ENTERTAINMENT CONSUMPTION

DSA 210 INTRODUCTION TO DATA SCIENCE
TERM PROJECT REPORT

Nur İdil MÜFTÜOĞLU
34292

January 2025

01 MOTIVATION

MOTIVATION

The motivation for this project is to explore **how entertainment consumption patterns reflect time management, stress and preferences across different life stages**. Analyzing my **Letterboxd, TV Time and Instagram data** provides insights into my behavior during three distinct phases: working full-time, studying for university entrance exams while working and being a full-time student. The project aims to highlight the **link between time constraints and entertainment activities.**

02

MAIN RESEARCH QUESTIONS AND HYPOTHESES

MAIN RESEARCH QUESTIONS AND HYPOTHESES

01

Content Type Preferences

- **Research Question:** How do my preferences for movies and social media content vary across different life stages?
  - **Hypothesis:** During busier life phases (exam preparation), I will consume shorter content (Instagram posts) more frequently, while longer content (movies) will be consumed during less busy periods (university and full-time work).
- **Research Question:** Is there a shift towards shorter or more easily consumable content during busier periods?
  - **Hypothesis:** The average duration of entertainment content consumed will decrease during high-stress periods (exam preparation) compared to low-stress periods (university and full-time work).

02

Stress Correlation

- **Research Question:** How does Instagram usage correlate with periods of high stress, such as exam preparations?
  - **Hypothesis:** Instagram usage will increase significantly during periods of high stress.

03

DATA SOURCES

DATA SOURCES

Entertainment data is exported directly from Letterboxd and Instagram as a personal data request.

INSTAGRAM

Instagram data is used to represent short-time content for entertainment.

Instagram likes data were loaded from the JSON file to be processes. The data structure was nested, so the nested structure was flatten to extract only title and timestamp. Then Unix timestamps were converted to readable datetime format.

LETTERBOXD

Letterboxd data is used to represent long-time content for entertainment.

Movie data were loaded form the .csv file to be processed. Watch date timestamps were converted to readable datetime format.

INITIAL SETUP AND DATA LOADING

Initially all required libraries for data analysis and visualization were imported. This includes pandas for data manipulation, matplotlib and seaborn for visualization, and various other utilities needed for data processing.

Then, Google Drive was mounted and paths for both Instagram and Letterboxd data files were defined to access the datafiles.

```
import pandas as pd
import json
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import drive
import requests
from bs4 import BeautifulSoup
import time
import random
import re
from urllib.parse import urlparse
import calendar
from scipy import stats

drive.mount('/content/drive/')
```

```
base_directory = "/content/drive/MyDrive/meta-2024-Nov-25-11-52-50"
instagram_file_path = f'{base_directory}/instagram-gercekten.idil-2024-11-25-xmCJFTlJ/your_instagram_activity/likes/liked_posts.json'
letterboxd_file_path = f'{base_directory}/letterboxd-idilmuftuoglu-2024-12-11-15-57-utc/diary.csv'
```

04

DATA ANALYSIS AND OUTCOMES

DEFINE LIFE PERIODS

A function was created to categorize timestamps into distinct periods of my life. Undefined periods were then removed from the data.

#	Name	Dates
1	1. University	December 2018 - June 2020
2	Full-time Work	October 2020 - September 2022
3	Exam Preparation	October 2022 - June 2023
4	2. University	October 2023 - Novermber 2024

ANALYZE DAILY INSTAGRAM ACTIVITY

A comprehensive analysis function was created that:

1. Processes daily Instagram likes
2. Creates a complete timeline for each life period
3. Fills in missing days with zero counts
4. Calculates summary statistics for each period including:
   - Mean, median, min, and max likes per day
   - Total number of likes
   - Standard deviation
   - Date range and total days in period

Summary statistics by period:

