



National University - Manila

Behavioral Patterns in Web Activity: A Time and Website Analysis

RESEARCH PROJECT PRESENTATION

Data Science

RAINNAND MONTANIEL



INTRODUCTION

Browsing behaviour

- Excessive or unbalanced digital consumption are common nowadays, can affect productivity, focus, overall time management.
- Inability to track time spent on digital platforms (time management).
- Consumption varies on content type (games, education) and time-related factors (day, time of day, week)

The Problem

- Existing studies use commercial & large-scale datasets to perform analysis, but not specialized for personal tracking of digital consumption.
- Limited focus on individualized statistical analysis of browsing behaviour



INTRODUCTION

The Goal

- Collect datasets consisting of time stamped website activities
- Record the website activities at daily, different time of the day, and weekly to enable systematic & aggregated statistical evaluation.
- Examine dominant website categories, peak digital activity, and time-based differences in browsing behavior.

Research Question

- Which type of websites accounts for the highest total usage time in the collected dataset?
- What temporal patterns of digital consumption are observed across different periods of the day?
- Is there a statistically significant difference in total website usage time across URL categories under different temporal aggregations (daily, time-of-day, and weekly)?



RELATED LITERATURE

Digital Activity Monitoring

- Automated activity loggings were utilized to measure how different tasks reduce cognitive performance and productivity.
- Large-scale studies analyzes productivity through time measurement across website categories, this supports the goal of the study on segmenting websites on multiple categories.

Temporal Patterns

- Chrono-biological research suggest that cognitive alertness fluctuates depending on their circadian rhythm, which means digital consumption could vary depending on time of the day.
- Digital consumption differs across different categories. Self-reported studies reported that there are bias on time allocation on educational, entertainment, and communication apps.
- These studies are self-reported and generalized, which rises the need for individualized analysis.






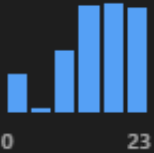
M E T H O D O L O G Y

- Data Collection: Utilizes typescript, sqlite, and powershell for website extraction (category were labeled manually). (45-days)
- Initial Variables: URL, Date, Seconds, Category
- Data Processing: Derived time of day, week number, hour, day of week from Date Variable. Aggregated datasets were also made for statistical analysis
 - time of day
 - daily
 - weekly summary
- Data Cleaning:
 - Removed null url, null dates, 0-second
 - Date variables were standardized (manila time)
 - No outliers removed to properly measure time usage

```
export default defineBackground(() => {  
  let _socket: WebSocket | null = null;  
  let lastLogTime = new Date();  
  let timeList = new Map<string, TimeListData>();  
  // let lastPingTime = new Date();  
  browser.runtime.onInstalled.addListener((details) => {  
    console.log("Extension installed:", details);  
  });  
  async function init_intervals() {  
    // pingServer();  
    incrementor();  
    await log_to_server();  
  }  
  init_intervals().catch(console.error);  
  
  // sends information to the server every 4 to 5 seconds  
  async function log_to_server() {  
    const currentTime = new Date();  
    const elapsedGeneral = currentTime.getTime() - lastLogTime.getTime();  
  
    // if its not yet 5 secs come back again shortly  
    if (elapsedGeneral < 5000) {  
      setTimeout(log_to_server, 100);  
      return;  
    }  
    // console.log("log to server runnin");  
    try {  
      // this will run after 5 seconds elapsed from the lastLogTime and the currentTime  
      browser.tabs.query({ active: true }).then(async (tabs) => {  
        for (let tab of tabs) {  
          // make sure the active tab is not a system tab  
          if (  
            tab &&  
            tab.url &&  
            tab.id &&  
            /^https?:\/\//.test(tab.url)  
          ) {  
            // get the content of that tab  
            const feedback = (await reqManipulate({  
              tabId: tab.id,  
            })).feedback;  
          }  
        }  
      });  
    }  
  }  
});
```




M E T H O D O L O G Y

base_url	full_url	date_object	# secon...	date	day_of...	# day...	is_we...	# hour	time...	# week...	week...	# time...	url_category
Missing: 0 (0%) Disti... 405 (<1%)	Missing: 0 (0%) Distinct: 17000 (5%)	Missing: 0 (0%) Distinct: 238731 (74%)	Missing: 0% Distinct: <1%	Missing: 0% Distinct: <1%	Missing: 0% Distinct: <1%	Missing: 0% Distinct: <1%	Missing: 0% False: 75% True: 25%	Missing: 0% Distinct: <1%	Loading...	Loading...	Loading...	Loading...	Loading...
facebook.c... 18% reddit.com 15% youtube.co... 12% Other 55%	17000 Distinct values	238731 Distinct values		45 Distinct values	Tuesday 19% Wednes... 18% Friday 15% Other 48%								
youtube.com	https://www.youtube.com/	2025-12-20 22:40:22.129000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:27.131000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:32.135000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:37.146000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:42.152000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:47.165000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:52.174000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:40:57.184000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:07.196000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:12.202000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:17.206000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:22.211000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:27.220000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:32.227000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:37.238000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:41:42.243000+08:00	3	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/resul	2025-12-20 22:41:47.252000+08:00	1	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/watc	2025-12-20 22:41:52.258000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/watc	2025-12-20 22:41:57.259000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/watc	2025-12-20 22:42:02.271000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:12.282000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:17.295000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:22.305000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:27.317000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:32.320000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/	2025-12-20 22:42:37.331000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/resul	2025-12-20 22:42:47.347000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/resul	2025-12-20 22:42:52.352000+08:00	1	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment
youtube.com	https://www.youtube.com/watc	2025-12-20 22:42:57.356000+08:00	4	2025-12-20	Saturday	5	True	22	evening	51	weekend	4	entertainment

developer, educational, entertainment, games, others, social media



M E T H O D O L O G Y

Statistical Analysis

- Computed total and mean browsing time; visualized category dominance and temporal trends using bar charts, line plots, and heatmaps.
- Aggregated variants of the dataset were made for statistical analysis of significance (cumulative sec).
- Conducted one-way ANOVA to test differences in mean browsing time across URL categories.
- Applied the Kruskal–Wallis test and Dunn’s post-hoc test with Bonferroni correction ($\alpha = 0.05$) to validate significant differences.

