# Browsing History Analysis Report

Date:11-03-2025
Prepared by: Prachurya Phukan
Tools Used: Python (Pandas, Matplotlib, Seaborn)

## 1. Executive Summary

Objective: Analyze browsing history data to uncover patterns, derive insights, and present findings.

Key Insights:

- Communication & Social Engagement: Messaging and social media (e.g., Facebook) are among the most frequently visited platforms, highlighting strong engagement in online communication.
- Travel Research & Flight Bookings: Frequent visits to Google Flights and searches for flights to Tokyo indicate active travel planning.
- Work & Career Exploration: The user is engaged in tech-related activities, including AWS DynamoDB, job searches (QA roles), and business acquisition research (LoopNet).
- Browsing Patterns & Productivity Trends: Browsing activity is highest in the early morning (4-6 AM) and late evening (5-7 PM, 9-11 PM), with the lowest engagement occurring around 9 AM and mid-afternoon. Weekday activity is noticeably higher than weekends, suggesting work-related usage.
- Navigation Behavior: Most visits come from link-based transitions, meaning the user predominantly clicks on search results, embedded links, or bookmarks rather than typing URLs.

]

Approach: Data cleaning, trend analysis, visualization, and interpretation of key metrics.

## 2. Data Overview & Cleaning

Dataset Information:

Number of records: 5110
Columns present:

- Under the summary section: OrgId, ParticipantId, DeviceId, InstalledDate, AcceptanceDate, Extension, BrowsingCount, BookmarkCount, CookieCount
- Under the main data section: OrgId, ParticipantId, DeviceId, url, eventtimeutc, transition, title, visitId, referringVisitId, eventtime

Missing values handling:

Preprocessing Steps

I. Removed unnecessary rows (first 5 rows including summary data)
II. Converted timestamps to datetime format
III. Extracted domain names from URLs

✔ Identified and removed duplicate or erroneous entries

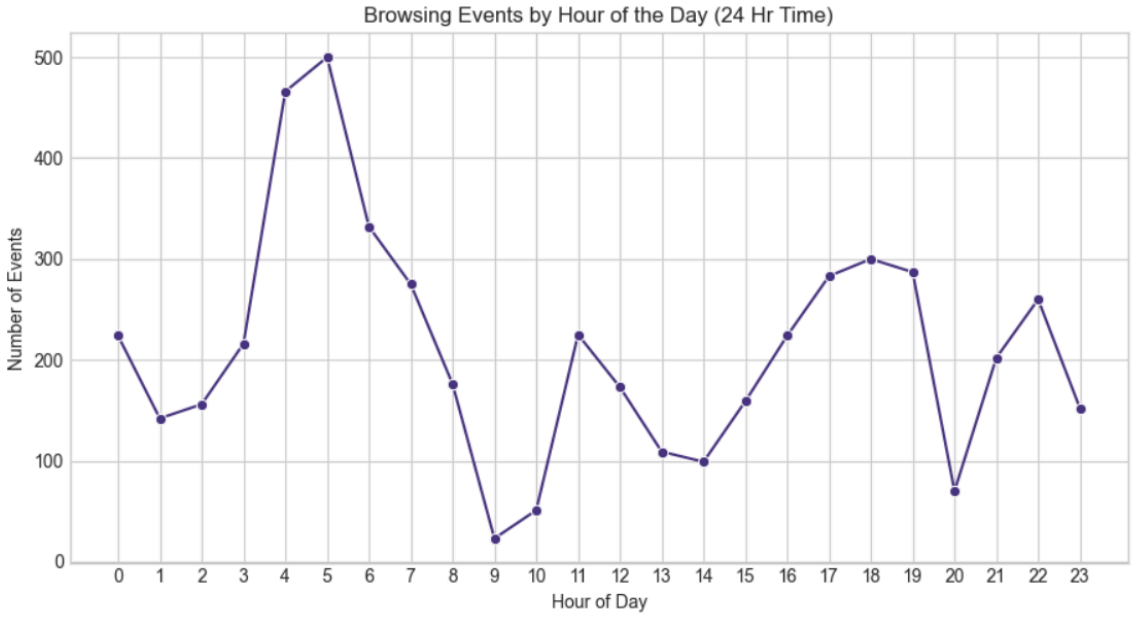
## 3. Exploratory Data Analysis (EDA) & Key Insights

This section uncovers trends and user behavior through visual analysis.

