PY Insights (Power of You) Data Analyst Pre-Task Description

Browsing History Analysis & Storytelling

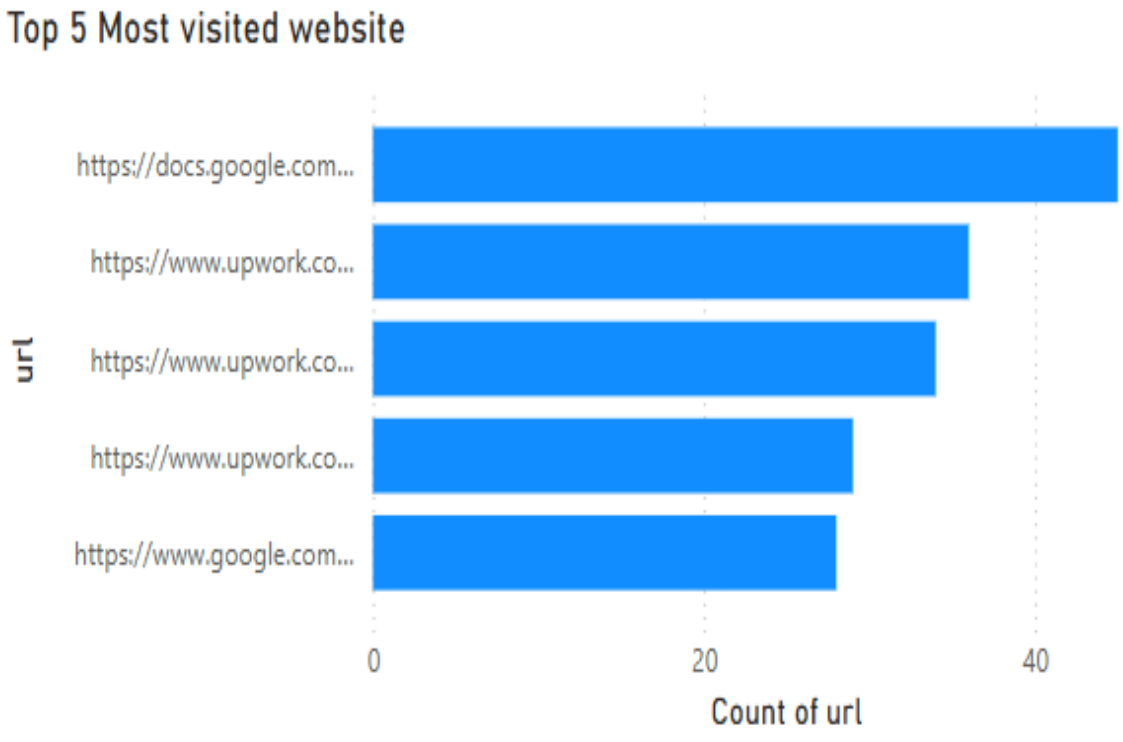
Objectives:

Extract the insights in Browsing history data set. Through Python as well as Powerbi