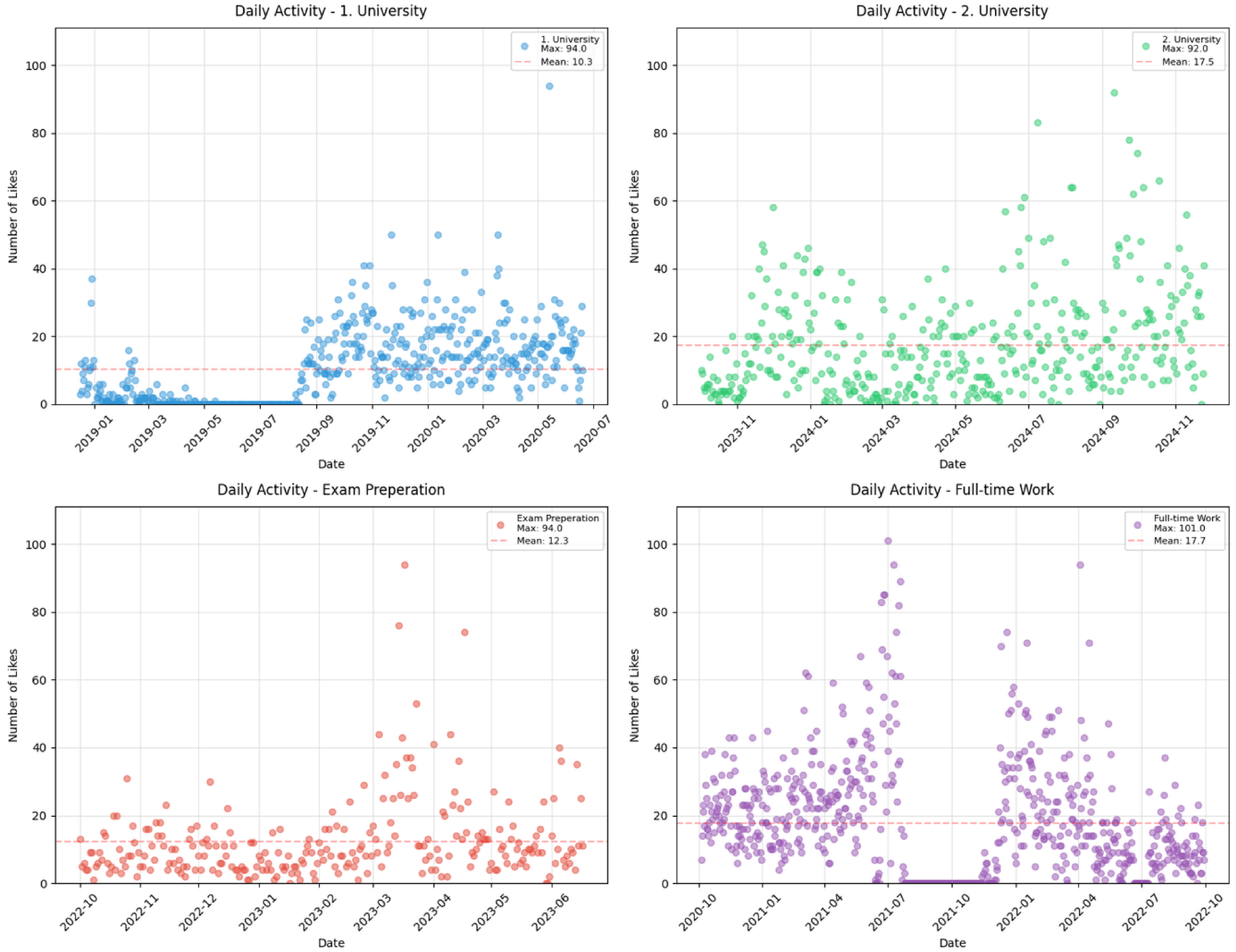
	count						date		
	mean	median	min	max	sum	std	min	max	count
custom_period									
1. University	10.30	9.0	0.0	94.0	5686.0	10.82	2018-12-16	2020-06-19	552
2. University	17.52	14.0	0.0	92.0	7374.0	15.19	2023-10-02	2024-11-25	421
Exam Preperation	12.28	9.0	0.0	94.0	3192.0	11.91	2022-10-01	2023-06-17	260
Full-time Work	17.67	14.0	0.0	101.0	12811.0	17.12	2020-10-05	2022-09-29	725

VISUALIZE DAILY INSTAGRAM ACTIVITY

Scatter plots of daily Instagram activity for each life period were created.

Standardized y-axis scale was used across the plots for easy comparison between periods.

Mean activity line was shown as a dashed red line for better visualization.