3.1 Browsing Trends Over Time

Q: When are users most active?

📊 Visualization: Line chart showing browsing events by hour of the day

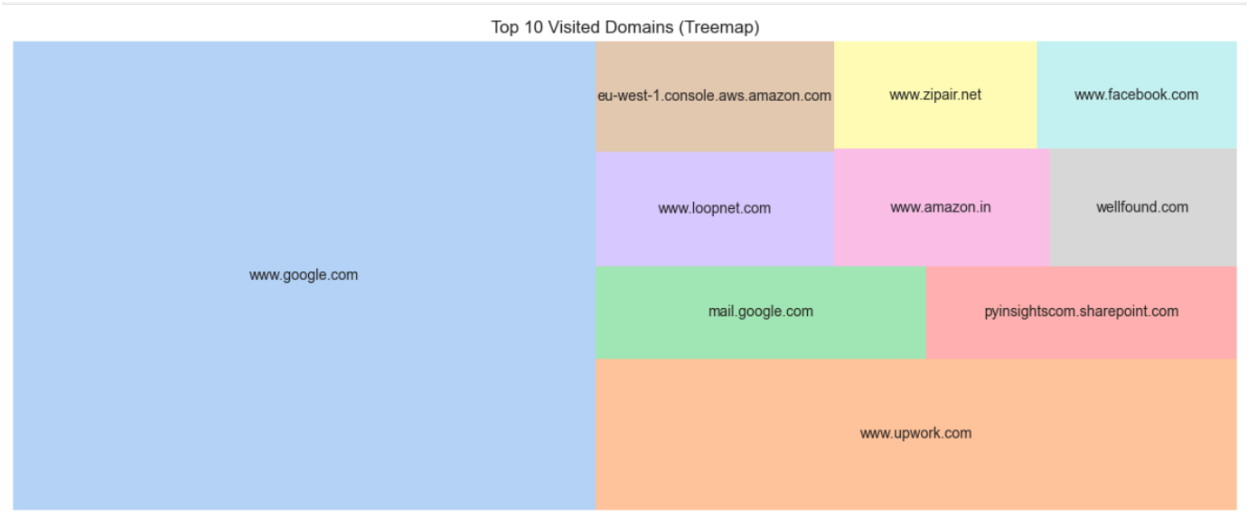



📈 Insight: Peak activity occurs at 5 AM, with the lowest activity noticed  at 9 AM. The user has an irregular browsing pattern, with spikes in activity during early mornings and late afternoons. There is a work-related rhythm, with high browsing before typical work hours and another surge after traditional working hours. The low afternoon activity suggests a structured break, possibly indicating an office-like schedule.