- **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.



RESULTS AND DISCUSSION

Default Variant of the dataset (non aggregated)

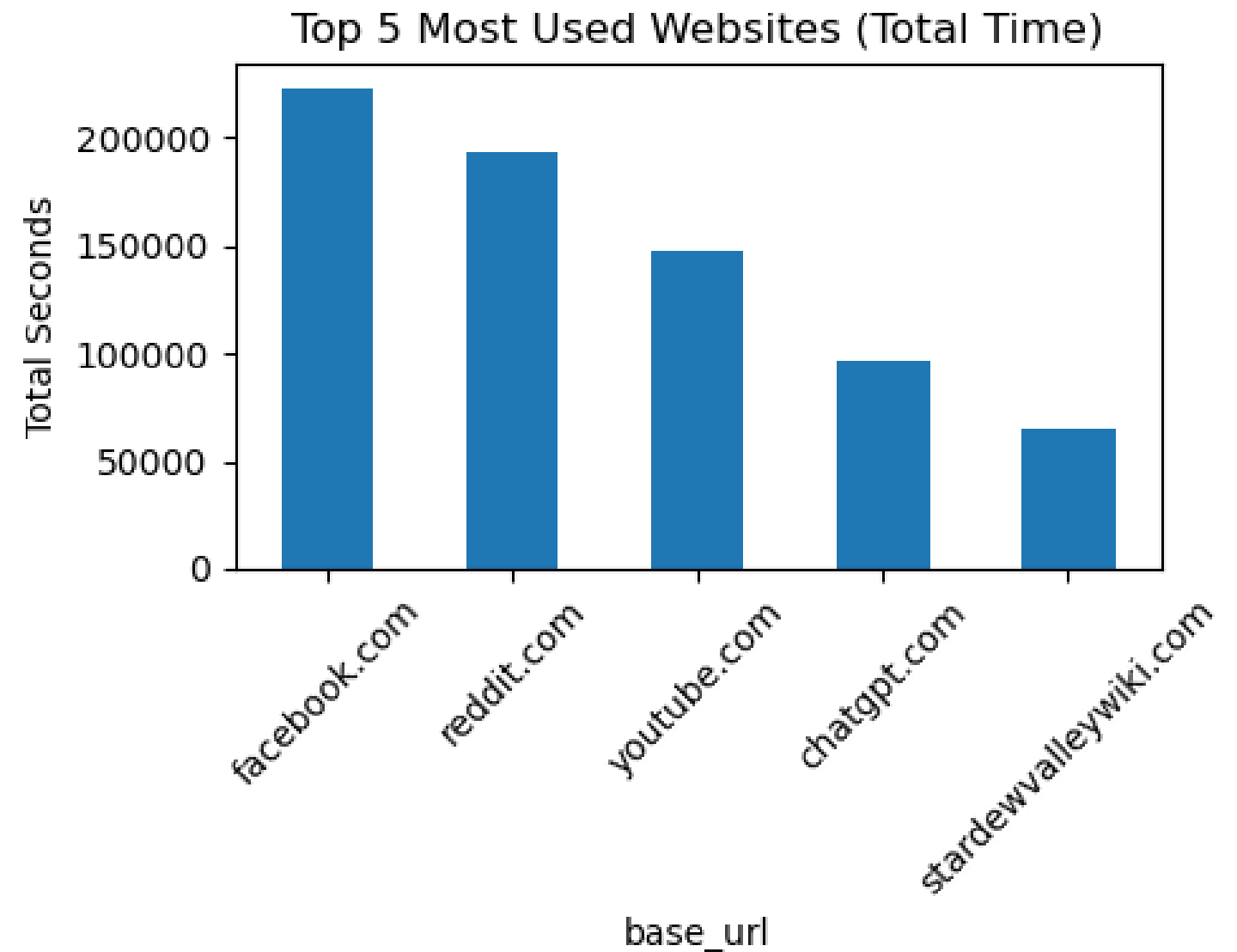
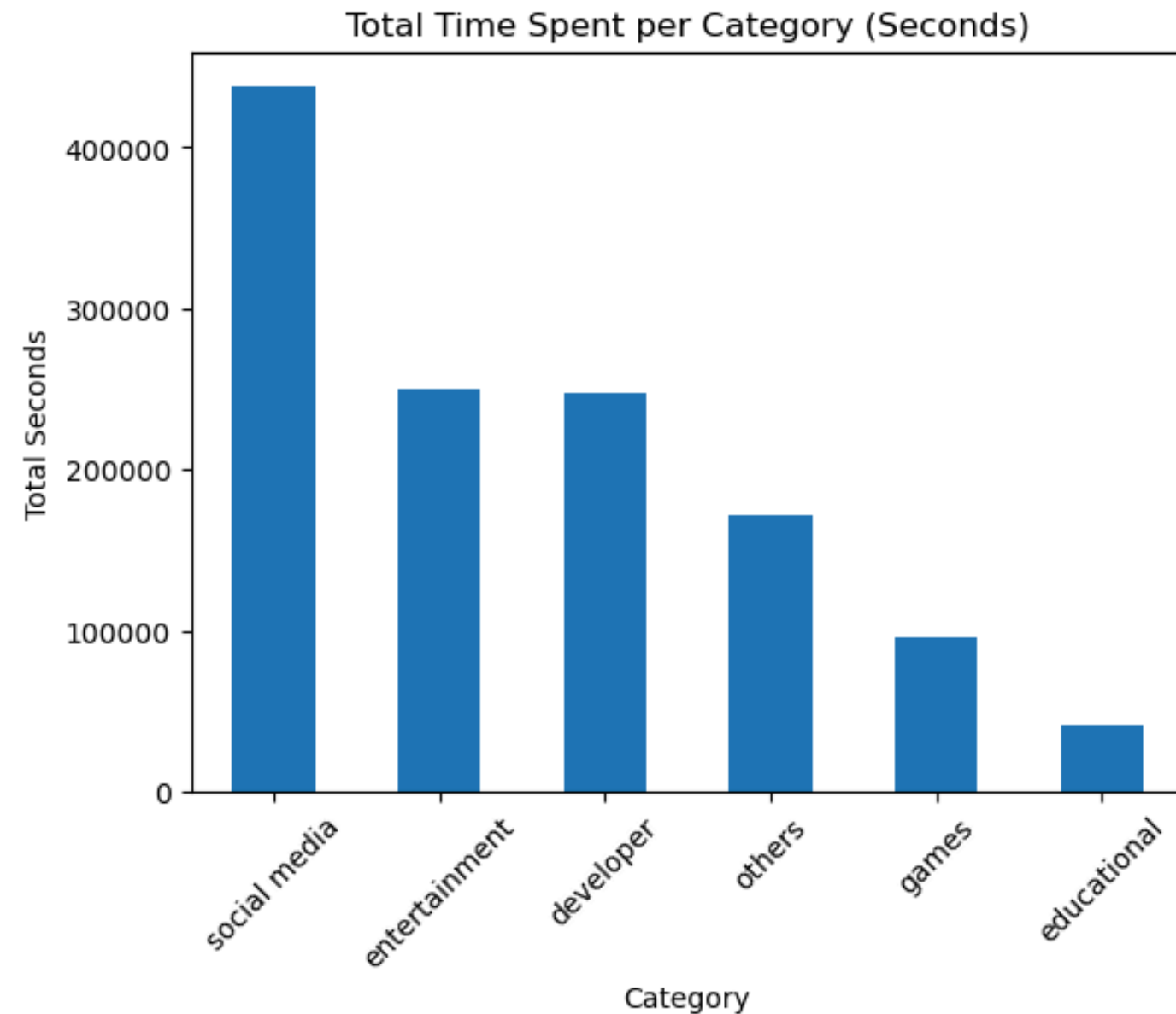
DESCRIPTIVE STATISTICS OF THE DATASET (DEFAULT VARIANT)