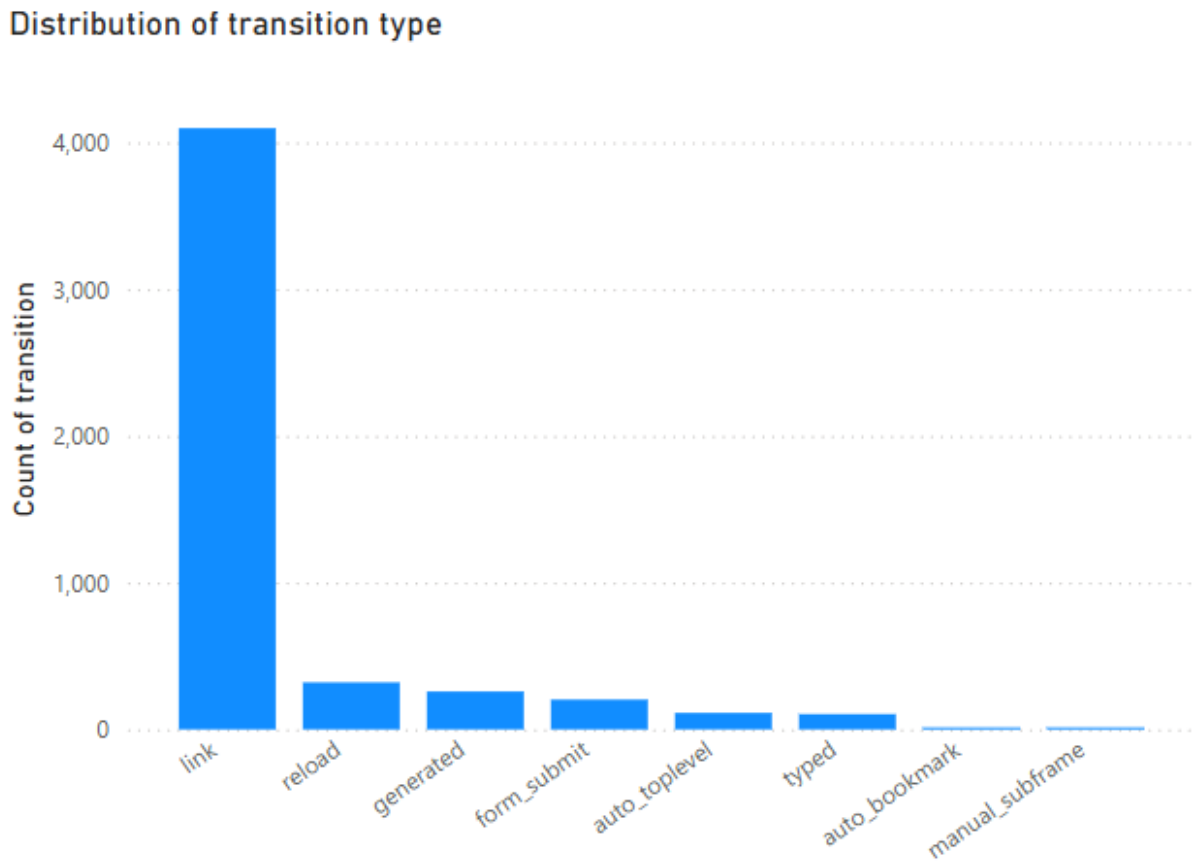
Powerbi (Visualization)

We finds five insight from datasets:

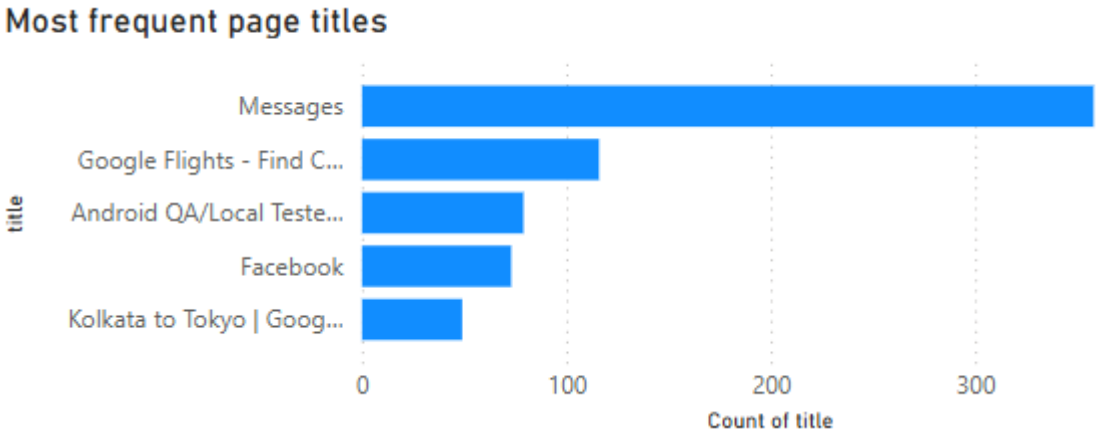
1. Top 5 most visited websites.