3.3 Website Engagement & Popular Domains

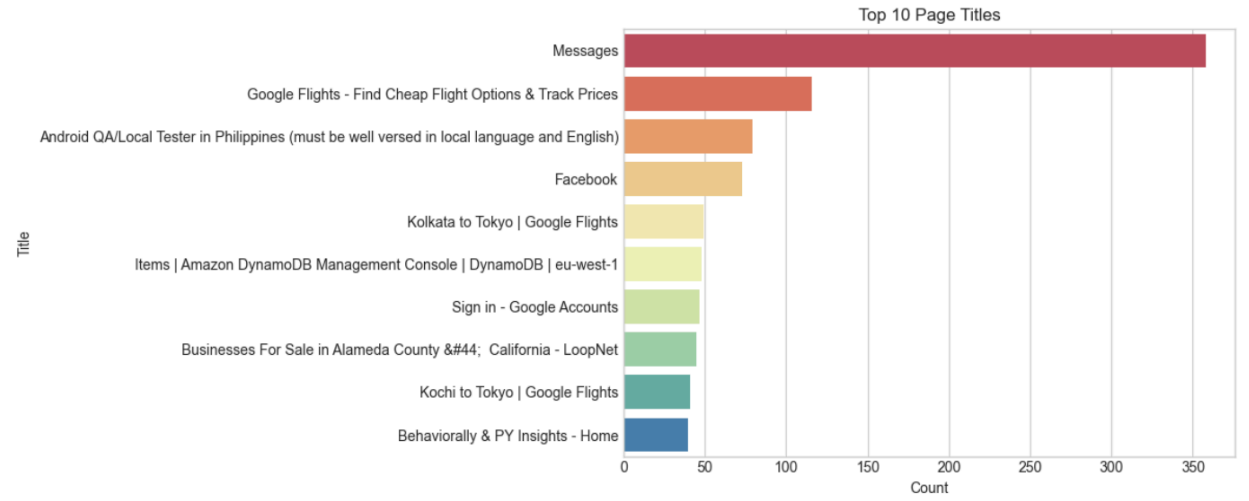
Q: Which websites are visited the most?

📊 Visualization: Tree map of Top 10 most visited domains




📈 Insight: Google and Upwork(Freelance website) dominate user activity. Many domains indicate work-related activity, suggesting a user focused on job platforms, cloud services, and professional communication.

📊 Visualization: Barplot of the top 10 page titles

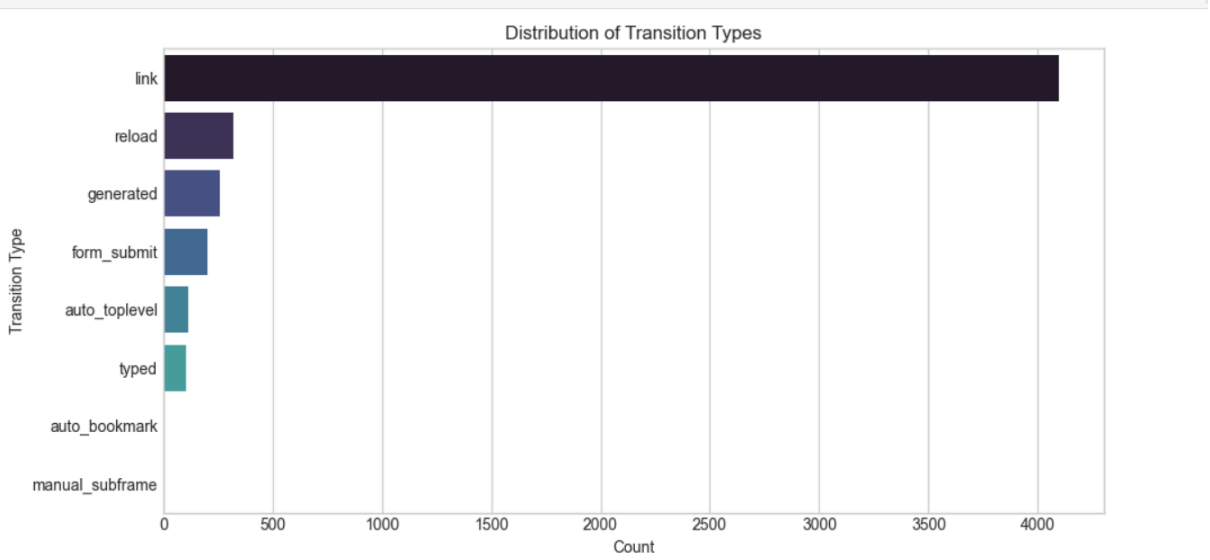



📈 Insight: Big majority of the pages were of the Messages type. Primary Activities of the user include messaging, travel planning, work-related tasks (AWS, job search), and social media. Secondary Interests include business acquisitions, real estate, and data analytics, suggesting a user engaged in tech-related work, travel, and business exploration, while maintaining active social and professional interactions.