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



RESULTS AND DISCUSSION

Data Visualization

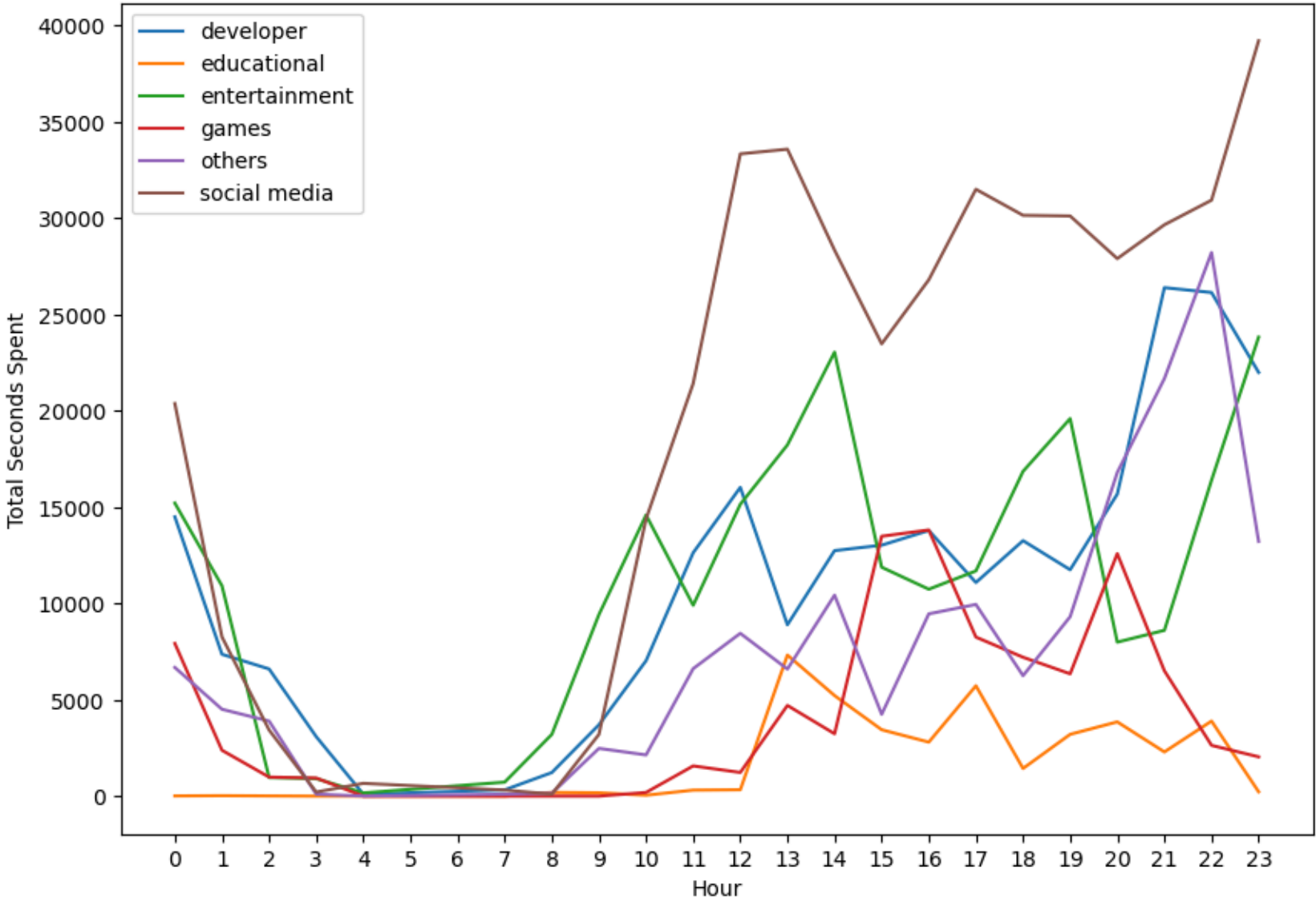
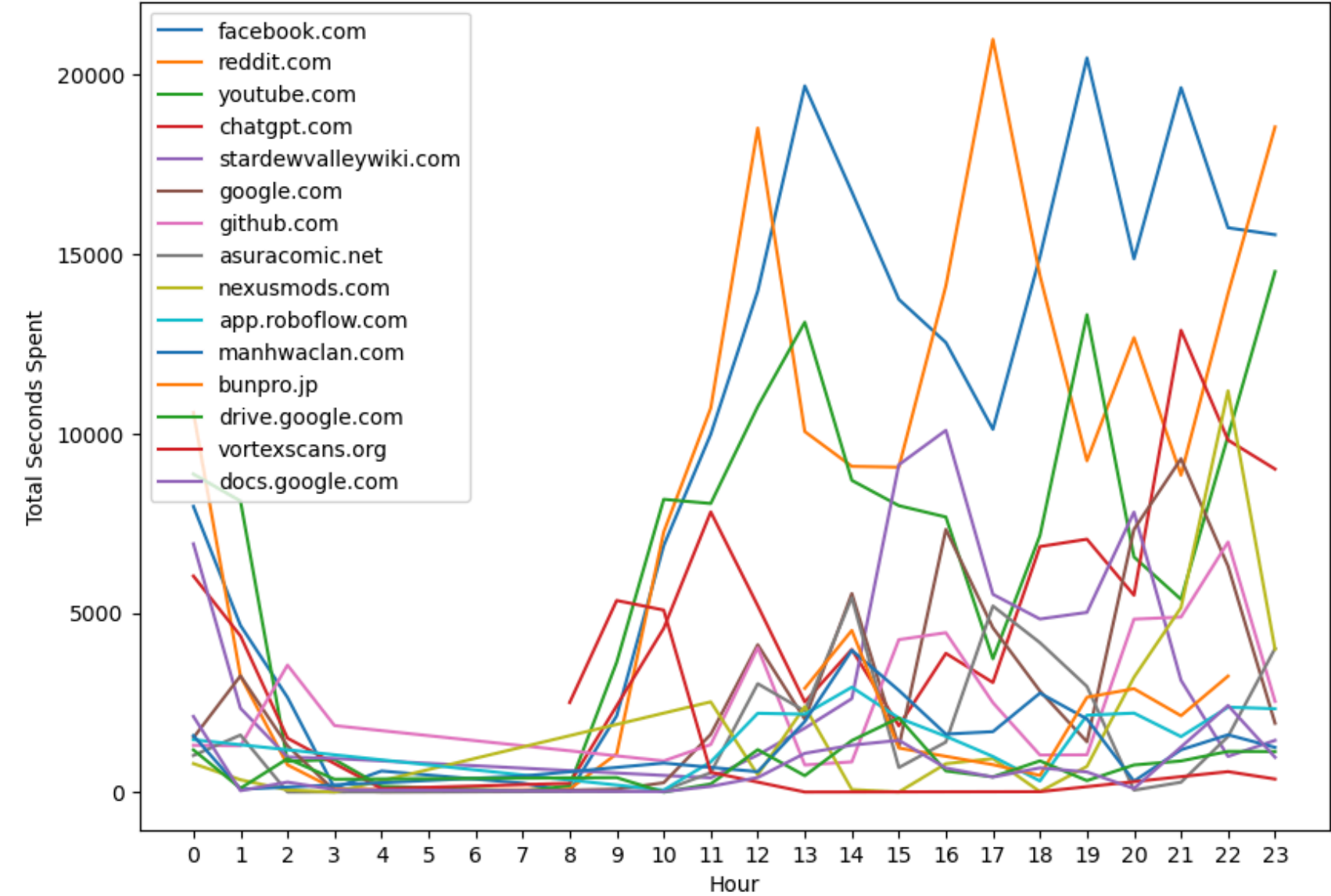


- Social Media has the highest total of consumption (437,446s => 121.51hr), while education is the least (40,544s => 11.26hr) in a span of 45 days.
- This answers the research question of the highest total usage time across all categories