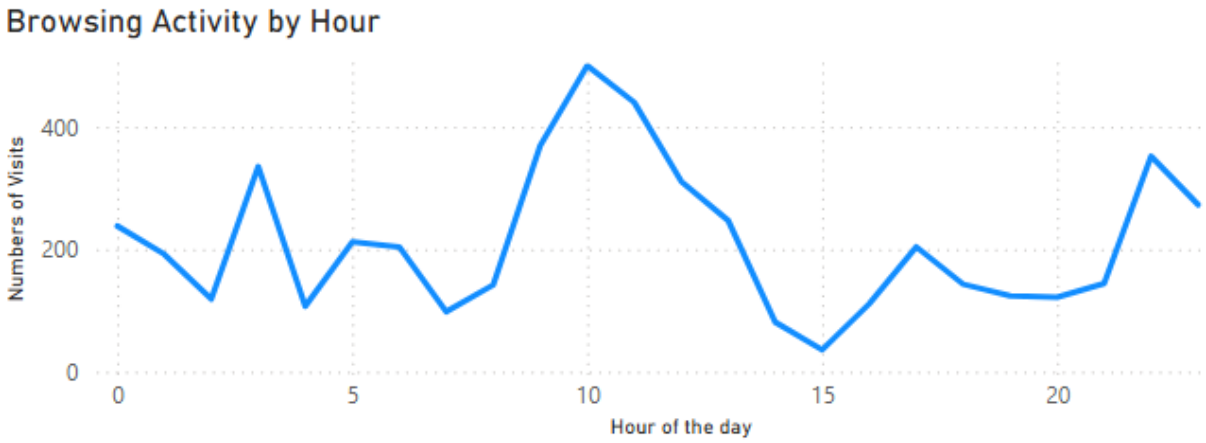
2. Distribution of transition type



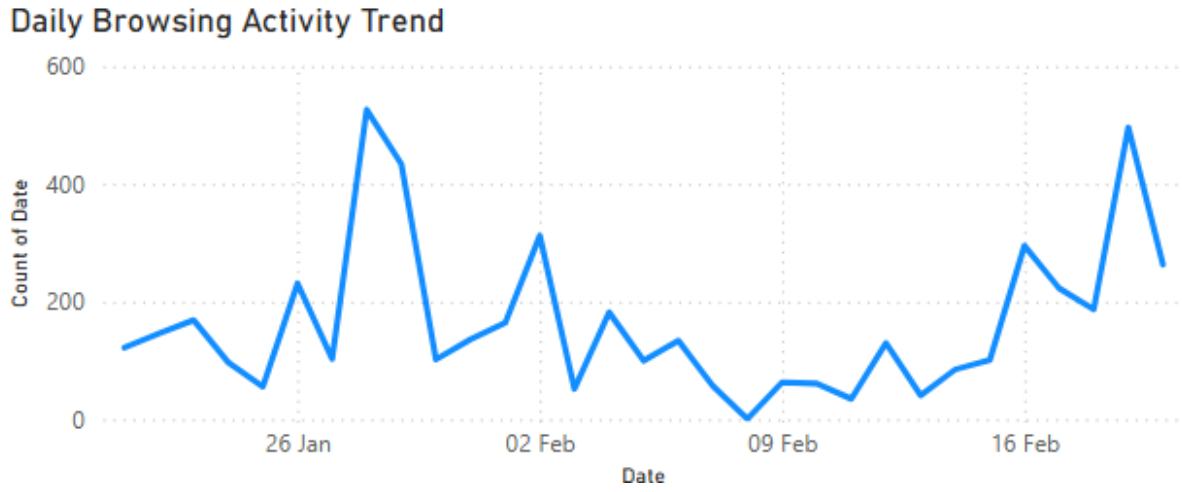
3. Most frequent page Titles



