Behavioral Patterns in Web Activity: A Time and Website Analysis

Rainnand Montaniel
College of Computing and Information Technology
National University
Manila, Philippines
montanielrp@students.national-u.edu.ph

***Abstract*—This study analyzes personal web browsing behavior using interval-based activity logging and statistical analysis. A custom-developed Google Chrome extension was used to record active tab usage at fixed 4-second intervals, generating a timestamped dataset categorized into social media, entertainment, developer, games, educational, and other website types. The dataset was examined in non-aggregated and aggregated forms, including daily, time-of-day, and weekly summaries. Descriptive statistics and visualizations were used to identify behavioral patterns, while inferential tests including one-way ANOVA, Kruskal–Wallis, and Dunn's post-hoc analysis with Bonferroni correction were conducted to determine statistically significant differences in browsing time across categories. Results indicate that browsing behavior varies significantly by website category, with social media consistently dominating total usage. Temporal analysis shows peak activity during noon, afternoon and evening hours. The findings demonstrate that digital behavior is strongly influenced by website type and time context, highlighting the value of data-driven self-analysis for improving awareness of personal screen-time habits.**

***Index Terms*—Web browsing behavior, Screen time analysis, ANOVA, Kruskal–Wallis test, Dunn post-hoc test, Temporal analysis, Self-tracking**

I. Introduction

The widespread use of digital platforms has made web browsing behavior an important indicator of productivity, attention allocation, and daily routine patterns. Individuals regularly divide their time among academic, professional, and entertainment websites, making browsing data a valuable source of behavioral insight. Analyzing how time is distributed across websites and temporal categories such as time of day and weekday versus weekend provides measurable evidence of digital habits and usage intensity.

Understanding browsing behavior is important because excessive or unbalanced digital consumption can affect productivity, focus, and overall time management. Prior research in behavioral analytics and human–computer interaction has shown that user engagement varies depending on content type and temporal rhythms influenced by circadian patterns. Studies suggest that platform type often determines duration of engagement, while time-related factors may influence consistency and frequency of use. However, many existing studies rely on large-scale commercial datasets or aggregated platform statistics. There is limited focus on structured, self-collected behavioral logs that allow precise, individualized analysis.

This project addresses that gap by examining a personally collected dataset of timestamped web activity recorded at fixed intervals. The raw logs were transformed into daily, time-of-day, and weekly aggregated measures to enable systematic statistical evaluation.

The objective of this study is to analyze personal web browsing behavior by identifying the most consumed website categories, determining peak periods of digital activity, and evaluating whether browsing time significantly differs across website categories under different temporal aggregations.

- **RQ1:** Which URL category accounts for the highest total website usage time in the collected dataset?
- **RQ2:** During which time-of-day periods does peak digital consumption occur?
- **RQ3:** Is there a statistically significant difference in total website usage time across URL categories under different temporal aggregations (daily, time-of-day, and weekly)?

Using visualization techniques and inferential statistical tests, the study evaluates patterns in personal web usage across URL categories and temporal groupings. By statistically examining behavioral patterns under different time aggregations, the research demonstrates how structured self-tracking data can provide meaningful insights into individual digital consumption patterns.

II. Literature Review

Previous studies have examined digital browsing behavior as a measurable indicator of productivity, attention allocation, and daily routine patterns. Research in human–computer interaction [1] investigated how frequent task switching during computer use affects cognitive performance and stress levels. Using automated activity logging and statistical modeling, their findings showed that interruptions and rapid transitions between digital tasks reduce sustained attention and efficiency. These results established web activity data as a reliable behavioral metric for evaluating productivity. These findings support the use of structured web activity logs as reliable behavioral indicators, which aligns with the present study's approach of analyzing timestamped browsing data to identify patterns in website usage and temporal consumption in different digital task or website categories.

Other studies have explored large-scale digital engagement using behavioral telemetry data. Analytics-based research platforms such as RescueTime [7] analyzed aggregated user activity logs to measure time spent across website categories, distinguishing between productive and distracting platforms. These studies primarily used descriptive statistics, regression analysis, and categorical grouping of applications. Their findings consistently indicate that users allocate a significant portion of time to entertainment and social media platforms, often exceeding perceived estimates. However, such research typically relies on commercial datasets and population-level aggregation, limiting individualized statistical evaluation.

Previous work has also investigated temporal patterns in digital activity. Chrono-biological researches [2] demonstrated that cognitive alertness fluctuates across the day according to circadian rhythms. Extending this perspective to online behavior, [3] analyzed large-scale timestamped social media data and identified systematic diurnal and weekly variations in engagement and mood. These studies commonly applied time-series analysis and statistical comparisons across hourly intervals. Their findings suggest that digital activity intensity differs significantly between morning and evening periods, as well as between weekdays and weekends.

Research on engagement duration further examined how platform type influences time consumption. Another study [4] analyzed gaming behavior and reported that engagement levels vary depending on content characteristics and user context. Surveys conducted by research centers [5] similarly highlighted differences in time allocation across educational, entertainment, and communication platforms. Many of these studies relied on self-reported screen-time data, which may introduce recall bias and measurement inaccuracies.

Methodologically, prior research in digital behavior analysis has commonly employed descriptive statistics, correlation analysis, regression modeling, and group comparison techniques. Analysis of Variance (ANOVA) has been widely recognized as an appropriate method for testing mean differences across categorical groups [6]. Despite the availability of such inferential methods, many studies emphasize large-scale aggregated datasets rather than structured, self-collected behavioral logs analyzed at the individual level.