3.3 User Navigation Patterns

Q: How do users move between pages?

📊 Visualization: Bar chart of transition types (typed URL, link click, form submission, etc.)

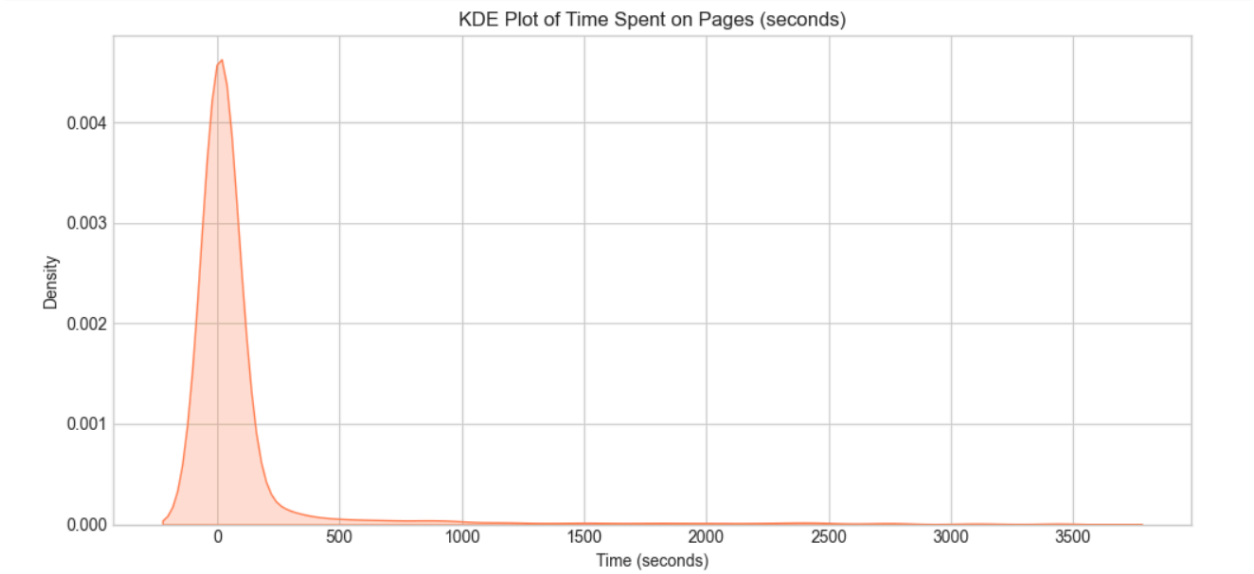



📈 Insight: Overwhelming majority of transitions were done through links. The small proportion of "typed" transitions suggests the user relies more on links and search rather than direct URL entry. Some transitions come from submitting forms, possibly login or data entry.

3.4 Time Spent on Pages

Q: How long do users spend on a page?