RESULTS AND DISCUSSION

Multi-Series Line Chart





RESULTS AND DISCUSSION

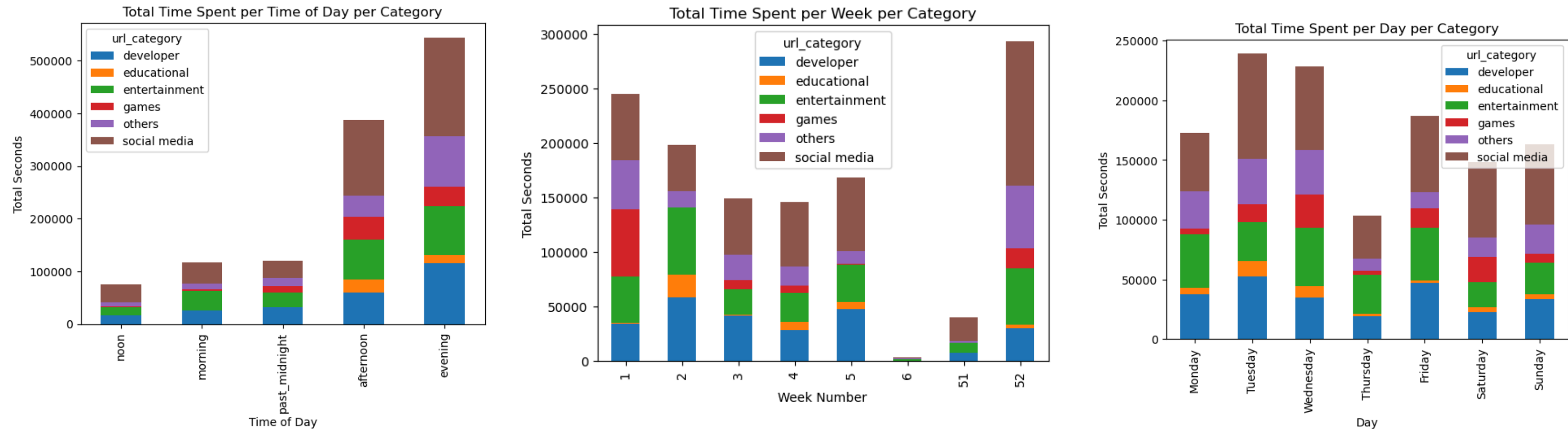
Multi-Series Line Chart

- Social media is still the highest consumption each nth hour.
- Entertainment and Developer also shows strong evening activities (18:00 - 23:00).
- Games shows spikes in mid-afternoon (15:00 - 16:00)
- Educational remains relatively low (with afternoon spikes).
- Temporal digital consumptions were found, peak hours of consumption are revealed to be noon (lunch time), afternoon, and evening (12:00 - 23:00) (2nd half of the day)
- Academic, Professional, Other obligations occur earlier in the day. But leisure-oriented browsing occurs later in the day.



RESULTS AND DISCUSSION

Stacked Bar Chart



- Increases progressively throughout the day and peaks during evening hours (daily routine)
- consistent usage on week 1-5 Jan, minimal activity on 6, week 52 is highest (holidays)
- Tuesday exhibits the highest, while Thursday is the lowest due to long class schedules.
- Based on this we can make aggregations (daily, hourly, weekly) to further find out if there's any significant differences between each variants.



Descriptive analysis

- Across all categories and variants, social media maintained the highest mean (3835, 1440, 9718) of total cumulative seconds, indicating that it is the most visited category in the dataset.
- Educational has the lowest totals and count, indicating that it is not the most visited category across all variants.
- All category maintained high STD (1000+), meaning the dataset is highly dispersed.