While previous research has established that digital engagement varies across platforms and time periods, limited attention has been given to systematically analyzing personally collected browsing logs using formal hypothesis testing. The present study builds upon prior findings by applying descriptive statistics and ANOVA to a structured dataset containing URLs, timestamps, and time-consumption measures. By transforming raw web logs into categorical temporal variables and website groupings, this research extends existing literature by demonstrating how individualized browsing data can be evaluated using rigorous statistical methods.

III. Methodology

The study investigates the individual web browsing behavior through systematic logging and quantitative analysis of browser activities. Over a continuous 45-day period, browsing interactions were recorded at fixed time intervals to capture detailed temporal patterns of website usage. The collected data were structured, aggregated, and statistically examined to evaluate differences across websites and time-related categories. This section outlines the participant profile, data acquisition process, variable construction, preprocessing procedures, and analytical techniques implemented to ensure reproducibility and methodological transparency.

A. Participants

The participant in this study was a 22-year-old male Bachelor of Science in Computer Science student at National University Manila. The participant regularly uses a personal computer for academic tasks, software development, research, and personal online activities. A single-subject design was selected to enable detailed longitudinal tracking of individual browsing behavior without external variability from multiple users.

B. Data Collection Methods

Browsing activity was collected using a custom-developed chromium and mozilla extension implemented in Typescript. The extension recorded the currently active browser tabs at a fixed interval of 4 seconds and stored the logs locally using SQLite. Additionally, powershell were utilized for persistent communication between the local database and the extension.

Only active tabs were recorded to ensure that the logged time reflected actual user interaction rather than background or inactive browser sessions. Each log entry corresponds to a fixed 4-second interval of browsing activity.

Additionally, URL categories were manually labeled to segregate websites that are related to social media, entertainment, developer resources, games, educational platforms, and other miscellaneous content. This manual categorization enabled structured comparison across website types and ensured that browsing behavior could be analyzed at the category level rather than solely at the individual URL level.

Variables Collected:

- Base URL
- Full URL
- Timestamp (date and time)
- Seconds Elapsed (fixed at 4 seconds per entry)
- URL Category

Data were collected continuously over a 45-day period whenever the participant actively used the Google Chrome browser.

C. Data Processing

The raw dataset consisted of timestamped active-tab logs recorded at fixed 4-second intervals. The *date_object* variable was converted into a standardized datetime format to enable temporal decomposition.

From *date_object*, additional structured variables were generated, including *date*, *day_of_week*, *day_of_week_num*, *hour*, *week_number*, and categorical classifications such as *time_of_day* and *week_type*. The *is_weekend* indicator was derived from the numerical day encoding.

Following feature extraction, aggregated datasets were created by grouping records based on combinations of *date*, *base_url*, *time_of_day*, and *week_number*. Total browsing duration within each group was calculated by summing *seconds_elapsed*. Derived time metrics in minutes and hours were computed through unit conversion for reporting purposes in the results section.

D. Operational Definitions

To ensure clarity and reproducibility, the primary variables and columns used in the study are defined as follows:

- **base_url**: The primary domain of the visited website, used as the categorical variable for website-level analysis (string).
- **full_url**: The complete webpage address including its path and parameters (string).
- **date_object**: The exact timestamp datetime value recorded at the moment the active tab was logged (YYYY-MM-DD:HH format).
- **seconds_elapsed**: Fixed 4-second duration assigned to each recorded active-tab entry (number).
- **date**: The calendar date extracted from *date_object* (YYYY-MM-DD format).
- **day_of_week**: The categorical and textual representation of the day corresponding to the date (Monday–Sunday).
- **day_of_week_num**: Numerical encoding of the day of the week (Monday = 0, ..., Sunday = 6).
- **is_weekend**: Boolean indicator where True represents Saturday or Sunday, and False represents Monday to Friday (boolean).
- **hour**: The hour component extracted from *date_object* using 24-hour format in manila timezone (0–23).
- **time_of_day**: Categorical classification of *hour* into defined time segments (past_midnight, morning, noon or lunch hour, afternoon, evening).
- **week_number**: ISO-standard calendar week index (1–52/53) derived from *date_object* (number).
- **week_type**: Binary categorical classification indicating weekday (Monday–Friday) or weekend (Saturday–Sunday).
- **url_category**: Categorical classification of the base_url (entertainment, social media, developer, games, educational, others).

E. Data Cleaning

Prior to analysis, the dataset underwent preprocessing to ensure consistency and reliability. Entries containing null or incomplete values were removed. Records with zero time allocation were excluded. For aggregated variants of the dataset, total seconds of less than 60 were removed to reduce data noise. Additionally, Timestamp fields were standardized to a uniform datetime format. Derived categorical variables were validated to ensure accurate temporal classification. No outliers were removed, as all entries represented actual recorded browsing activity.

F. Experimental Setup

Data processing and statistical analysis were conducted using Python version 3.12.9 within Jupyter Notebook in Visual Studio Code. The following libraries were utilized and is using the latest versions:

- Pandas for data manipulation and aggregation
- NumPy for numerical computation
- Matplotlib and Seaborn for visualization
- SciPy for inferential statistical testing
- Scikit Posthocs for Posthoc Dunn testing

G. Statistical Analysis

This study collected browsing activity data at the URL level to analyze how time is allocated across website categories and temporal contexts. The primary variables recorded were: timestamp, base_url, url_category, and seconds_elapsed. The timestamp variable was used to derive additional temporal features, including hour, day_of_week, week_number, and time_of_day classifications (morning, noon, afternoon, evening, and past_midnight). The purpose of collecting these variables was to examine whether browsing behavior varies by website type and time-related factors.

