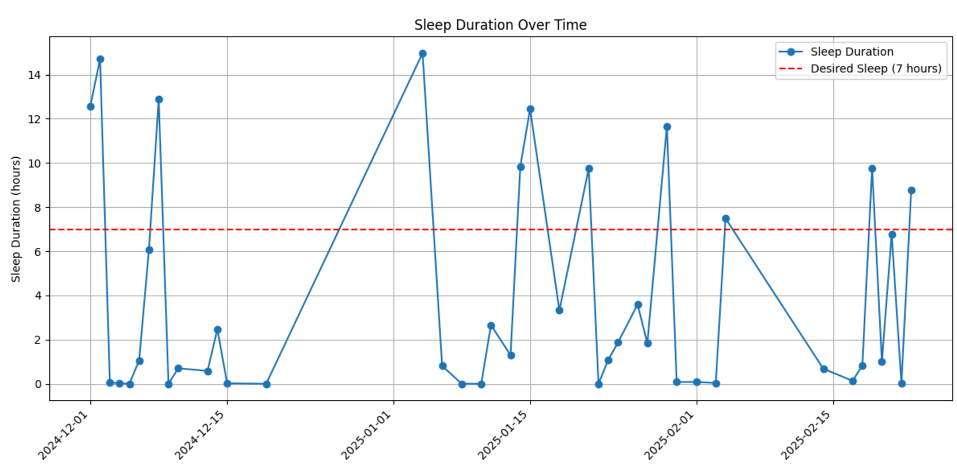

Figure 1: Line graph of Sleep Duration over Time

Firstly, the line graph of sleep duration over time showed significant fluctuations in sleep duration, with multiple days of extremely short sleep, less than 1 hour, and occasional long sleep exceeding 10 hours. This suggests inconsistent sleep habits, possibly driven by workload, deadlines, or recovery from sleep deprivation, especially during the semester. The horizontal red line at 7 hours represents the recommended sleep duration, and most of the recorded sleep sessions fall below this threshold. There appears to be more extreme fluctuations during the semester, whereas dates before the semester durations, while still irregular, are somewhat more stable. This could indicate that academic stress and responsibilities led to different sleep patterns. There are also gaps in the data where no sleep duration is recorded, which was due to data reaching extremely long durations when sleep could not be accurately calculated.