VISUALIZE CLEANED INSTAGRAM ACTIVITY

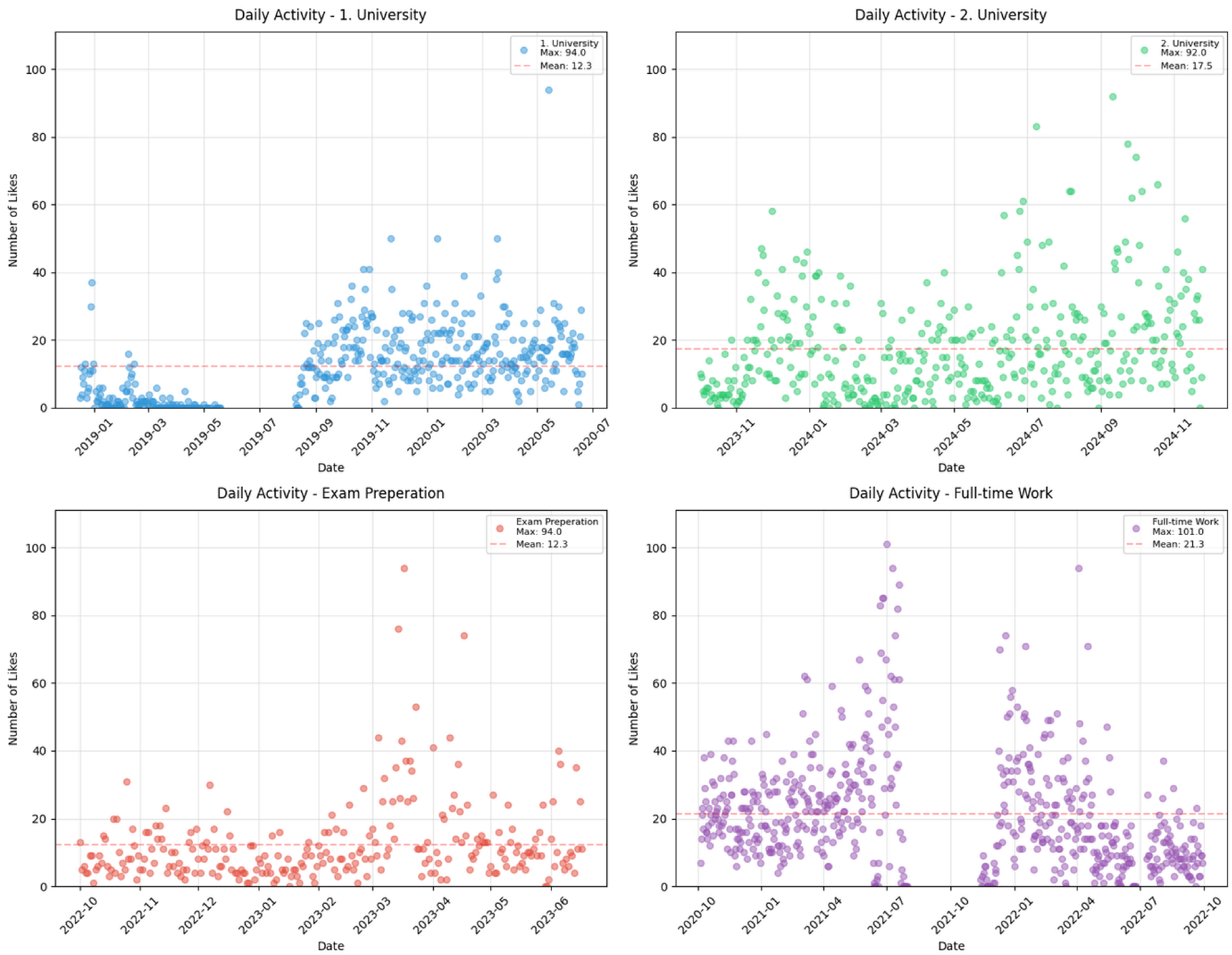
Previous analysis revealed some outlier sequences where I deliberately did not use Instagram. These periods are not normal and should be removed.

```
Removing outlier sequence from 1. University:
Removed 89 days
Date range: 2019-04-28 to 2019-08-08

Removing outlier sequence from Full-time Work:
Removed 123 days
Date range: 2021-07-31 to 2022-07-10

Before cleaning:
                  count       mean  min    max
custom_period
1. University       552  10.300725  0.0   94.0
2. University       421  17.515439  0.0   92.0
Exam Preperation    260  12.276923  0.0   94.0
Full-time Work      725  17.670345  0.0  101.0

After cleaning:
                  count       mean  min    max
custom_period
1. University       463  12.280778  0.0   94.0
2. University       421  17.515439  0.0   92.0
Exam Preperation    260  12.276923  0.0   94.0
Full-time Work      602  21.280731  0.0  101.0
```