The variable seconds_elapsed represents the duration of each logged browsing interval. Time was measured using a fixed-interval logging mechanism that records active browsing sessions in seconds. Website categories were assigned based on predefined classification rules (e.g., social_media, entertainment, developer, games, educational, others). Temporal variables were programmatically derived from the original timestamp using date-time parsing and transformation procedures.

The dataset covers a multi-week observation period, including Weeks 1 through 6 and Weeks 51 through 52 of the recorded year. This time span allows for short-term weekly comparisons, although it does not represent a full-year dataset.

Data preprocessing included converting timestamps into structured datetime format, generating derived temporal variables, checking for missing or incomplete entries, and aggregating raw interval-level data into three summarized dataset variants: (1) daily totals, (2) time-of-day totals, and (3) weekly summaries. Aggregation was performed to total the seconds per variant, and to better capture meaningful differences in cumulative time allocation across categories. These aggregated variants would be strictly used for statistical or inferential analysis.

Descriptive statistics, including total time and mean time, were computed to summarize browsing behavior. Visualizations such as bar charts, line plots, and heatmaps were generated to illustrate category dominance, total seconds allocated, and temporal trends.

For inferential analysis, one-way Analysis of Variance (ANOVA) was conducted to test whether mean total browsing time differed significantly across URL categories. Because browsing time distributions may not strictly satisfy normality assumptions, the non-parametric Kruskal–Wallis test was also performed as a robustness check. The significance level was set at $\alpha = 0.05$. When omnibus tests indicated significant differences, Dunn's post-hoc test with Bonferroni correction was applied to identify specific pairwise differences while controlling for Type I error. Dunn's test compares ranked distributions between groups rather than group means, making it appropriate for non-normally distributed continuous data. Non-significant results indicate insufficient evidence of difference but do not imply equality between categories Pairwise comparisons with $p > 0.05$ were interpreted as lacking sufficient statistical evidence to conclude a difference, rather than indicating equality between categories.

These three statistical methods were implemented due the aggregated variant's null hypothesis:

- **Daily Hypothesis**
  H_0: There is no statistically significant difference in total website usage time across URL categories using daily aggregated totals.
  H_a: At least one URL category differs significantly in daily aggregated usage time.
- **Time-of-Day Hypothesis**
  H_0: There is no statistically significant difference in total website usage time across URL categories using time-of-day aggregated totals.
  H_a: At least one URL category differs significantly in time-of-day aggregated usage time.
- **Weekly Hypothesis**
  H_0: There is no statistically significant difference in total website usage time across URL categories using weekly aggregated totals.
  H_a: At least one URL category differs significantly in weekly aggregated usage time.

Addionally, potential sources of bias and measurement error were considered. First, the dataset represents a single participant (n = 1), limiting generalizability. Second, fixed-interval logging may not perfectly capture passive or background browsing behavior. Third, manual or rule-based category classification may introduce minor categorization bias. Finally, the relatively short observation window may not fully represent long-term behavioral patterns.

IV. Results and Discussion

This chapter presents the empirical findings derived from the collected browsing dataset. The analysis includes descriptive statistics, visualization-based pattern identification, inferential statistical testing, and structured interpretation of results across daily, time-of-day, and weekly aggregations.

A. Dataset Overview

Table I presents the descriptive statistics of the original, non-aggregated dataset consisting of 321,460 timestamped records. The variable *sec_elapsed* has a mean of 3.87 seconds with a standard deviation of 0.58, indicating that most recorded browsing intervals cluster closely around four seconds. The minimum and maximum values range from 1 to 14 seconds, suggesting limited variability per recorded event due to the fixed-interval logging mechanism. Temporal features derived from timestamps include day of week (*dow*), hour of day (*hr*), and week number (*wk_num*). The average hour value of 15.48 indicates that a substantial portion of browsing activity occurred during the afternoon period. The interquartile ranges for *dow* and *hr* demonstrate that activity was distributed across both weekdays and weekends and throughout most daily hours, with concentration between 12:00 and 20:00.

TABLE I
Descriptive Statistics of the Dataset (Default Variant)

Stat	sec_elapsed	dow	hr	wk_num
count	321460	321460	321460	321460
mean	3.87	2.80	15.48	15.93
std	0.58	1.99	6.21	21.78
min	1.00	0.00	0.00	1.00
25%	4.00	1.00	12.00	2.00
50%	4.00	2.00	16.00	4.00
75%	4.00	5.00	20.00	51.00
max	14.00	6.00	23.00	52.00

TABLE II
Descriptive Statistics by Category (Daily Total)

Category	Count	Mean	Std	Min
developer	186.00	1314.55	2034.48	60.00
educational	34.00	1169.64	2061.46	62.00
entertainment	137.00	1817.23	2515.04	63.00
games	65.00	1453.49	3107.00	62.00
others	188.00	892.72	1668.40	60.00
social media	114.00	3835.07	4618.10	68.00

TABLE III
Descriptive Statistics by Category (Time Of Day Total)

Category	Count	Mean	Std	Min
developer	287.00	846.12	1295.83	60.00
educational	39.00	1018.97	1705.46	62.00
entertainment	228.00	1088.82	1420.76	63.00
games	78.00	1210.19	1975.34	62.00
others	242.00	685.49	1201.74	60.00
social media	303.00	1440.42	2051.64	60.00