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

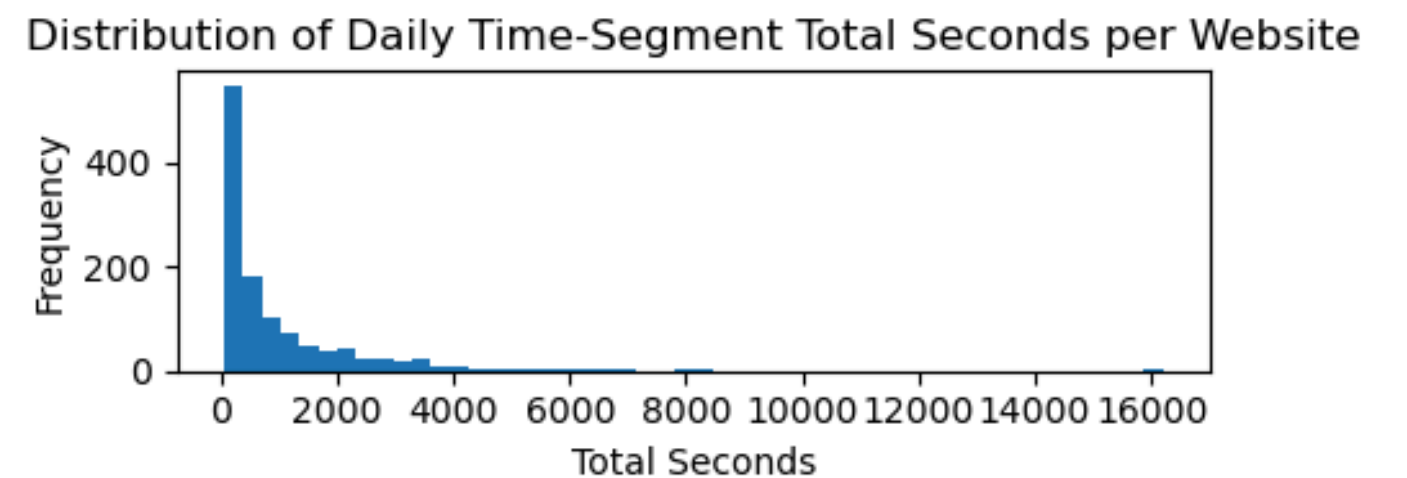
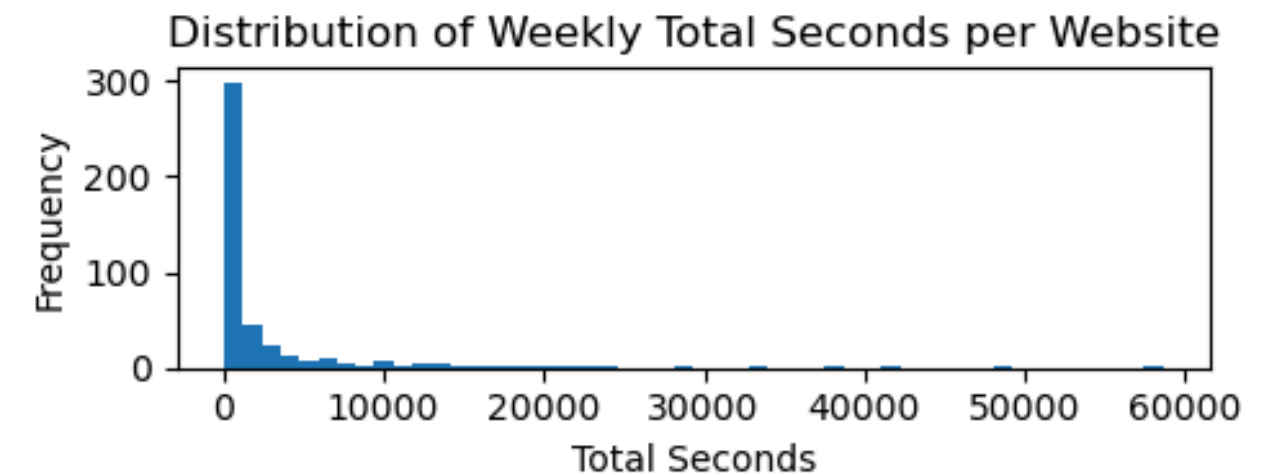
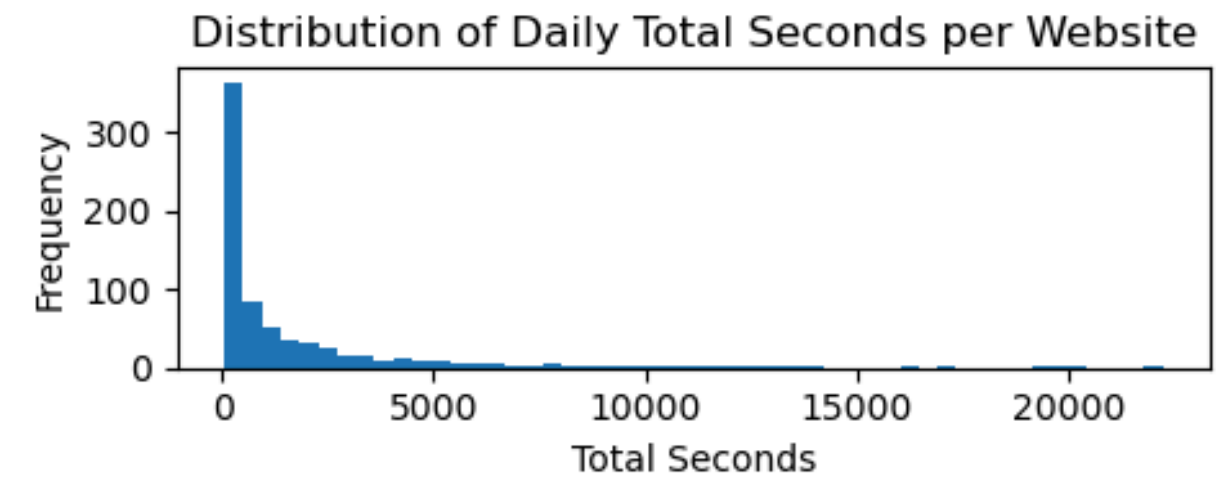
TABLE III DESCRIPTIVE STATISTICS BY CATEGORY (TIME OF DAY TOTAL)				
Category	Count	Mean	Std	Min
developer	287.00	846.12	1295.83	60.00
educational	39.00	1018.97	1705.46	62.00
entertainment	228.00	1088.82	1420.76	63.00
games	78.00	1210.19	1975.34	62.00
others	242.00	685.49	1201.74	60.00
social media	303.00	1440.42	2051.64	60.00

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



RESULTS AND DISCUSSION

- Dataset distribution from the histogram suggests that the data is right-skewed.
- Small number of visits have extremely high total_seconds duration (>6000)
- Majority of visits are concentrated at low total_seconds values (<2000)
- Note that outliers were not removed because it is important to track the total usage.