VISUALIZE CLEANED INSTAGRAM ACTIVITY

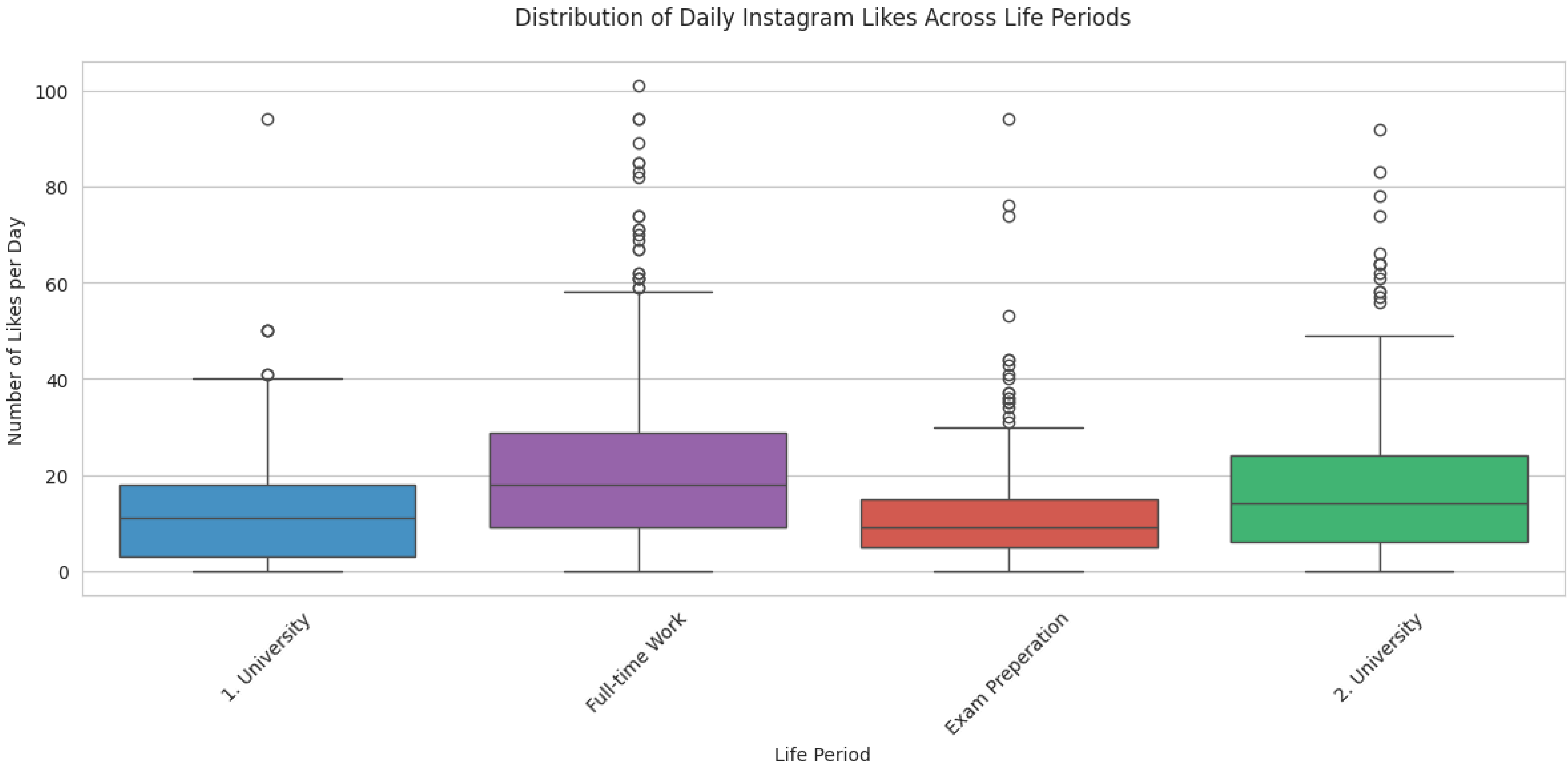
Then scatter plots using the cleaned data that excludes outlier sequences (greater than or equal to 7 days of inactivity) were created.

This provided a clearer view of typical Instagram usage patterns.



VISUALIZE CLEANED INSTAGRAM ACTIVITY

In order to compare all periods side-by-side boxplots were created next.



Detailed Statistics by Period:

custom_period	count	mean	std	min	25%	50%	75%	max
1. University	463.0	12.280778	10.736866	0.0	3.0	11.0	18.00	94.0
2. University	421.0	17.515439	15.191723	0.0	6.0	14.0	24.00	92.0
Exam Preperation	260.0	12.276923	11.906661	0.0	5.0	9.0	15.00	94.0
Full-time Work	602.0	21.280731	16.613626	0.0	9.0	18.0	28.75	101.0

ANALYZE LETTERBOXD MOVIE DATA

In order to understand how I engage with longer-duration entertainment content my movie-watching habits were analyzed. By comparing two types of entertainment consumption (quick Instagram posts vs. full-length movies), the aim is to understand:

- How my entertainment preferences changed during different life periods
- Whether stress or free time affected my choice between short vs. long-form content
- If there are any patterns in when I prefer to watch movies vs. browse Instagram

Summary Statistics by Period:

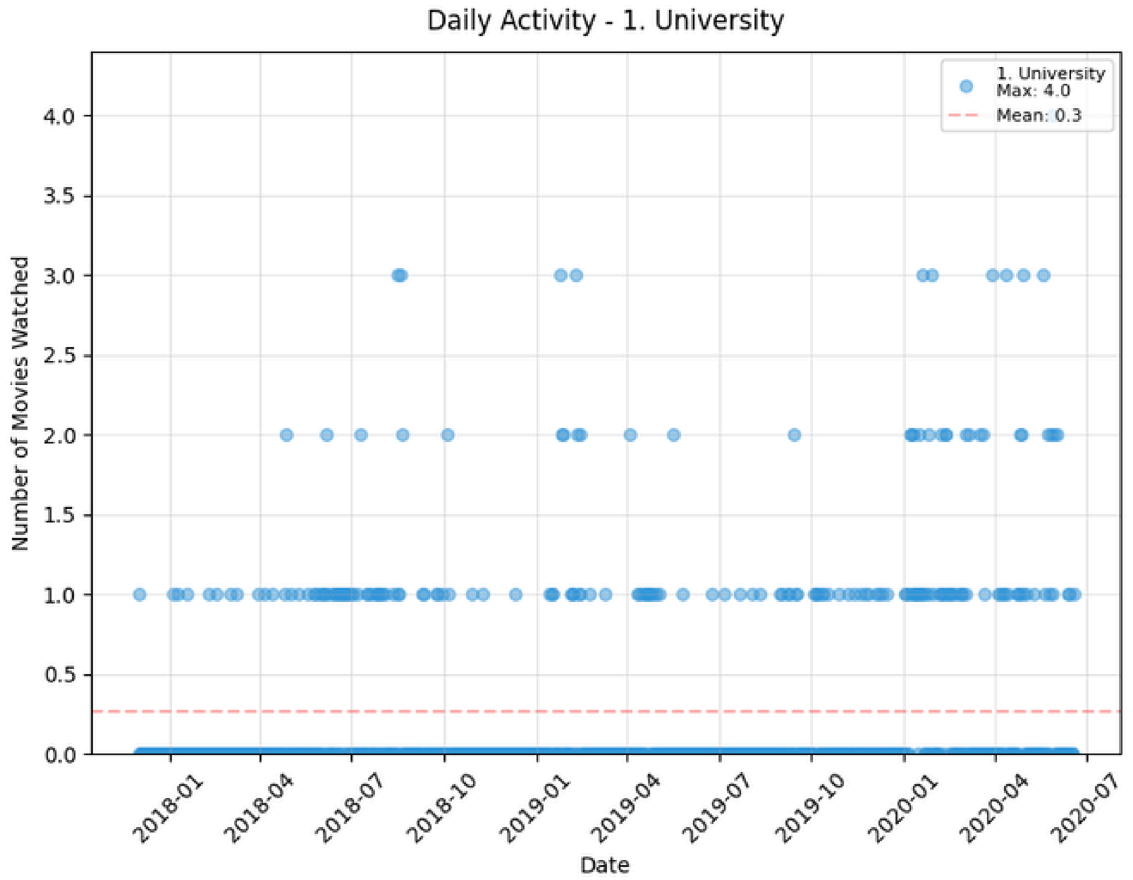
	count						date		
	mean	median	min	max	sum	std	min	max	count
custom_period									
1. University	0.26	0.0	0.0	4.0	245.0	0.58	2017-12-01	2020-06-18	931
2. University	0.18	0.0	0.0	3.0	88.0	0.44	2023-07-11	2024-11-18	497
Exam Preperation	0.25	0.0	0.0	2.0	58.0	0.52	2022-09-30	2023-05-23	236
Full-time Work	0.29	0.0	0.0	4.0	237.0	0.56	2020-06-25	2022-09-28	826