Tables II to IV show the aggregated versions of the dataset at daily, time-of-day, and weekly levels, respectively. After aggregation, substantial variability emerges across website categories. In the daily total aggregation (Table II), social media exhibits the highest mean total time (3835.07 seconds) and the largest dispersion (std = 4618.10), indicating both frequent and highly variable engagement. Entertainment and

TABLE IV
DESCRIPTIVE STATISTICS BY CATEGORY (WEEKLY SUMMARY TOTAL)

Category	Count	Mean	Std	Min
developer	115.00	2136.79	3708.96	60.00
educational	26.00	1531.69	4032.99	62.00
entertainment	75.00	3328.53	6117.53	63.00
games	55.00	1722.18	5710.53	62.00
others	129.00	1307.40	2641.78	60.00
social media	45.00	9718.33	13161.96	68.00

developer categories also show relatively high average totals compared to educational and others.

In the time-of-day aggregation (Table III), social media remains dominant in terms of mean total usage (1440.42 seconds), followed by games and entertainment. The standard deviations across categories remain large relative to their means, suggesting inconsistent usage intensity across different time segments from morning until past midnight.

At the weekly aggregation level (Table IV), differences between categories become more pronounced. Social media demonstrates a markedly higher mean weekly total (9718.33 seconds) compared to all other categories, along with the highest variability (std = 13161.96). Entertainment and developer categories follow, while educational and others maintain comparatively lower weekly totals.

B. Data Visualization

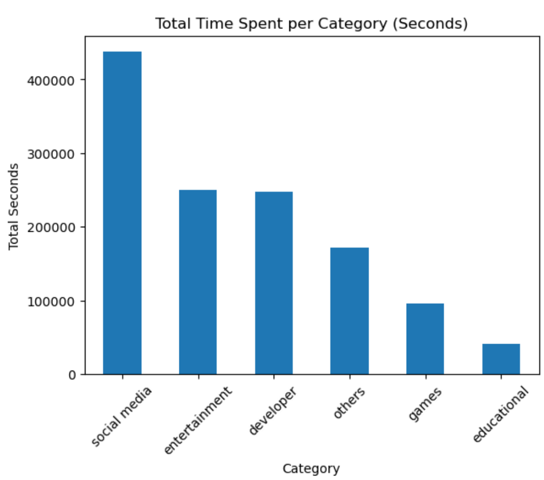

Fig. 1. Total Time Spent per Category (Seconds)

The non-aggregated dataset variant represents the raw, interval-level browsing records before any temporal consolidation. In this form, total usage is reflected by the accumulation of short, fixed-duration entries per website and category rather than by long continuous sessions. Figure 1 illustrates the total time spent per category. Social media records the highest total time (437,446 seconds), followed by entertainment (250,067 seconds) and developer-related websites (247,380 seconds). The others category accounts for 171,323 seconds, while games (96,065 seconds) and educational websites (40,544 seconds) contribute comparatively smaller totals. This distribution indicates that browsing activity is heavily concentrated on social and entertainment-related platforms.

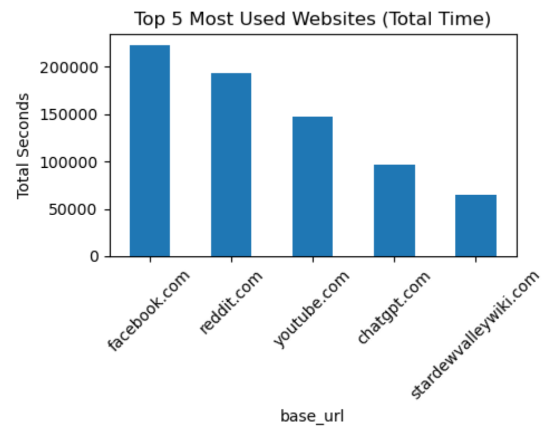

Fig. 2. Top 5 Most Used Websites

The top five most used websites, shown in Figure 2, further reinforce this pattern. Facebook.com ranks first with 223,272 seconds, followed by reddit.com (193,405 seconds), youtube.com (147,784 seconds), chatgpt.com (96,999 seconds), and stardewvalleywiki.com (64,969 seconds). These results show that a small number of platforms account for a substantial portion of total browsing time, suggesting concentrated usage behavior rather than evenly distributed engagement across many websites.

Fig. 3. Correlation Matrix of the Non-Aggregated dataset variant

The correlation matrix presented in Figure 3 shows minimal relationships between most numerical features. The correlation between *seconds_elapsed* and temporal variables such as *day_of_week_num* (-0.02), *hour* (-0.00), and *timeofday_numeric* (0.00) is negligible, indicating that per-interval duration remains largely constant regardless of time context. However, a strong positive correlation (0.95) is observed between *hour* and *timeofday_numeric*, which is expected since the time-of-day variable is directly derived from the hour variable. This confirms that the dataset does not exhibit artificial multi-collinearity beyond derived temporal features.

