Jalen Wang

GitHub Repository: https://github.com/wangjalen7/CS4501-Sleep-Project-1

Data Sources and Hypotheses

This report examines my browsing and YouTube watch activity over a one-month period (January 2025 to February 2025). The goal of this analysis is to infer my sleep patterns based on my online behavior using Chrome browsing history and YouTube watch history data extracted from Google Takeout. The Chrome Browsing History contains timestamps and URLs of visited websites, and YouTube Watch History includes timestamps of watched videos. The histories were parsed from JSON and HTML files respectively. The timestamps in the data were converted from UTC timestamps to human-readable datetime format and adjusted to Eastern Standard Time (EST) for consistency. The data was then grouped by the hour of the day to visualize activity patterns.

Before analyzing the data, I hypothesized that my inferred sleep schedule would be between 4 AM and 12 PM, which is when I usually tend to sleep each day. Third, I anticipated that sudden drops and increases in activity could correspond to transitions between online and offline activities, such as sleeping. Finally, I expected to see a noticeable gap in activity during my sleep hours.

Methods

The analysis was performed using Python and involved several techniques. Data cleaning and processing included extracting timestamps and converting them into readable datetime format, converting UTC timestamps to Eastern Standard Time (EST), and creating hourly and daily totals of activity counts. For visualization, histograms were plotted to show Chrome browsing and YouTube watch activity by hour, as well as a combined histogram to analyze overall online activity. These plots helped to reveal not only when activity was at its peak but also when significant decreases occurred, providing insight into daily routines and potential offline engagements. The histograms were designed with a bin size of one hour to ensure specific insight into activity trends throughout the day.

To further understand activity fluctuations, an activity change analysis was performed to identify the top three hours with the biggest increases and decreases in activity. This was done by calculating the hour-to-hour difference in browsing and watching activity, allowing for a clear identification of transitional periods such as waking up and going to sleep. These transitions were identified based on sharp increases or decreases in activity levels, which provided a strong indication of when offline activities (such as sleeping or being away from devices) occurred. This method proved useful in isolating specific timeframes when major changes in behavior took place.

Another critical analysis conducted was the identification of the 8-hour window with the lowest activity for each day of the week. This was achieved by iterating through the hourly activity data and summing the activity within a rolling 8-hour window for each day. The specific window of 8 hours was chosen to represent the usual amount of time I sleep per night. The window with the smallest total activity count was designated as the inferred sleep period. This approach took into account the possibility of variability in sleep schedules across different days, ensuring that each day was analyzed independently.

Jalen Wang

Results

The histograms reveal distinct patterns in my online activity. Chrome browsing peaks occur between 3 PM - 4 PM, 7 PM - 8 PM, and a smaller peak at midnight. This pattern generally aligns with the times that I am engaged in work or general web browsing, which typically takes place during the afternoon and into the evening. The peak at midnight suggests that I often continue browsing late at night, possibly for leisure, winding down before sleep, or staying up late to study. YouTube watching shows heavy activity peaking around 4 AM and again around 2 PM, with moderate activity throughout the rest of the day. This aligns with my tendency to watch YouTube as a way to relax before bed, as well as periodically throughout the day during breaks from work or other tasks. The combined activity histogram reveals a significant gap in activity between 4 AM and 12 PM, which strongly indicates my probable sleep hours, as I am not engaging in online activity during this time.

Furthermore, the top three biggest increases in activity were from 2 PM to 3 PM, where usage surged by 351 events (561 to 912) (events represent sum of activity like sites browsed and videos watched); from 11 AM to 12 PM, where usage rose by 264 events (374 to 663); and from 1 PM to 2 PM, where usage increased by 205 events (356 to 561). These increases suggest significant transitions from offline to online activities, likely corresponding with key daily moments. The 11 AM to 12 PM surge likely represents my wake-up period, during which I check my phone, browse updates, or watch YouTube. The 1 PM to 2 PM and 2 PM to 3 PM increases coincide with my shift into productivity mode after lunch, marking the time when I begin working or focusing on structured activities.

Conversely, the top three biggest drops in activity were observed from 8 PM to 9 PM, where usage decreased by 473 events (780 to 307); from 1 AM to 2 AM, where usage dropped by 198 events (481 to 283); and from 3 AM to 4 AM, where usage fell by 173 events (308 to 135). These decreases pinpoint key times when I disengage from online activity, either for sleep or offline responsibilities. The drop from 3 AM to 4 AM strongly suggests the time when I put my phone down and officially go to sleep. Similarly, the decline between 1 AM and 2 AM is another indicator of bedtime preparation, likely when I start reducing screen usage or sleep a bit earlier. The 8 PM to 9 PM drop aligns with my gym schedule, a period when I am typically engaged in offline physical activity, explaining the reduced online activity.

Finally, by identifying the 8-hour window with the least total online activity for each weekday, I estimated my sleep schedule with high accuracy. On Monday, Wednesday, Thursday, and Friday, my inferred sleep schedule was from 4 AM to 12 PM. On Tuesday and Saturday, the sleep window shifted slightly earlier, from 3 AM to 11 AM. On Sunday, it was from 5 AM to 1 PM, reflecting a slight deviation, likely due to weekend habits. This estimated sleep schedule aligns perfectly with my self-reported sleep pattern, further validating the accuracy of the inferred data. It also suggests that I maintain a consistent sleep routine, with only minor variations between weekdays and weekends. Unlike some individuals who may experience drastic shifts in sleep timing between workdays and weekends, my sleep schedule remains relatively stable, although not healthy.

Overall, the inferred sleep schedule is a strong match with my actual sleeping pattern (4 AM to 12 PM). This confirms that my browsing and YouTube history serve as reliable indicators for estimating my sleep schedule. The data suggests that sudden increases in online activity at certain times are due to transitions between offline and online activities. For example, low activity in the morning is a strong indicator that I am asleep, while specific periods of inactivity in the afternoon or evening likely correspond with offline engagements such as gym sessions or other responsibilities. The most significant drops in activity typically occur late at night or during the early morning hours, further reinforcing the notion that these timeframes represent my transition into and out of sleep. This analysis not only provides insights into my sleep schedule but also highlights how my online engagement aligns with my daily routines and offline commitments.

Insights

This analysis underscores the potential of online activity tracking as a valuable means for mapping daily behavioral patterns. Even without access to complex or detailed data sources, clear trends emerge, allowing for reasonable inferences about transitions between work, leisure, and rest. However, several improvements could enhance the depth of this analysis. First, incorporating data from mobile devices and additional applications such as TikTok, iMessage, or other frequently used platforms would offer a more holistic view of my digital habits. Given that these platforms constitute a significant portion of my screen time, their inclusion would likely refine sleep estimates and further differentiate my offline and online transitions. Additionally, expanding the timeframe beyond a single month would help distinguish between temporary behavioral fluctuations and more consistent, long-term habits. This could be particularly useful for detecting seasonal variations, such as increased activity during academic semesters versus breaks. Furthermore, a content-based analysis of browsing and media consumption patterns could provide insight into how my focus and interests shift throughout the day and how they might relate to my sleep patterns.

In conclusion, this analysis successfully identified clear patterns in my online activity and provided an accurate estimate of my sleep schedule. The results confirmed my self-reported sleep pattern and showcased how web activity data can serve as a useful tool for behavioral analysis. The study also demonstrates how digital footprints can provide insights into everyday habits and schedules, even without self-reported data.