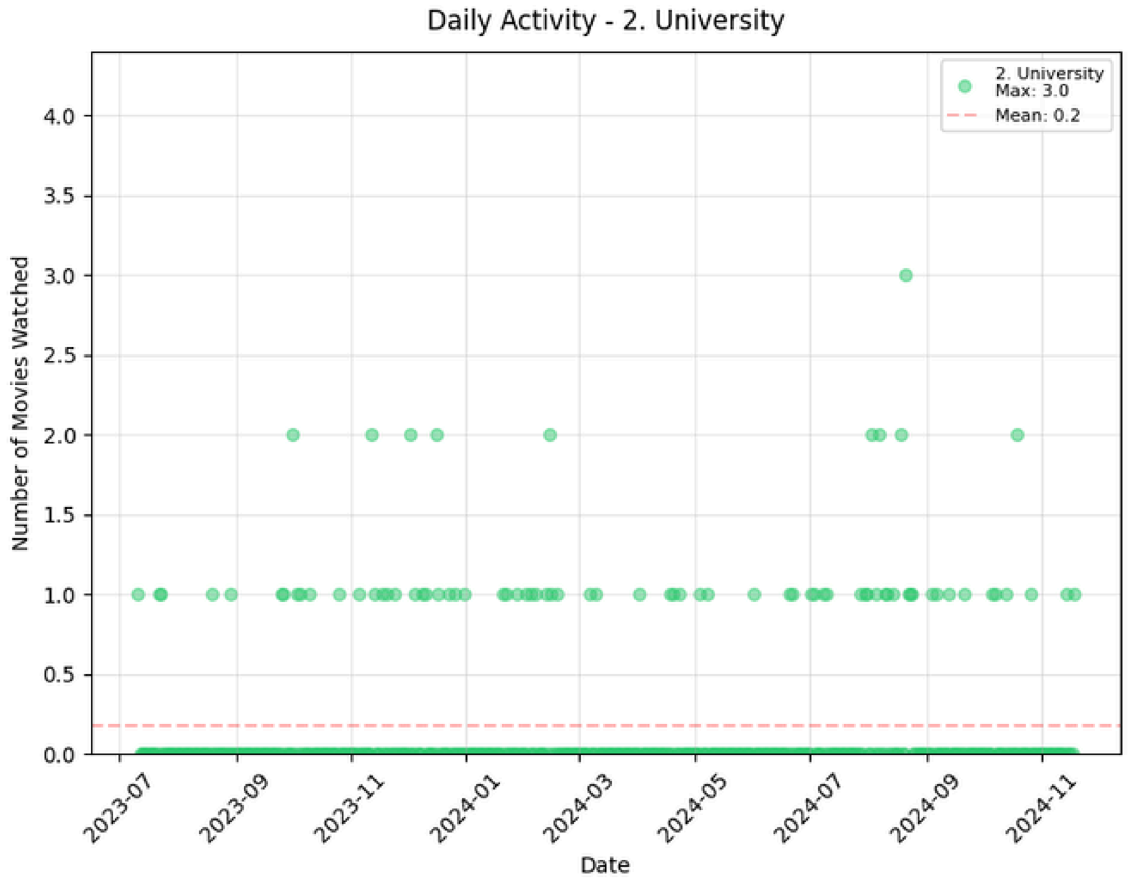
VISUALIZE LETTERBOXD MOVIE DATA

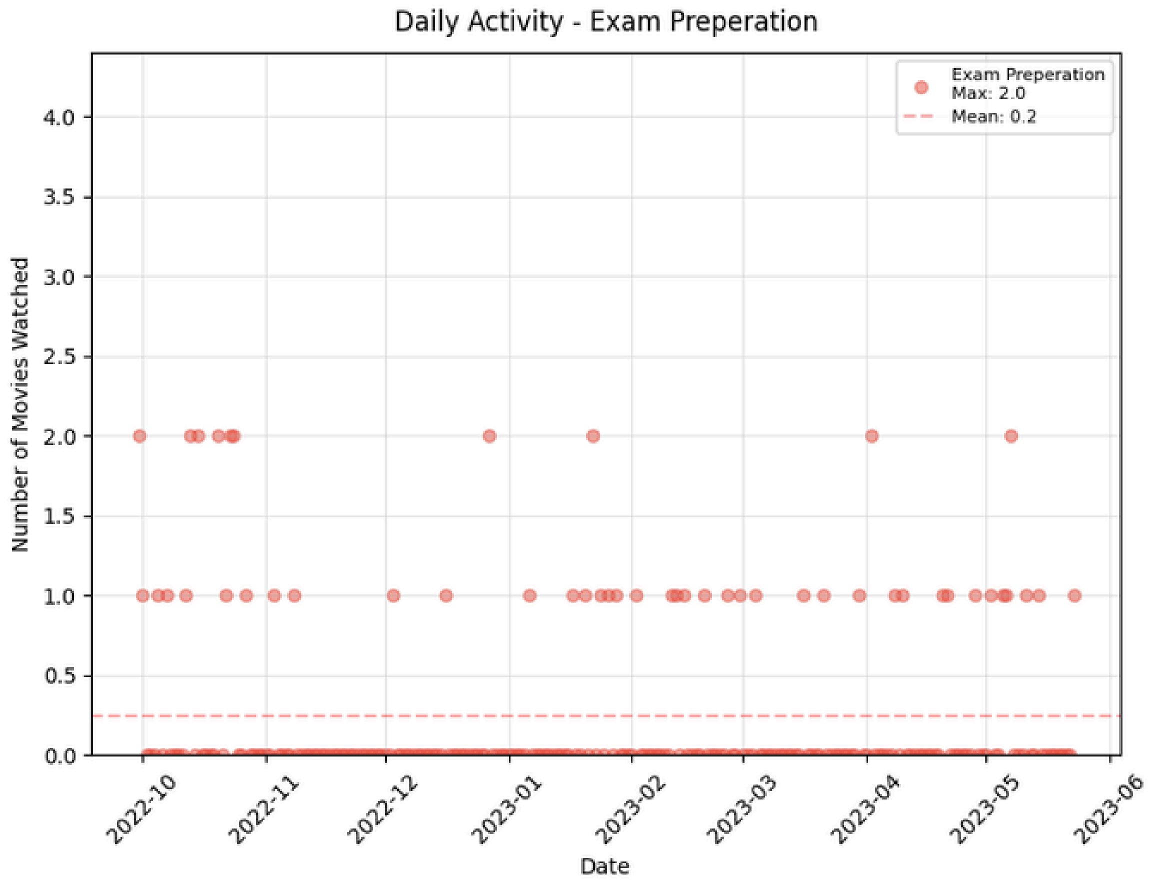
Scatter plots of daily movie watching activity for each life period were created.