RESULTS AND DISCUSSION

Statistical analysis

KRUSKAL-WALLIS TEST RESULTS

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

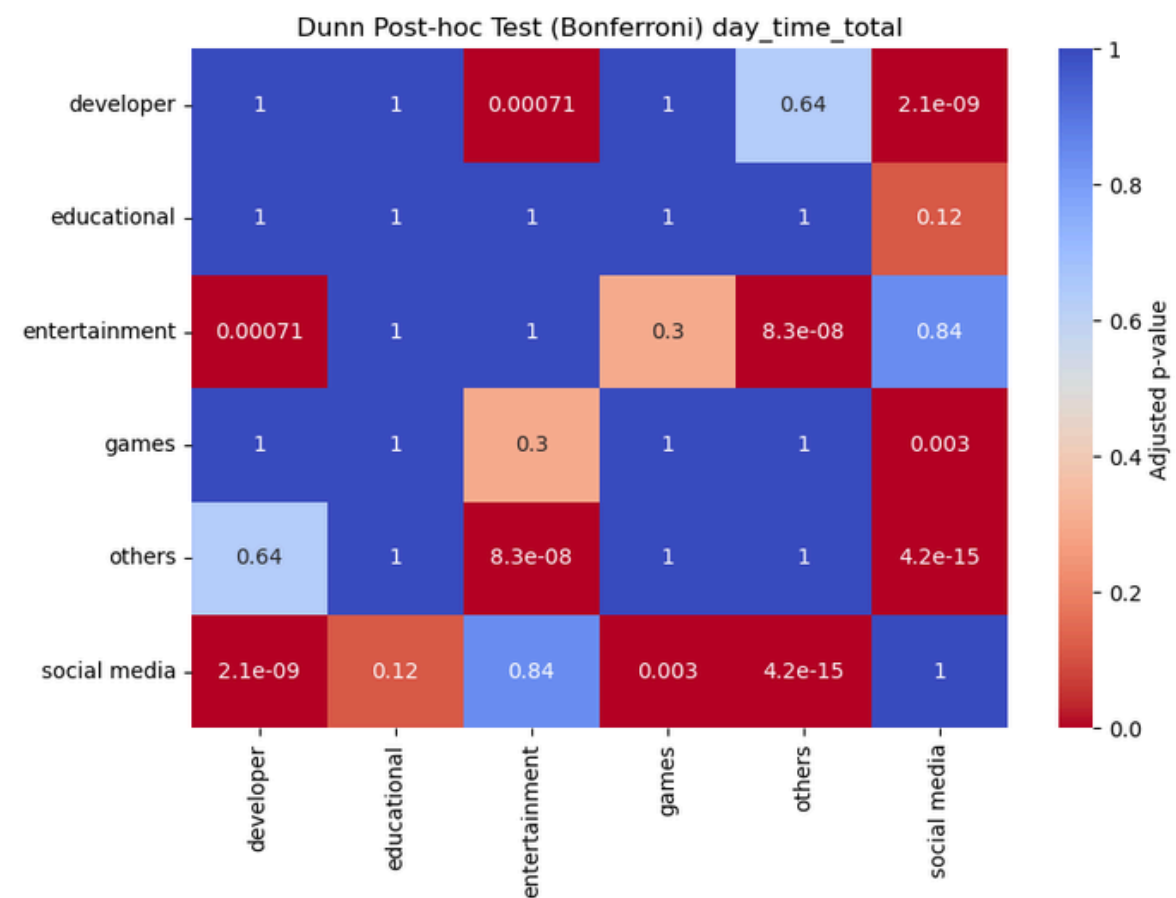
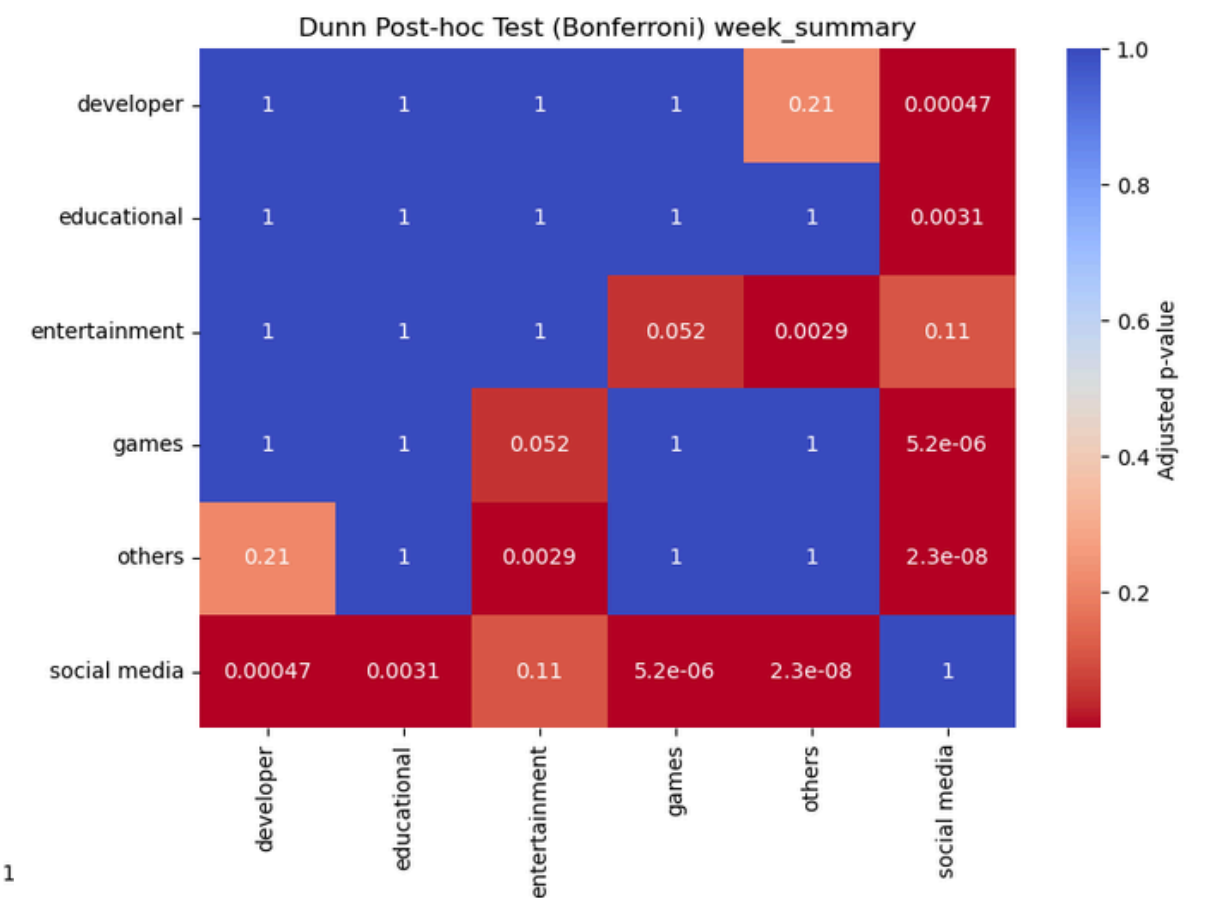
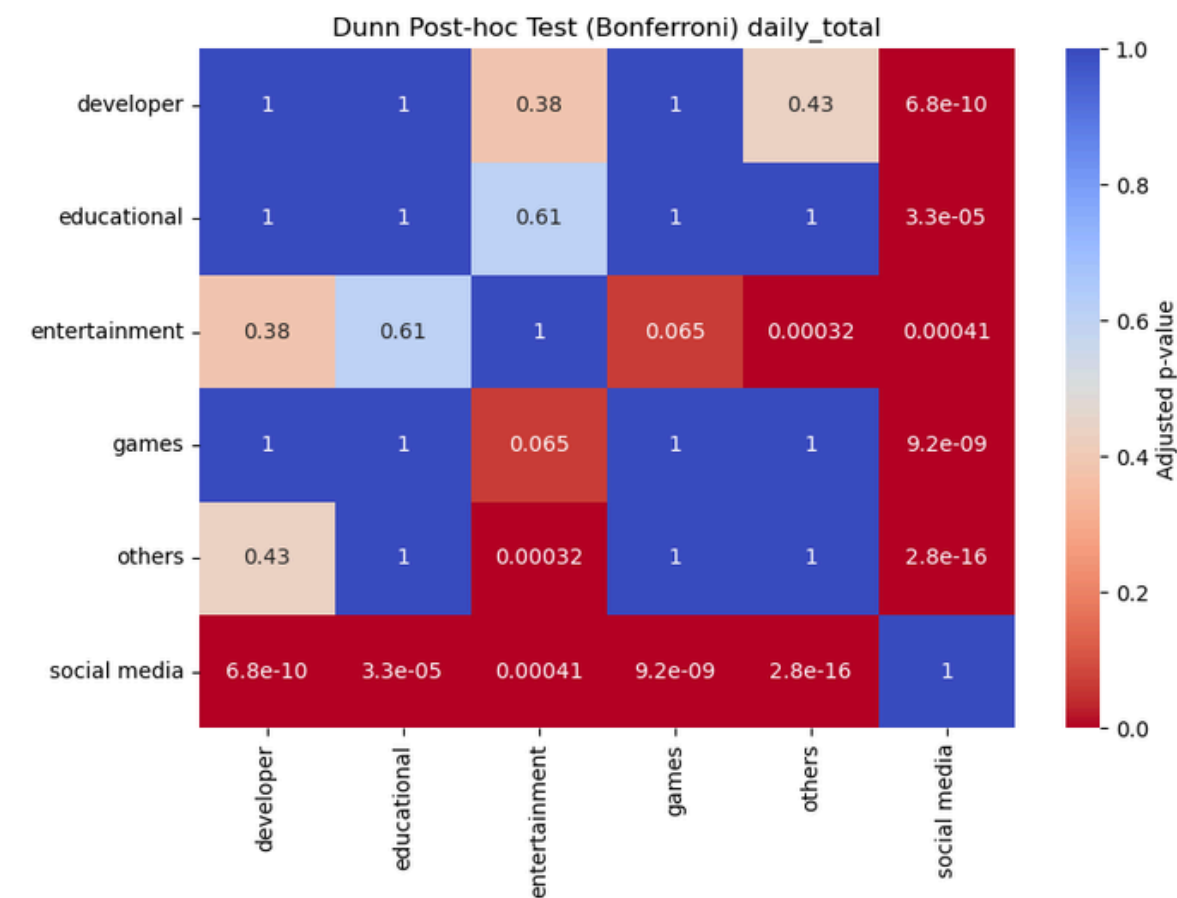
ANOVA TEST RESULTS