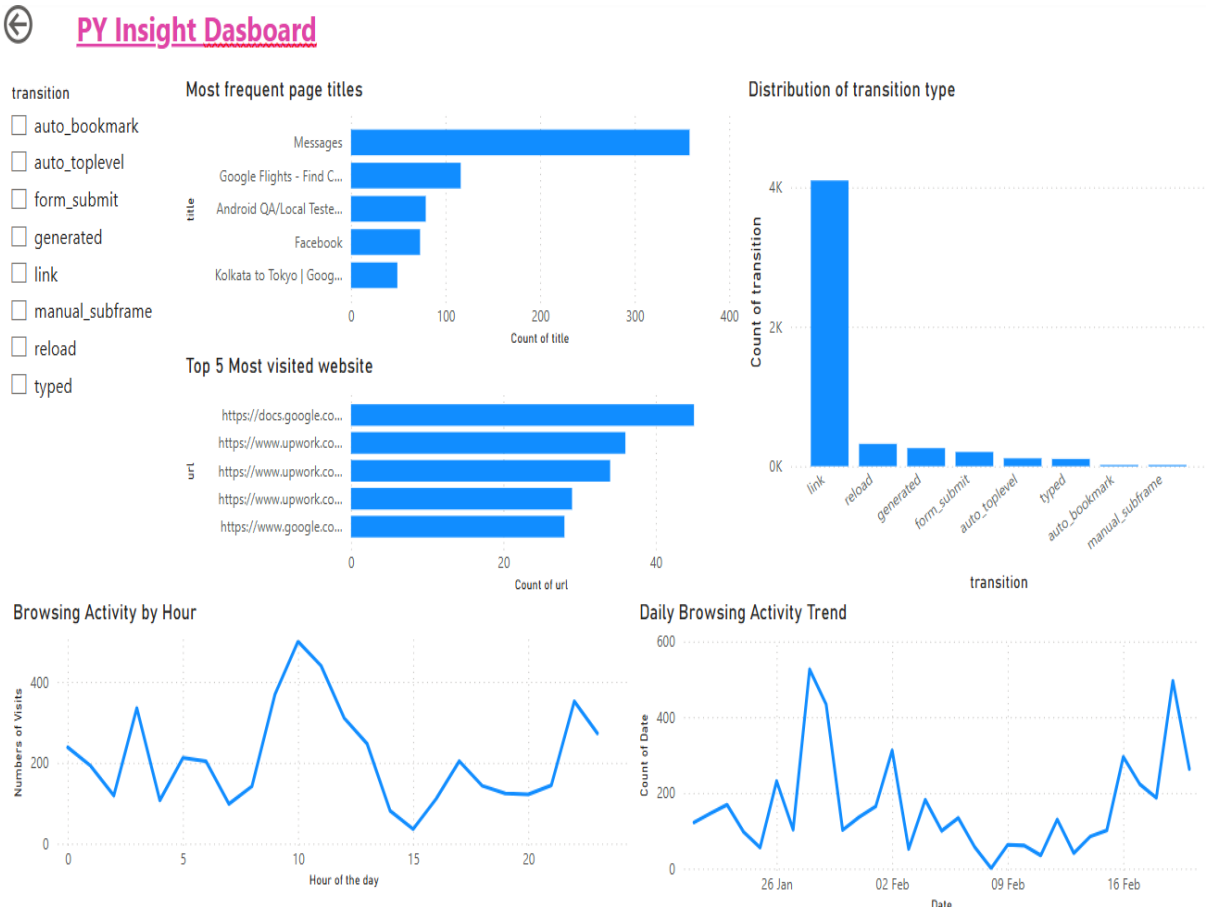
4. Browsing Activity by Hour



5. Daily Browsing Activity Trends



PY Insight Dashboard



Python
(code, insights and Visualization)