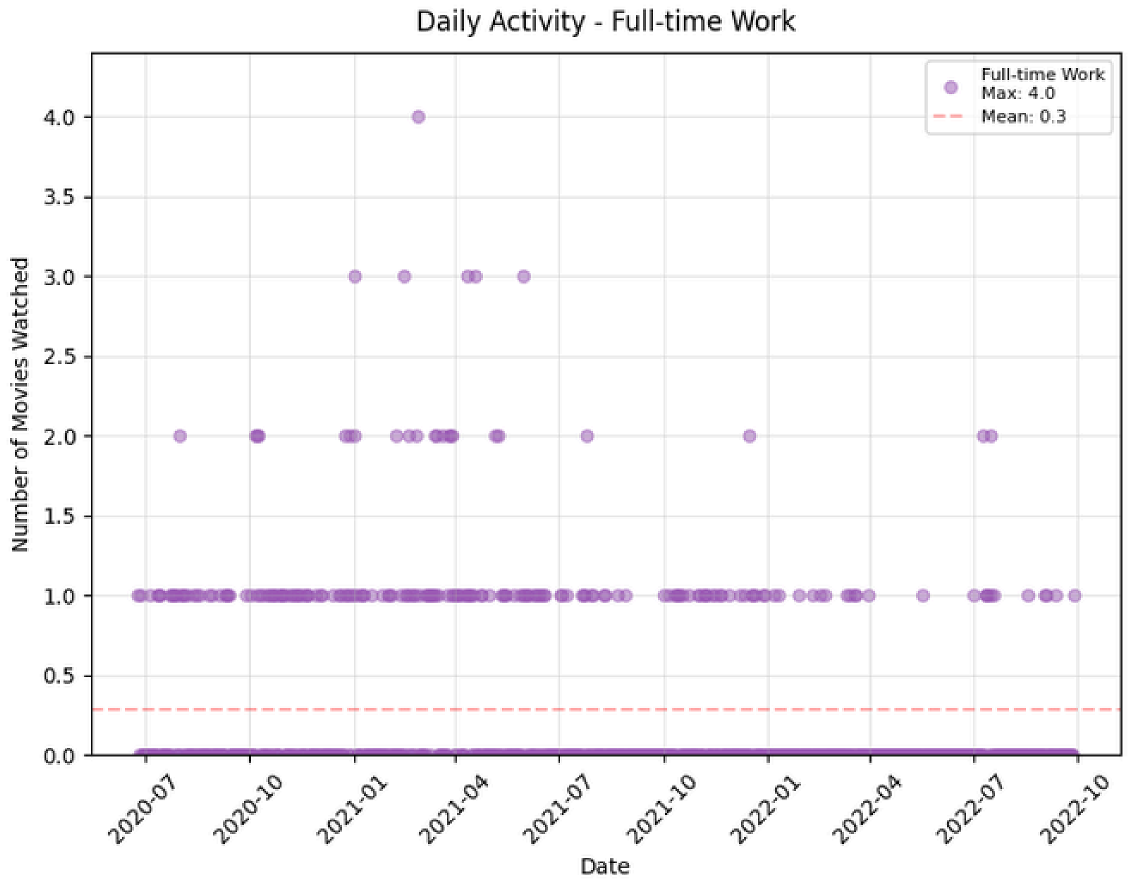
Standardized y-axis scale was used across the plots for easy comparison between periods.

Mean activity line was shown as a dashed red line for better visualization.









ENHANCE MOVIE DATA

To further understand the movie-watching patterns, what kinds of movies I chose during different periods were also analyzed in addition to when I watched movies:

1. Additional movie details (genres and duration) were scraped from Letterboxd
2. The results were cached to avoid repeatedly hitting their website to avoid being banned and increase the development speed

Including movie genres and their duration enabled me to understand if my movie preferences changed during different life periods. For example:

- Did I watch more comedies during stressful exam periods?
- Were my movie choices different during university vs. work life?
- Did my average movie duration change based on how busy I was?