📊 Visualization: KDE plot of time spent per page visit

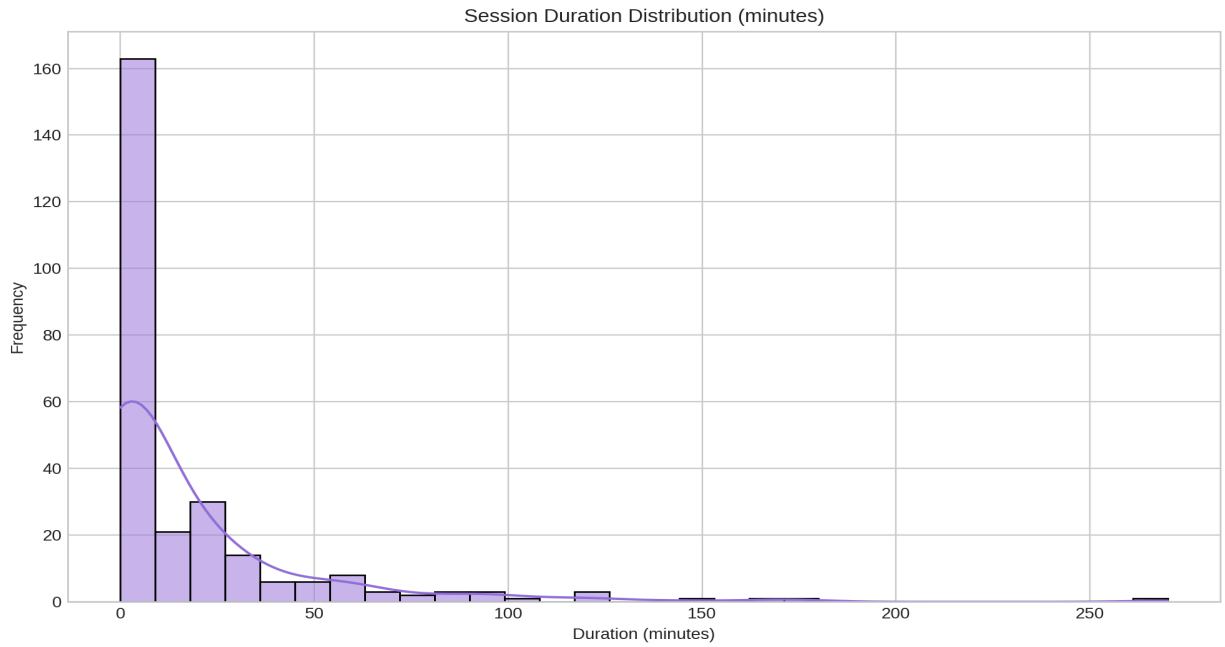



📈 Insight:The distribution is highly skewed to short bursts of time, showing that most users quickly navigate between pages. This could indicate information-seeking behavior rather than deep engagement.

3.5 Session Analysis & User Behavior

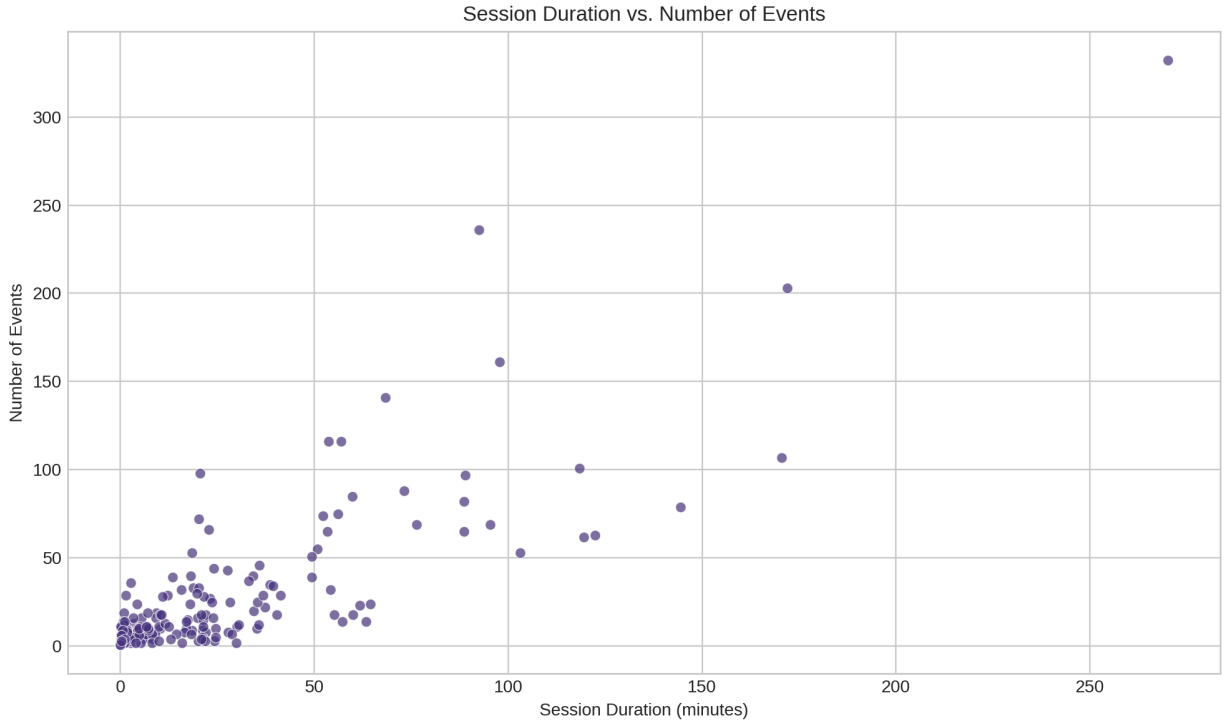
Q: How long are user sessions?