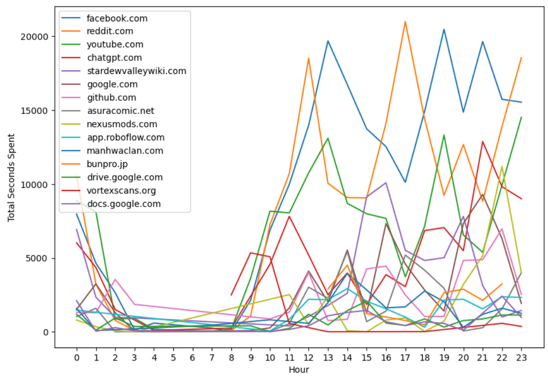

Fig. 4. Total Time Spent on Websites per Nth Hour

Figure 4 shows total time spent per hour across the top websites. Minimal activity is observed during early morning hours (approximately 02:00–08:00), followed by a steady increase beginning around 09:00. Usage peaks between 12:00 and 23:00, indicating that browsing behavior is concentrated in the noon, afternoon, and evening periods. A limited number of websites dominate total time per hour, while others display lower engagement levels.

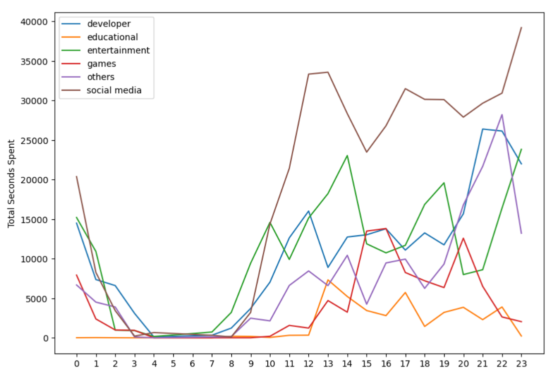

Fig. 5. Total Time Spent on Websites by Categories per Nth Hour

When examining hourly usage by category in Figure 5, the same temporal concentration pattern is observed. Social media consistently records the highest hourly totals, particularly during afternoon and evening hours. Entertainment and developer categories also show strong evening activity, whereas games exhibit moderate spikes in mid-afternoon and early evening. Educational usage remains comparatively low across most hours, with only modest increases in the afternoon. Overall, internet engagement is predominantly concentrated in the latter half of the day which also includes the noon period.

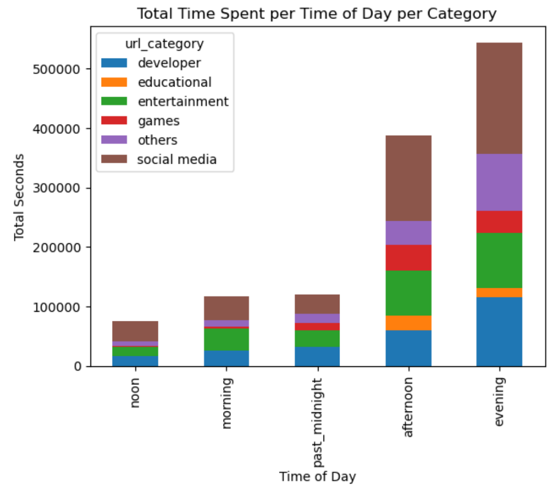

Fig. 6. Total Time Spent per Time of Day per Category

Time-of-day aggregation in Figure 6 further highlights these trends. Evening records the highest totals across most categories, particularly social media (187,995 seconds) and developer websites (115,235 seconds). Afternoon also shows substantial engagement, especially for social media (143,720 seconds) and entertainment (75,599 seconds). Morning and noon periods reflect moderate usage, while past midnight activity is relatively lower, except for noticeable contributions from developer (31,649 seconds) and social media (32,953 seconds). These patterns confirm that browsing intensity increases progressively throughout the day and peaks during evening hours.

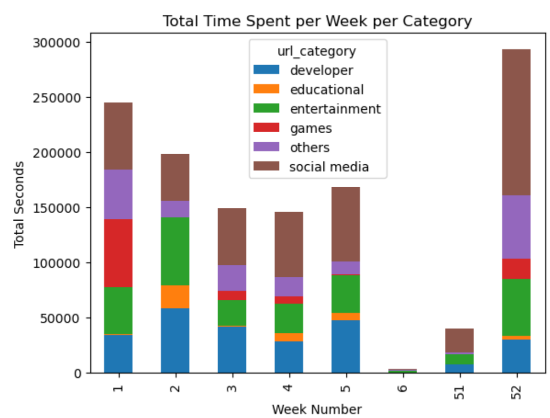

Fig. 7. Total Time Spent per Week per Category

Weekly totals shown in Figure 7 indicate variability across different weeks. Week 52 exhibits the highest totals across several categories, including social media (132,766 seconds) and others (57,212 seconds), suggesting a possible seasonal or end-of-year effect. Weeks 1 through 5 show relatively consistent but fluctuating engagement levels, while week 6 records minimal activity due to examination seasons.

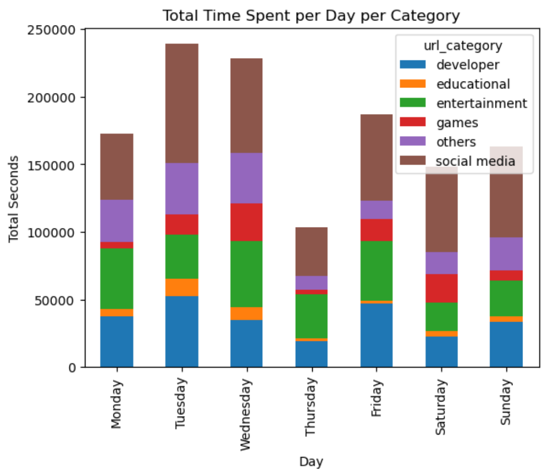

Fig. 8. Total Time Spent per Day per Category

Daily distribution presented in Figure 8 reveals that browsing activity is not evenly spread across weekdays. Tuesday records the highest social media usage (88,252 seconds), Thursday is relatively low due to class schedules, while Sunday and Wednesday also show elevated totals. Developer-related activity is highest on Tuesday (52,398 seconds) and Friday (47,356 seconds), suggesting stronger productivity-related browsing on specific weekdays. Weekend days, particularly Saturday and Sunday, maintain high social media engagement but show varied patterns in other categories.