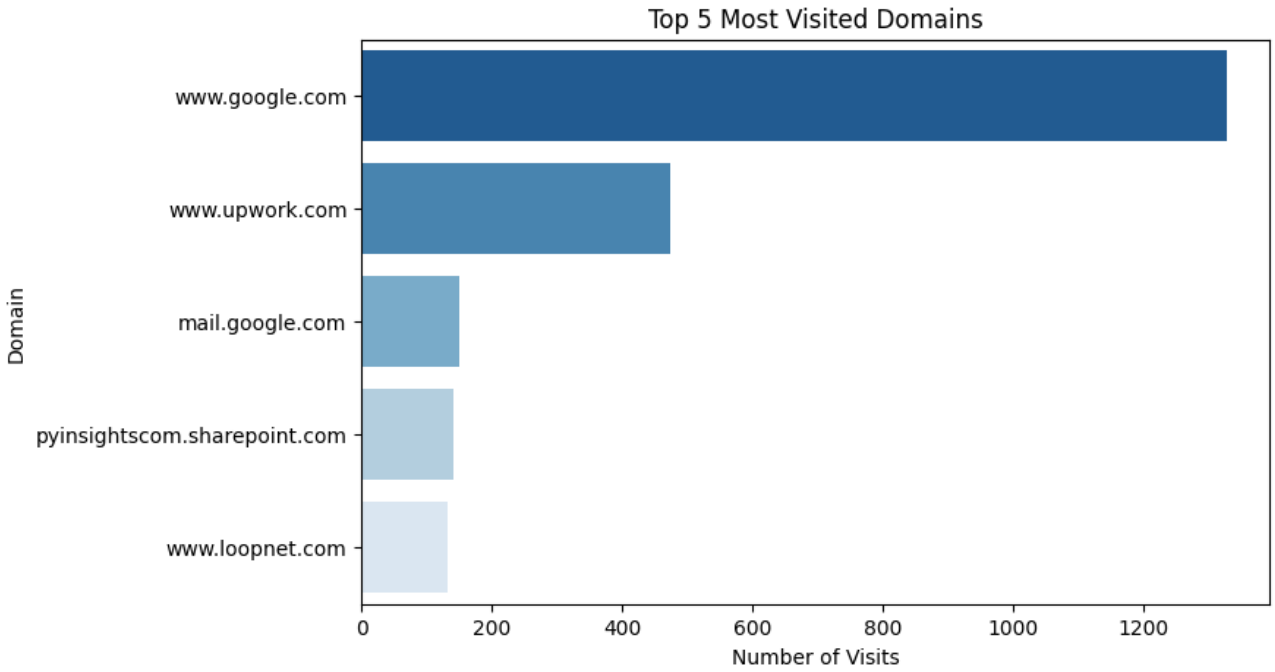
 # import and read excel File.