📊 Visualization: Histogram of session durations



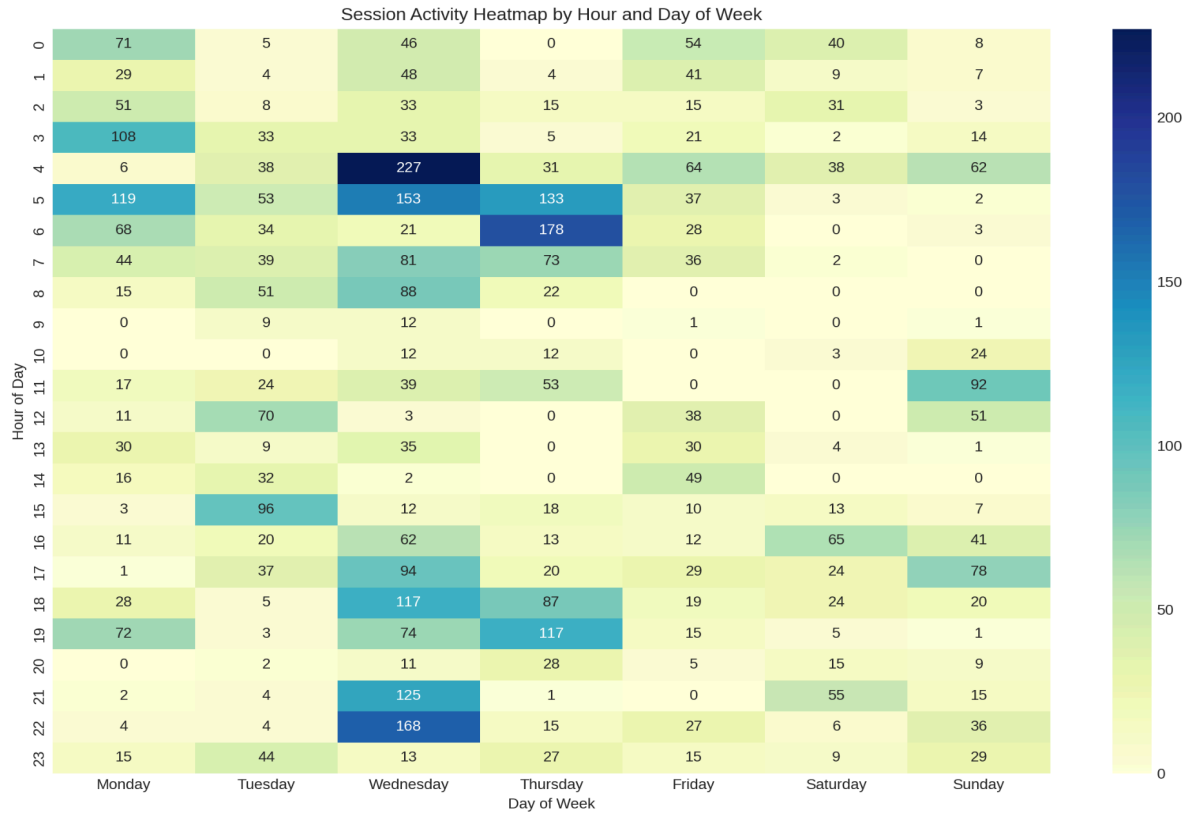

📈 Insight: Most users spend only a short time in a session, with the highest frequency occurring near 0-10 minutes, suggesting either quick interactions or potential issues (e.g., users leaving quickly). Investigating engagement strategies might be helpful.