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

- For ANOVA, all p-values are less than 0.05, thus, the null hypothesis is rejected that There is no statistically significant difference in total website usage time across URL categories using time-of-day / daily / weekly aggregated totals.
- That is also the same for Kruskal Wallis Test that rejects the null hypothesis and proposing that there is at least one category that is significantly different from another category in terms of total usage.



RESULTS AND DISCUSSION





RESULTS AND DISCUSSION

Dunn Post-hoc Test

- Daily aggregation: Social media shows statistically significant differences with nearly all other categories (developer, educational, games, others). Entertainment also differs significantly from social media and others, indicating strong daily variation across categories.
- Time-of-day aggregation: Social media again differs significantly from developer, games, and others. Entertainment shows significant differences from developer and others, while educational does not exhibit widespread significant differences at this level.
- Weekly aggregation: Social media continues to differ significantly from developer, educational, games, and others. The “others” category also differs from entertainment.
- Trend across levels: The number of significant pairwise differences decreases at the weekly level compared to daily and time-of-day levels. Since spike/drops are not visible, it becomes less distinct.
- Overall implication: Social media consistently exhibits the most distinct time allocation pattern across all aggregation levels.



RESULTS AND DISCUSSION

Dunn Post-hoc Test

- 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 (e.g education vs developer) are different, thus failing to reject the null hypothesis. This is primarily due to 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 (social media & education → 0.12 p-value). Overlapping distributions between social media and educational usage reduce the separation in ranked values, resulting in a higher p-value and weaker statistical evidence of difference.



RESULTS AND DISCUSSION

Limitations

- The dataset represents a single participant ($n = 1$), limiting generalizability, only having a single individualized logs, and reflecting only individual browsing behavior.
- The observation period is relatively short, preventing analysis of long-term, seasonal, or workload and categorial related behavioral changes.
- Possible logging gaps, inactive background states, and manual URL category labeling may introduce recording errors or classification bias.
- Small sample sizes at higher aggregation levels (especially weekly) reduce statistical power, and temporal factors were not modeled in a multi-factor or interaction analysis.



RESULTS AND DISCUSSION

Conclusion

- Browsing time differs significantly across URL categories, with social media consistently recording the highest total usage across daily, time-of-day, and weekly aggregations.
- Usage is strongly concentrated in the noon, afternoon and evening (12:00–23:00), while early morning activity remains minimal.
- Social media and entertainment dominate overall screen time, outweighing productivity-related categories such as developer or educational websites.
- The results highlight the value of data-driven self-monitoring, enabling better time management, self-regulation, and improved balance between productivity and recreation.



**THAT'S ALL! THANK YOU FOR
LISTENING**