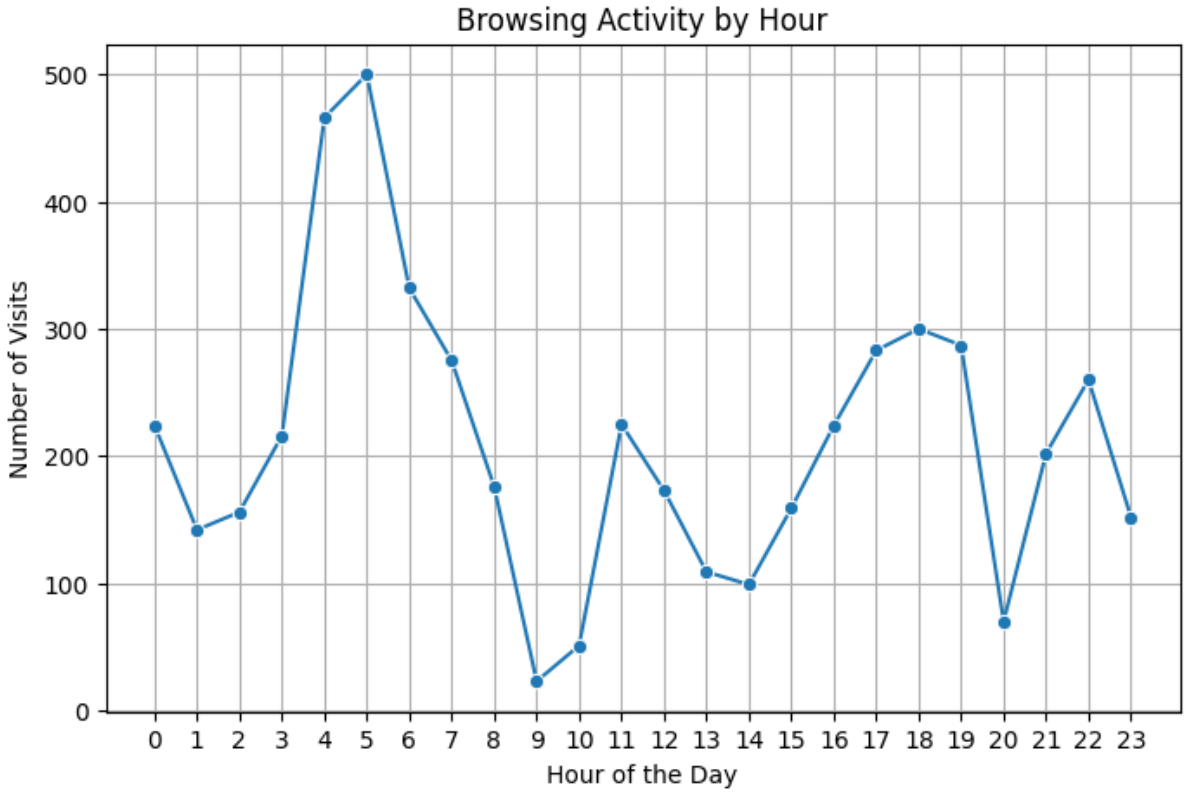
```
import pandas as pd
from google.colab import drive
drive.mount('/content/drive')
df = pd.read_excel('/content/drive/MyDrive/py/py_demo.xlsx')
print(df)
```

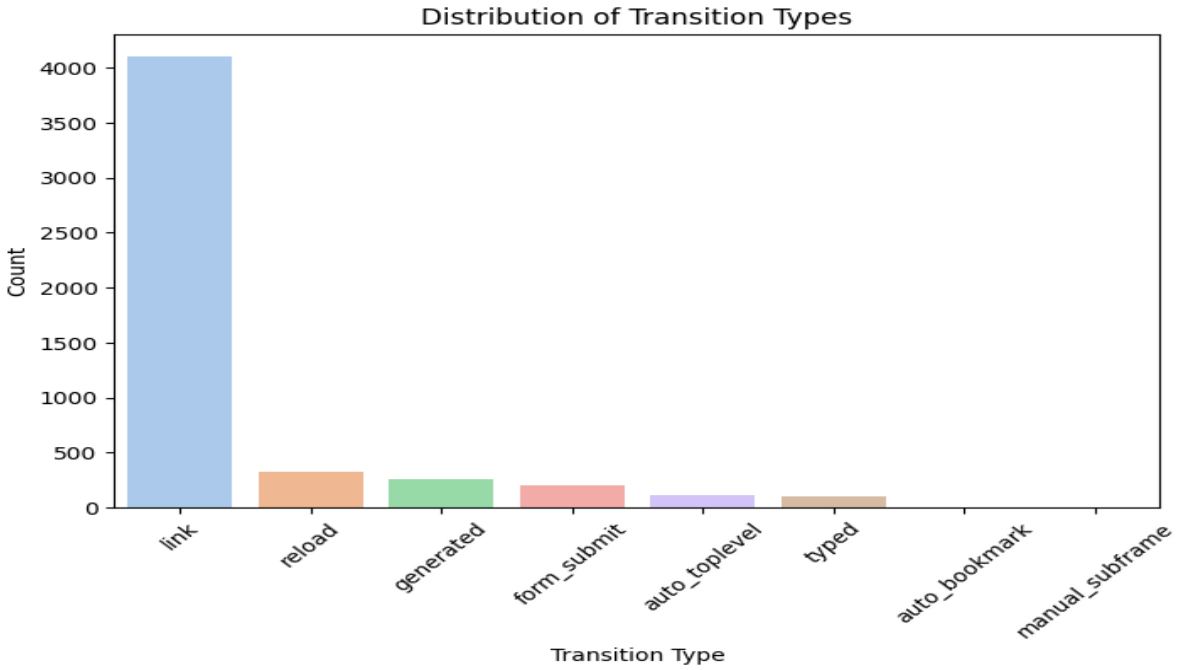
```
# For insights and Visualization
# we import matplotlib and seaborn library

import matplotlib.pyplot as plt
import seaborn as sns

# Convert eventtimeutc to datetime for analysis
df['eventtimeutc'] = pd.to_datetime(df['eventtimeutc'])

# Extract date and hour for time-based analysis
df['date'] = df['eventtimeutc'].dt.date
df['hour'] = df['eventtimeutc'].dt.hour

# Insight 1: Most visited websites (Top domains)
df['domain'] = df['url'].apply(lambda x: x.split('/')[2] if '//' in x else
x)
top_domains = df['domain'].value_counts().head(5)

# Insight 2: Peak browsing hours
peak_hours = df['hour'].value_counts().sort_index()
```