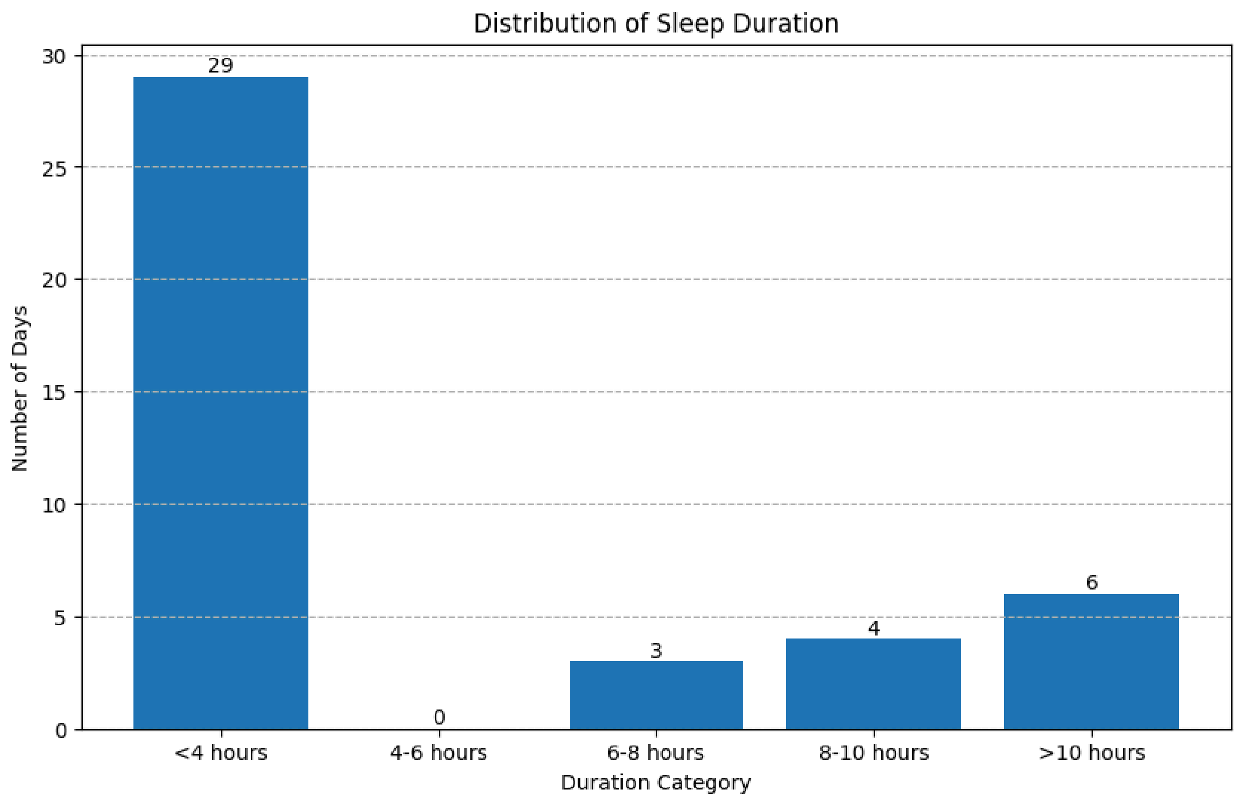

Figure 2: Bar chart of Distribution of Sleep Duration

The bar chart categorizing sleep duration into different time frames showed a significant pattern of extremely short sleep periods, with 29 nights recorded at under 4 hours of sleep. In contrast, only 3 nights fell within the 6-8 hour range, which is typically recommended for healthy sleep. Additionally, 4 nights had sleep durations between 8-10 hours, indicating occasional longer rested days. Sleep durations exceeding 10 hours were recorded on 6 nights, showing instances of recovery sleep, possibly after prolonged sleep deprivation. One of the most surprising results is that there are no sleep durations between 4-6 hours. This indicates that on nights when sleep was short, it was often very short, rather than falling into a moderate sleep reduction.

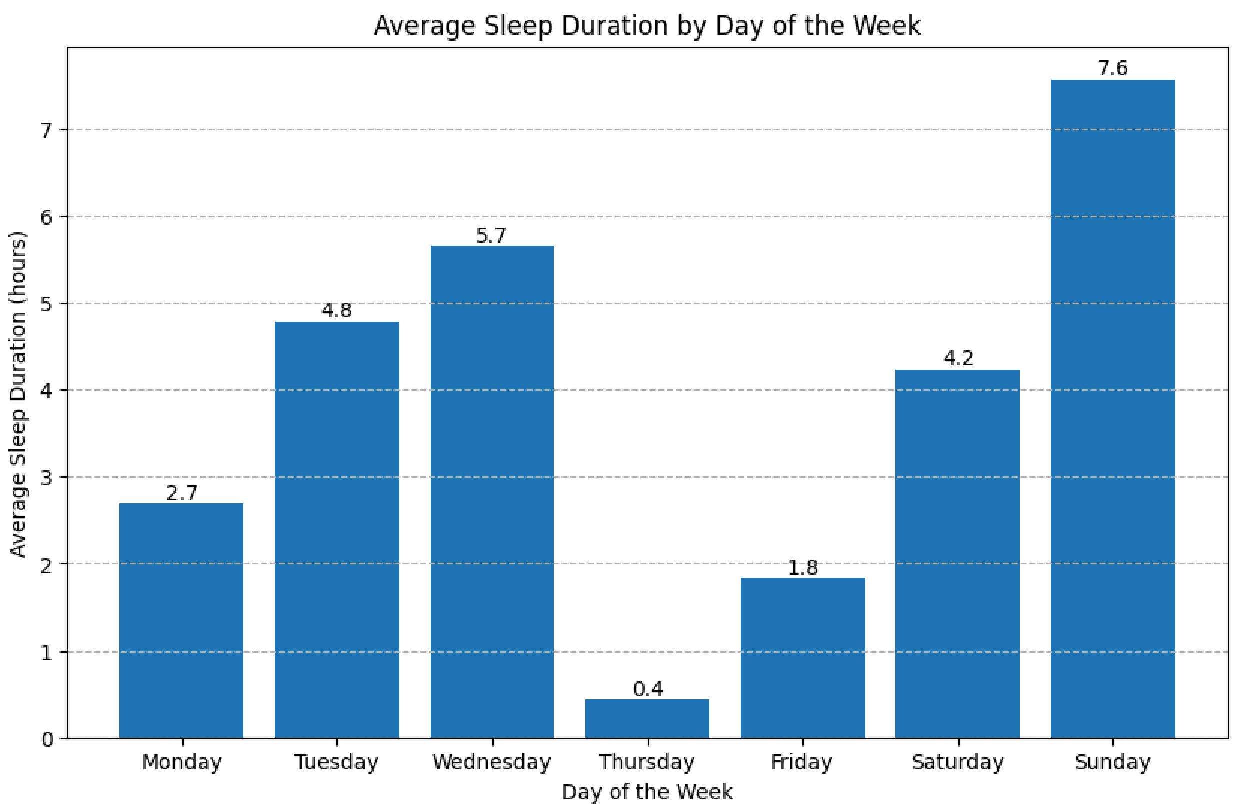

Figure 3: Bar chart of Average Sleep Duration by Day of the Week

The average sleep duration by day of the week showed inconsistency of sleep across different days. Sundays showed the highest average sleep duration at 7.6 hours, while Thursdays had the lowest at only 0.4 hours. Sleep was generally more consistent and slightly higher on weekends compared to weekdays. The results during the week appear to be consistent with my current schedule of classes and expected deadlines before each class.