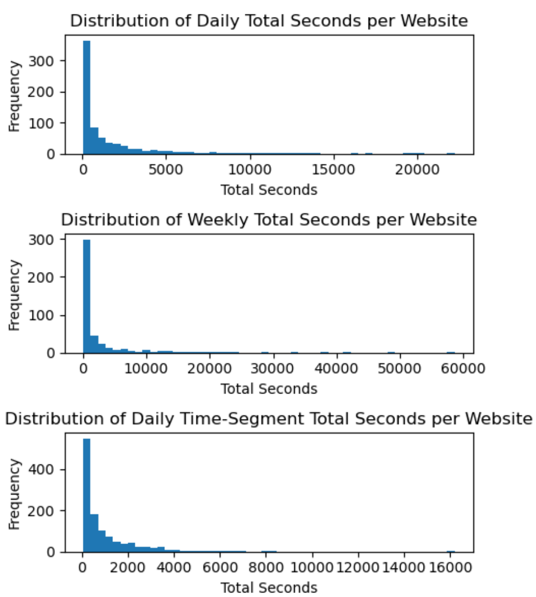

Fig. 9. Total Time Spent per Day per Category

Figure 9 shows the distribution of total seconds per website aggregated at the daily level, together with the weekly and daily time-segment summaries. Across all three histograms, the distributions exhibit strong right skewness. The majority of observations are concentrated at low total-second values, while only a small number extend toward extremely high durations, forming long right tails. In the daily aggregation, most websites accumulate relatively short usage times within a single day. The weekly aggregation displays a wider range of values, as usage accumulates across multiple days, yet it maintains the same positively skewed structure. Similarly, the daily time-segment aggregation demonstrates a dense clustering of low-duration observations with fewer high-duration cases.

The consistent right-skewed pattern across all aggregation levels indicates that website usage is unevenly distributed, with a small subset of website–period combinations accounting for disproportionately high total seconds. The presence of extreme values and the lack of symmetry suggest that the data do not follow a normal distribution. This distributional behavior reflects typical digital usage patterns, where user attention is concentrated on a limited number of websites, while most receive relatively brief interactions within a given day, week, or time segment.

C. *Statistical Analysis Result*

Statistical tests were conducted to determine whether total time spent differs significantly across URL categories for the three aggregated dataset variants: daily totals, time-of-day totals, and weekly summaries. Both parametric (ANOVA) and non-parametric (Kruskal–Wallis) tests were applied to ensure robustness of results.

TABLE V
KRUSKAL–WALLIS TEST RESULTS

Dataset	**H-statistic**	**P-value**
daily_total	85.1238	7.0912e-17
time_of_day_total	85.9967	4.6521e-17
week_summary	45.9848	9.1470e-09

Table V presents the Kruskal-Wallis test results. All three dataset variants produced highly significant results: daily_total ($H = 85.1238$, $p = 7.0912 \times 10^{-17}$), time_of_day_total ($H = 85.9967$, $p = 4.6521 \times 10^{-17}$), and week_summary ($H = 45.9848$, $p = 9.1470 \times 10^{-9}$). Since all p-values are far below the 0.05 significance threshold, the null hypothesis of equal median time across categories is rejected for all aggregation levels. This indicates that at least one category differs significantly in total time allocation.

Similarly, the ANOVA results shown in Table VI confirm these findings. The F-statistics for daily_total ($F = 18.5148$, $p = 2.5347 \times 10^{-17}$), time_of_day_total ($F = 7.3048$, $p = 9.4460 \times 10^{-7}$), and week_summary ($F = 15.3042$, $p = 7.0667 \times 10^{-14}$) are all statistically significant. The agreement between ANOVA and Kruskal-Wallis results strengthens the conclusion that browsing time distribution differs significantly across URL categories in all aggregated variants.

TABLE VI
ANOVA TEST RESULTS

Dataset	F-statistic	P-value
daily_total	18.5148	2.5347e-17
time_of_day_total	7.3048	9.4460e-07
week_summary	15.3042	7.0667e-14

To identify which categories differ, Dunn's post-hoc test with Bonferroni correction was applied.

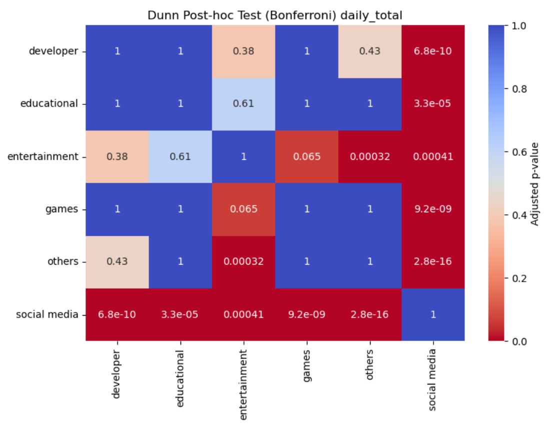

Fig. 10. Dunn Post-hoc Test (Bonferroni) For Daily_Total

Figure 10 shows the pairwise comparisons for the daily_total dataset. Social media demonstrates statistically significant differences with nearly all other categories, including developer ($p = 6.83 \times 10^{-10}$), educational ($p = 3.34 \times 10^{-5}$), games ($p = 9.15 \times 10^{-9}$), and others ($p = 2.83 \times 10^{-16}$). Entertainment also differs significantly from others ($p = 0.000322$) and social media ($p = 0.000410$). These results indicate that daily time spent on social media is substantially different from most other categories.