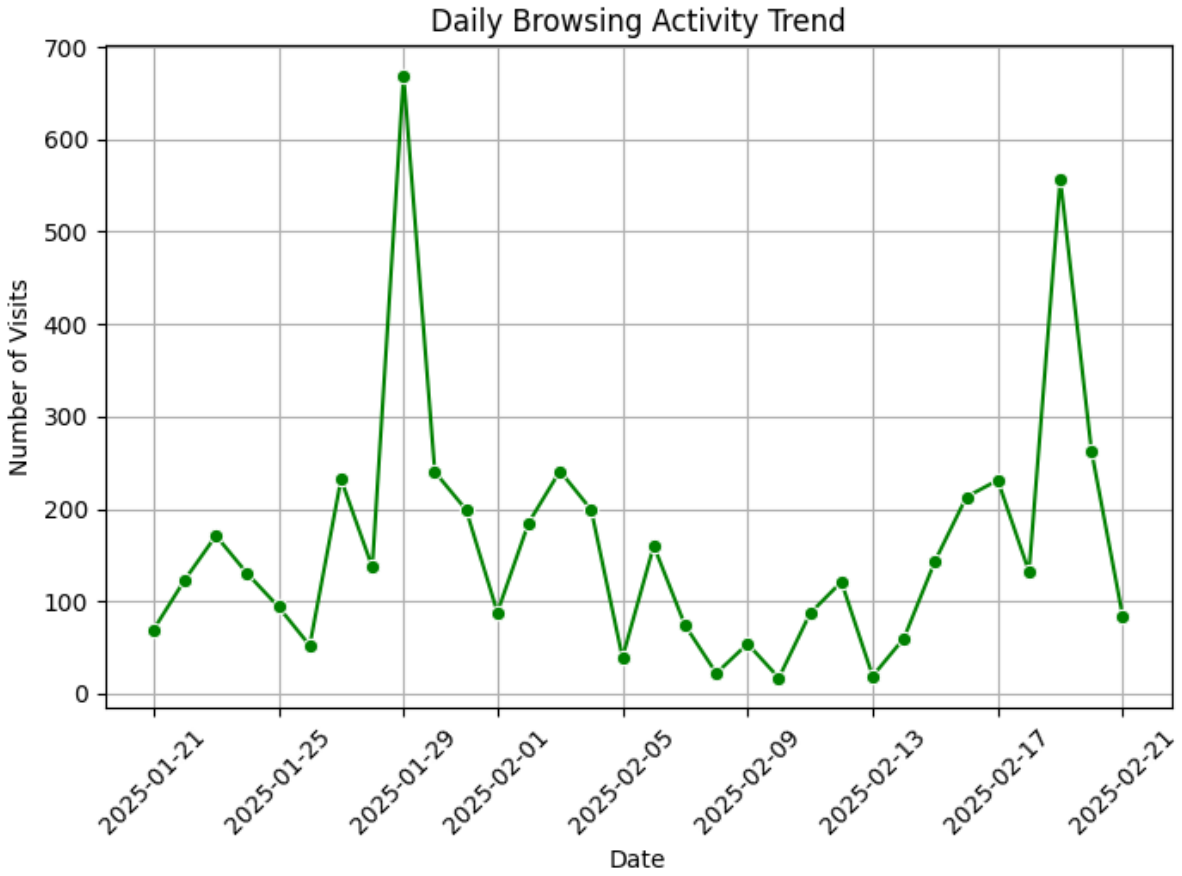
```
# Insight 3: Most common transition types (link, reload, typed, etc.)
transition_counts = df['transition'].value_counts()

# Insight 4: Daily browsing activity trend
daily_activity = df.groupby('date').size()

# Insight 5: Most frequent page titles
top_titles = df['title'].value_counts().head(5)

# Visualization 1: Top visited domains
plt.figure(figsize=(8, 5))
sns.barplot(x=top_domains.values, y=top_domains.index,
palette="Blues_r")
plt.xlabel("Number of Visits")
plt.ylabel("Domain")
plt.title("Top 5 Most Visited Domains")
plt.show()

# Visualization 2: Browsing activity by hour
plt.figure(figsize=(8, 5))
sns.lineplot(x=peak_hours.index, y=peak_hours.values,
marker="o")
plt.xlabel("Hour of the Day")
plt.ylabel("Number of Visits")
plt.title("Browsing Activity by Hour")
plt.xticks(range(0, 24))
plt.grid()
plt.show()

# Visualization 3: Transition types distribution
plt.figure(figsize=(8, 5))
sns.barplot(x=transition_counts.index, y=transition_counts.values,
palette="pastel")
plt.xlabel("Transition Type")
plt.ylabel("Count")
```

```
plt.title("Distribution of Transition Types")
plt.xticks(rotation=45)
plt.show()

# Visualization 4: Daily browsing activity
plt.figure(figsize=(8, 5))
sns.lineplot(x=daily_activity.index, y=daily_activity.values,
marker="o", color="g")
plt.xlabel("Date")
plt.ylabel("Number of Visits")
plt.title("Daily Browsing Activity Trend")
plt.xticks(rotation=45)
plt.grid()
plt.show()

# Visualization 5: Most frequent page titles
plt.figure(figsize=(8, 5))
sns.barplot(x=top_titles.values, y=top_titles.index,
palette="muted")
plt.xlabel("Number of Visits")
plt.ylabel("Page Title")
plt.title("Top 5 Most Frequent Page Titles")
plt.show()

# Return insights
top_domains, peak_hours.head(), transition_counts,
daily_activity.head(), top_titles
```


Top 5 Most Visited Domains
www.google.com
www.upwork.com
mail.google.com
pyinsightscom.sharepoint.com
www.loopnet.com
Domain
0
200
400
600
800
1000
1200
Number of Visits


Browsing Activity by Hour
Number of Visits
0
100
200
300
400
500
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of the Day


Distribution of Transition Types
Count
4000
3500
3000
2500
2000
1500
1000
500
0
link
reload
generated
form_submit
auto_toplevel
typed
auto_bookmark
manual_subframe
Transition Type


Daily Browsing Activity Trend
Number of Visits
700
600
500
400
300
200
100
0
2025-01-21
2025-01-25
2025-01-29
2025-02-01
2025-02-05
2025-02-09
2025-02-13
2025-02-17
2025-02-21
Date