The average sleep duration during the semester was 4.02 hours, while the days prior to the semester had a slightly lower average of 3.67 hours. This small difference implies that academic responsibilities may not have significantly altered overall sleep duration. These findings do not align with my expected sleep patterns between days during the breaks of the academic year and times during the semester. The estimated average sleep start time was about 12:00 AM, which aligns well with my current sleep schedule, and an average wake time was found to be about 9:00 AM. This wake time is later than expected, possibly due to irregular sleep habits shown in the data. The standard deviation of sleep start times was 0.64 hours, indicating relatively stable bedtime habits. However, the standard deviation of wake times was higher at 1.52 hours, showing greater inconsistency in morning routines. These results indicate that while bedtimes remain fairly predictable, wake-up times fluctuate more, which I believe correspond with my current sleep schedule. These results highlight the significant inconsistency in my sleep habits, and the need to find a more consistent sleep schedule.

Limitations and Discussion

One of the limitations of this project is the presence of extended online activity periods that spanned multiple days, suggesting that my computer may have been left on passively rather than reflecting actual usage. This could have resulted from leaving tabs open, watching videos overnight, or simply not shutting down my computer before going to sleep. As a result, some estimated sleep start and wake times may not accurately represent when I actually went to bed or woke up. Another challenge in estimating sleep patterns is the use of multiple devices. While this analysis relies on browsing activity from my primary computer, it does not account for phone or tablet usage before sleep or immediately after waking up. This means there could be instances where my online activity continued on another device after closing my computer, or I may have woken up and used a phone before resuming browsing on my computer. These gaps lead to uncertainty in determining the true times I went to sleep and woke up.

Additionally, this project lacks a reliable ground truth for validating the accuracy of the sleep estimates. Without a recorded sleep log or data from a dedicated sleep tracking device, it is difficult to quantify how well the inferred sleep patterns align with my actual sleep behavior. While the trends may provide insight into overall patterns, there is still the possibility of miscalculations. There are also potential inaccuracies in how sleep durations were calculated. While the analysis does calculate the duration with the gap between the last recorded activity of one day, even with the duration moving to the next day, and the first recorded activity of the next day representing the wake time, this may not always be the case. There may have been instances where I woke up briefly, worked for a short period, and then went back to sleep. This fragmented sleep pattern is not shown in the current methodology, meaning that some nights of interrupted sleep could have been incorrectly classified as a continuous work or sleep.

In addition to the limitations of the analysis, it was found that some of the findings align with my expectations of my sleep habits, but others do not. I initially anticipated a greater difference in sleep duration between weekdays and weekends, as I tend to have more flexibility on weekends to recover from lost sleep. However, the data did not show as much variation as expected. Additionally, the difference in sleep habits between the dates before the semester and semester itself was not greater than I had anticipated. The days of inactivity that were removed from the dataset were mostly found in the days before the semester, which may have played a role in inaccurately showing my sleep habits. If those days had been included, they might have provided a clearer picture of how my sleep patterns shifted from an academic break to the academic period. I also expected to see more instances of sleep durations between 4-6 hours, as this tends to be my typical sleep range. However, the data showed no durations in this range, with most nights falling either under 4 hours or well above 10 hours. While I do not doubt the frequent occurrence of nights with less than 4 hours of sleep, the presence of extremely short sleep durations, such as less than 1 hour, raises concerns about the accuracy of my sleep detection method. These instances suggest that another approach may be needed to determine sleep duration more reliably. On the other hand, I did expect the number of nights with more than 10 hours of sleep to be the second most common category after nights with less than 4 hours. This finding is understandable given that I often compensate for sleep deprivation with longer sleep sessions on certain days. The results generally confirm this expectation, but refining the way sleep is identified could provide a clearer distinction between normal short sleep, extreme sleep deprivation, and recovery sleep periods.

Future Work

To improve the accuracy of sleep inference, future work should incorporate additional data sources beyond browser history. Integrating YouTube watch history, other app usage, or wearable device data could refine sleep detection by showing screen time before bed and before waking up. Validating these estimates with a sleep tracker would help assess accuracy and refine detection criteria. Additionally, a deeper analysis of weekday versus weekend sleep variability could provide insight into how stress impacts sleep consistency. Future work should also account for fragmented sleep, as the current methodology does not capture multiple short sleep intervals.

Conclusion

This project was done to determine whether browser history could be used as an accurate way to infer sleep patterns. The findings highlighted frequent nights of extreme sleep deprivation, with an average semester sleep duration of only 4.02 hours, and notable fluctuations in wake-up times. While some results aligned with expectations, such as increased weekend sleep and recovery sleep, others, such as the lack of 4-6 hour sleep durations, raised questions about the methodology's accuracy. Limitations, including passive screen time, multiple device usage, and a lack of validation data, highlight the need for a more comprehensive approach. Future research incorporating additional data sources and wearable tracking could improve accuracy and provide deeper insights into sleep behaviors. Despite its limitations, this analysis demonstrates that browser history can offer a useful, though imperfect, method for tracking sleep trends and identifying patterns influenced by academic schedules.