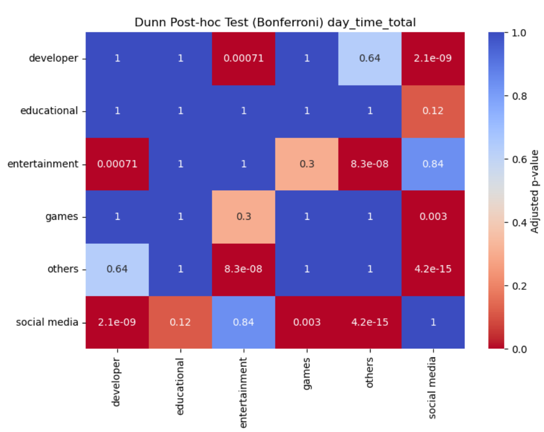

Fig. 11. Dunn Post-hoc Test (Bonferroni) For Time_of_Day_Total

Figure 11 presents the post-hoc results for the time_of_day_total dataset. Social media again shows significant differences with developer ($p = 2.12 \times 10^{-9}$), games ($p = 0.00299$), and others ($p = 4.16 \times 10^{-15}$). Entertainment differs significantly from developer ($p = 0.000711$) and others ($p = 8.33 \times 10^{-8}$). Educational, however, does not show widespread significant differences in this aggregation level, indicating relatively stable time allocation compared to other categories during different times of the day.

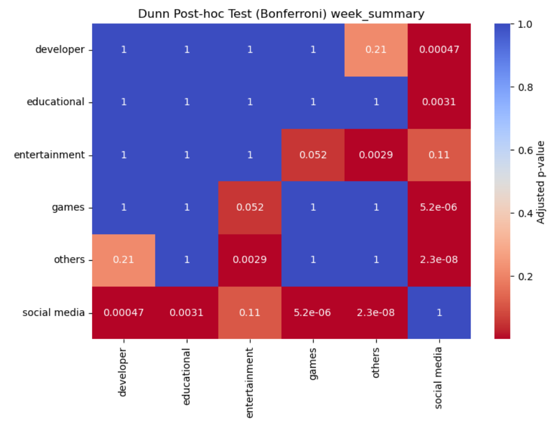

Fig. 12. Dunn Post-hoc Test (Bonferroni) For Week_Summary

For the week_summary dataset (Figure 12), social media continues to exhibit significant differences with developer ($p = 0.000468$), educational ($p = 0.00308$), games ($p = 5.18 \times 10^{-6}$), and others ($p = 2.25 \times 10^{-8}$). Others also differs significantly from entertainment ($p = 0.00294$). Compared to daily and time-of-day variants, fewer pairwise differences are observed at the weekly level, suggesting that aggregation over longer periods reduces variability between certain categories.

D. Discussion

A. Interpretation of Results: The results indicate that browsing behavior is not uniformly distributed across URL categories. Across all aggregation levels (daily, time-of-day, and weekly), social media consistently recorded the highest total time and demonstrated statistically significant differences from most other categories. This suggests that social platforms occupy a dominant portion of overall digital engagement.

Temporal analysis further reveals that usage peaks during noon, afternoon and evening hours (12:00–23:00), with minimal activity during early morning hours (02:00–08:00). This pattern likely reflects typical daily routines, where academic, professional, or personal obligations occur earlier in the day, followed by increased leisure-oriented browsing later. The strong presence of entertainment and gaming activity during evening periods may indicate recreational behavior after structured daytime activities.

Weekly variability, particularly the elevated totals in Week 52, may suggest contextual influences such as holidays, reduced workload, or increased leisure time. Conversely, the unusually low activity in Week 6 may reduced browsing time due to examination seasons.

The statistical tests confirm that these observed differences are not due to random variation. Both ANOVA and Kruskal–Wallis tests reject the null hypothesis across all aggregated variants, while Dunn's post-hoc analysis identifies social media as the most distinct category. This implies that category-based segmentation meaningfully captures behavioral differences in time allocation. However, p-values on greater than 0.05 on multiple pair-wise category indicates that it there are no statistical evidence that proves those two URL categories are different, primarily due the dataset being spread out, or being biased on certain day, week, or time of day. An example would be that some days, games is longer than entertainment, or in some weeks developer is longer than entertainment. The dataset is biased on social media, as to why we can see its significant differences on all other categories. However, social media category also has a high variability that overlaps with other much lower variability such as education, which in turn produces a much higher p-value.

Overall, the results suggest that browsing behavior is driven by habitual engagement with specific platforms, particularly social media, and is strongly influenced by time-of-day and weekly context.

B. Comparison to Related Work: The findings align with prior research on digital behavior patterns, which consistently report that social media platforms account for a substantial portion of daily screen time. Previous studies on internet usage behavior have shown increased activity during evening hours, particularly for entertainment and social networking services. The observed concentration of usage in the second half of the day is consistent with these established behavioral trends.

Additionally, research on digital engagement often reports that leisure-oriented categories dominate total usage time compared to educational or productivity-related websites. The comparatively lower totals observed for educational websites in this study are consistent with such findings.