📊 Visualization: Scatter plot of session duration vs. number of events per session



📈 Insight: A small core of users engages deeply, but most sessions are brief. Further analysis can determine whether these long sessions represent engaged users or technical issues (e.g., users leaving a tab open).

📊 Visualization: Heatmap of session activity by Hour and Day of the week




📈 Insight: Peak activity periods vary by day, but there is a notable spike around early morning and late evening. Some days have distinct high-activity hours (e.g., Wednesday at 5 AM has a large count). Activity during weekends is generally lower.

## 4. Key Insights drawn

4.1 High Engagement with Communication & Social Platforms
-The most visited page was "Messages," showing a strong emphasis on communication.
-Facebook was among the top sites, indicating social media engagement.
Implication: The user likely prioritizes messaging, communication, and networking.

4.2 Strong Interest in Travel & Flight Bookings
-Multiple visits to Google Flights, including specific searches for Tokyo (Japan) flights.
-The user appears to be actively researching and planning a trip.
Implication: There’s a potential upcoming trip, and the user is looking for the best deals on flights.

4.3  Work-Related & Tech Engagement
-The presence of AWS DynamoDB Console suggests cloud-related work or database management.
J-ob searches (Android QA/Local Tester) suggest career-related activities.
-Browsing LoopNet for businesses for sale suggests possible entrepreneurial interest.
Implication: The user is engaged in tech work, career exploration, and possibly business investments.

4.4 Browsing Patterns Indicate a Unique Productivity Schedule
-Peak browsing hours:
    Early morning (4-6 AM) – Highest activity
    Late afternoon to evening (5-7 PM, 9-11 PM) – Moderate activity
-Drop in browsing around mid-morning and early afternoon, possibly indicating focus on other tasks (e.g., work, meetings).
Implication: The user operates on an unusual schedule, possibly balancing multiple responsibilities.

4.5 Dominance of “Link” Transition Type
-The majority of browsing actions were clicking on links, rather than manually typing URLs.
-Other common transitions: Reload, Generated, Form Submit, suggesting frequent site revisits.
Implication: The user navigates primarily through clicking links (bookmarks, search results, embedded links) rather than directly entering URLs.

## 5. Actionable Recommendations:

5.1 For Personal Productivity & Efficiency

Optimize Communication Time:

- Since messaging is the top activity, consider batch processing messages at set times to avoid constant interruptions.

Streamline Flight Booking:

- Set price alerts for the flights being researched to receive notifications when ticket prices drop.

Use Bookmarks & Shortcuts More Effectively:

- Since most browsing happens via clicking links, organizing frequently visited sites into browser bookmarks or shortcut folders could save time.

5.2 For Career & Work Optimization

Explore Job Opportunities in Tech & Remote Work:

- Given the job searches (QA roles, AWS), consider setting up job alerts on platforms like LinkedIn, Wellfound (AngelList), and Upwork.

Deepen AWS & Cloud Computing Skills:

- If AWS services (DynamoDB) are frequently used, consider certification courses to enhance expertise and career growth.

Research Business & Investment Further:

- If exploring business acquisitions (LoopNet), look into market trends, financing options, and business valuation strategies.

5.3 For Browsing & Digital Well-being

Leverage Browser Extensions for Efficiency:

- Extensions like Pocket (for saving articles), OneTab (for managing tabs), and Productivity Blockers can improve time management.

Analyze and Adjust Browsing Habits:

- Since peak browsing occurs in early mornings and evenings, consider whether this aligns with optimal productivity or if adjustments are needed.

Use Incognito Mode for Job & Business Research:

- To prevent tracking & personalized ads while researching career or business options, use incognito/private browsing mode.

## 6. Areas for improvement:

- Implement personalized recommendations based on browsing behavior.
- Optimize website performance for peak browsing hours.
- Conduct A/B testing to increase engagement.