However, unlike large-scale behavioral studies that examine multiple participants, this study focuses on a single-user dataset. While the trends align with broader patterns reported in literature, the magnitude and specific platform dominance (e.g., Facebook and Reddit) reflect individual-level preferences rather than population-wide generalizations.

C. Limitations: Several limitations must be acknowledged for this study.

First, the dataset represents a single participant (n = 1). As a result, findings cannot be generalized beyond the observed user. The results reflect individual habits rather than broader behavioral norms.

Second, although the logging mechanism automatically captured browsing intervals, potential missing entries or inactive background states may have influenced recorded totals. Short interruptions or unlogged sessions could slightly distort true usage patterns.

Third, the data collection window is relatively limited in duration. Seasonal variations, academic workload changes, or long-term behavioral shifts cannot be fully captured within a short observation period.

Fourth, category assignment may introduce classification bias, particularly for websites that serve multiple purposes (e.g., educational content on social platforms). This could slightly affect comparisons between categories.

Fifth, the url categories were manually labeled, that could lead to unbalanced empirical results due to bias and misclassification.

Finally, while statistical tests show significant differences, the small sample size at aggregated levels (especially weekly summaries) reduces statistical power and limits inferential strength. Additionally, statistical tests were only conducted on differences across URL categories within each aggregated dataset, and temporal variables such as day, time-of-day, and week were not treated as independent factors in a multi-factor or interaction analysis.

D. Recommendations and Future Work: Future research could improve upon this study in several ways.

First, increasing the sample size to include multiple participants would allow for population-level comparisons and more robust statistical inference. Group-level analysis could identify whether the observed dominance of social media is consistent across users or varies significantly by demographic factors.

Second, extending the data collection period would capture long-term trends, seasonal effects, and academic or occupational cycles. A longitudinal design would provide deeper insight into behavioral stability and change.

Third, additional variables could be incorporated, such as device type, application foreground/background state, sleep duration, productivity metrics, or academic performance indicators. Including contextual variables would enable more comprehensive modeling of digital behavior.

Fourth, future studies may explore alternative analytical methods, including clustering techniques to identify behavioral profiles, time-series modeling for trend detection, or predictive modeling to forecast high-usage periods.

Fifth, utilizing systematic categorizer could greatly improve the variation of website categories, that leads to a more meaningful empirical results.

For students conducting similar research, careful planning of logging mechanisms, consistent category classification, and extended data collection are recommended to ensure reliable and meaningful results.

V. Conclusion

This study aimed to analyze personal browsing behavior by examining total time spent across URL categories using non-aggregated and aggregated dataset variants (daily, time-of-day, and weekly summaries). The primary objective was to determine whether browsing time differs significantly across website categories and to identify temporal patterns in usage behavior.

The key findings show that browsing activity is not evenly distributed across categories. Social media consistently recorded the highest total time across all aggregation levels and demonstrated statistically significant differences compared to most other categories. Usage patterns were strongly concentrated in the afternoon and evening hours (12:00–23:00), while early morning activity remained minimal. Weekly analysis also revealed fluctuations in engagement, with certain weeks exhibiting substantially higher totals, suggesting contextual or seasonal influences.

From a personal perspective, the analysis highlights a clear dominance of social media and entertainment platforms in overall screen time. Productivity-related browsing, such as developer websites, is present but does not outweigh leisure-oriented engagement. The temporal concentration of usage in the latter half of the day suggests that digital activity increases after structured daily responsibilities, reflecting a pattern of recreational browsing during free time.

These findings can be applied in real life by increasing awareness of time allocation across digital categories. Recognizing that a large portion of time is devoted to social media provides an opportunity to implement self-regulation strategies, such as time limits, scheduled usage periods, or intentional allocation toward educational or productivity-related platforms. Understanding peak usage hours may also support better time management and improved balance between academic, professional, and recreational activities.

In conclusion, this study demonstrates that personal browsing behavior exhibits statistically significant differences across website categories and is strongly influenced by temporal factors. While limited to a single-user dataset, the findings provide meaningful insight into digital habits and highlight the value of data-driven self-reflection. By quantifying and analyzing browsing patterns, individuals can gain greater awareness of their behavior and make informed adjustments to improve productivity and overall digital well-being.

References

[1] G. Mark, D. Gudith, and U. Klocke, "The cost of interrupted work: More speed and stress," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, Apr. 2008, pp. 107–110, doi: 10.1145/1357054.1357072.

[2] T. H. Monk, "The post-lunch dip in performance," *Clinics in Sports Medicine*, vol. 24, no. 2, pp. 15–23, 2005, doi: 10.1016/j.csm.2004.12.001.

[3] S. A. Golder and M. W. Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures" *Science*, vol. 333, no. 6051, pp. 1878–1881, 2011, doi: 10.1126/science.1202775.

[4] A. K. Przybylski, "Electronic gaming and psychosocial adjustment," *Pediatrics*, vol. 134, no. 3, pp. e716–e722, 2014, doi: 10.1542/peds.2013-4021.

[5] Pew Research Center, "Children's engagement with digital devices, screen time," Pew Research Center, Washington, DC, USA, 2021. [Online]. Available: https://www.pewresearch.org/.

[6] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. London, U.K.: Sage Publications, 2013.

[7] Connecteam, "Honest RescueTime Review 2026: Pros, Cons, Features & Pricing," Oct. 30, 2024. [Online]. Available: https://connecteam.com/reviews/rescuetime/. [Accessed: Feb. 17, 2026].