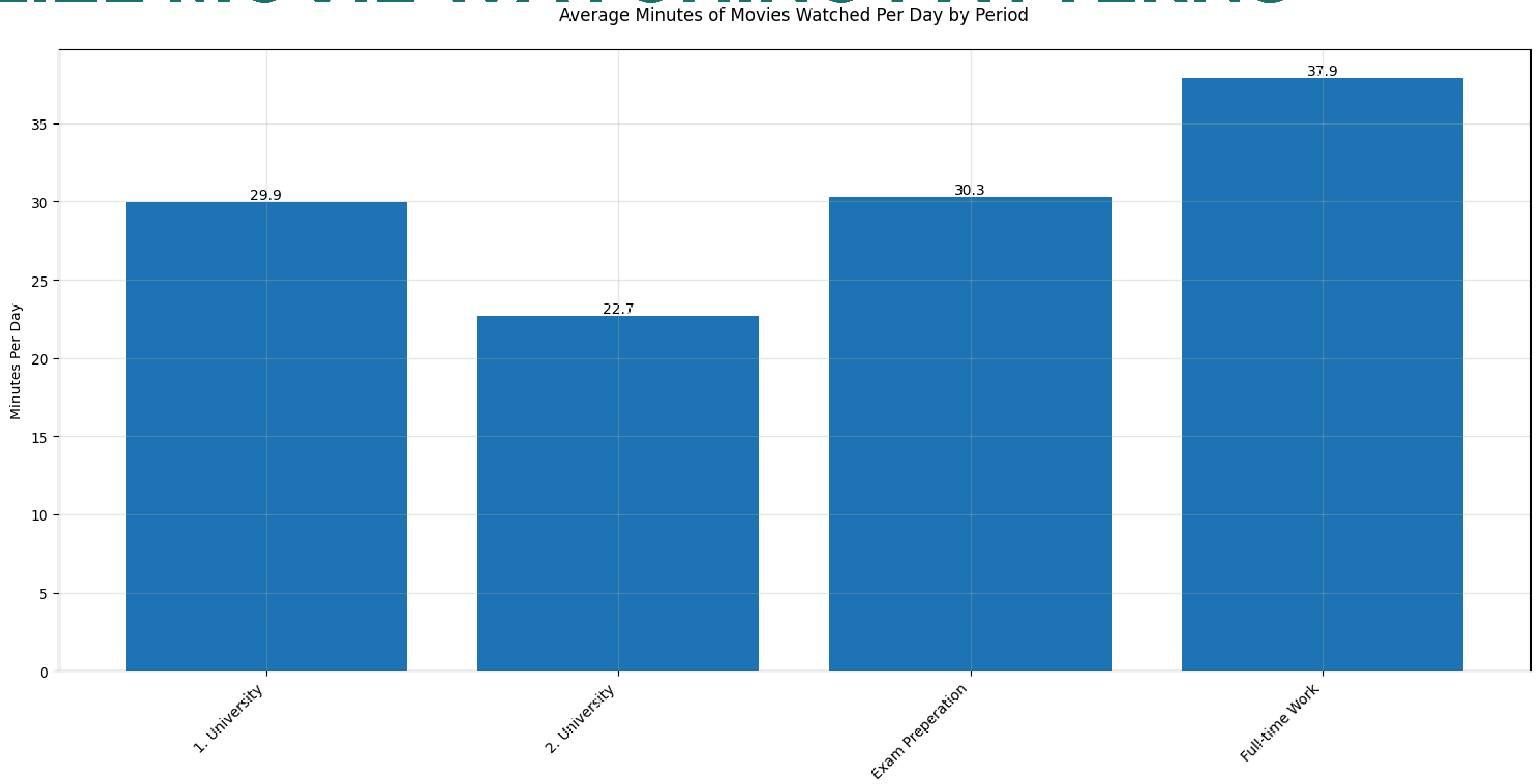
ANALYZE MOVIE WATCHING PATTERNS

1. Time Investment Analysis:
    - How many minutes per day I spent watching movies in each period was analyzed to see if my movie time commitment changed during busy periods.
2. Genre Preferences:
    - The seasonal patterns in my movie preferences were analuzed to see if my genre choices shifted during different life phases
3. Viewing Patterns:
    - Daily and monthly viewing habits were analyzed to look for changes in movie duration preferences and to identify binge-watching patterns if there is any.

Minutes Analysis by Period:

	duration				Watched Date		duration
	count	sum	mean	median	min	max	daily_avg
custom_period							
1. University	244	27865.0	114.20	114.0	2017-12-01	2020-06-18	29.96
2. University	88	11280.0	128.18	116.0	2023-07-11	2024-11-18	22.74
Exam Preperation	58	7146.0	123.21	113.5	2022-09-30	2023-05-23	30.41
Full-time Work	237	31284.0	132.00	112.0	2020-06-25	2022-09-28	37.92

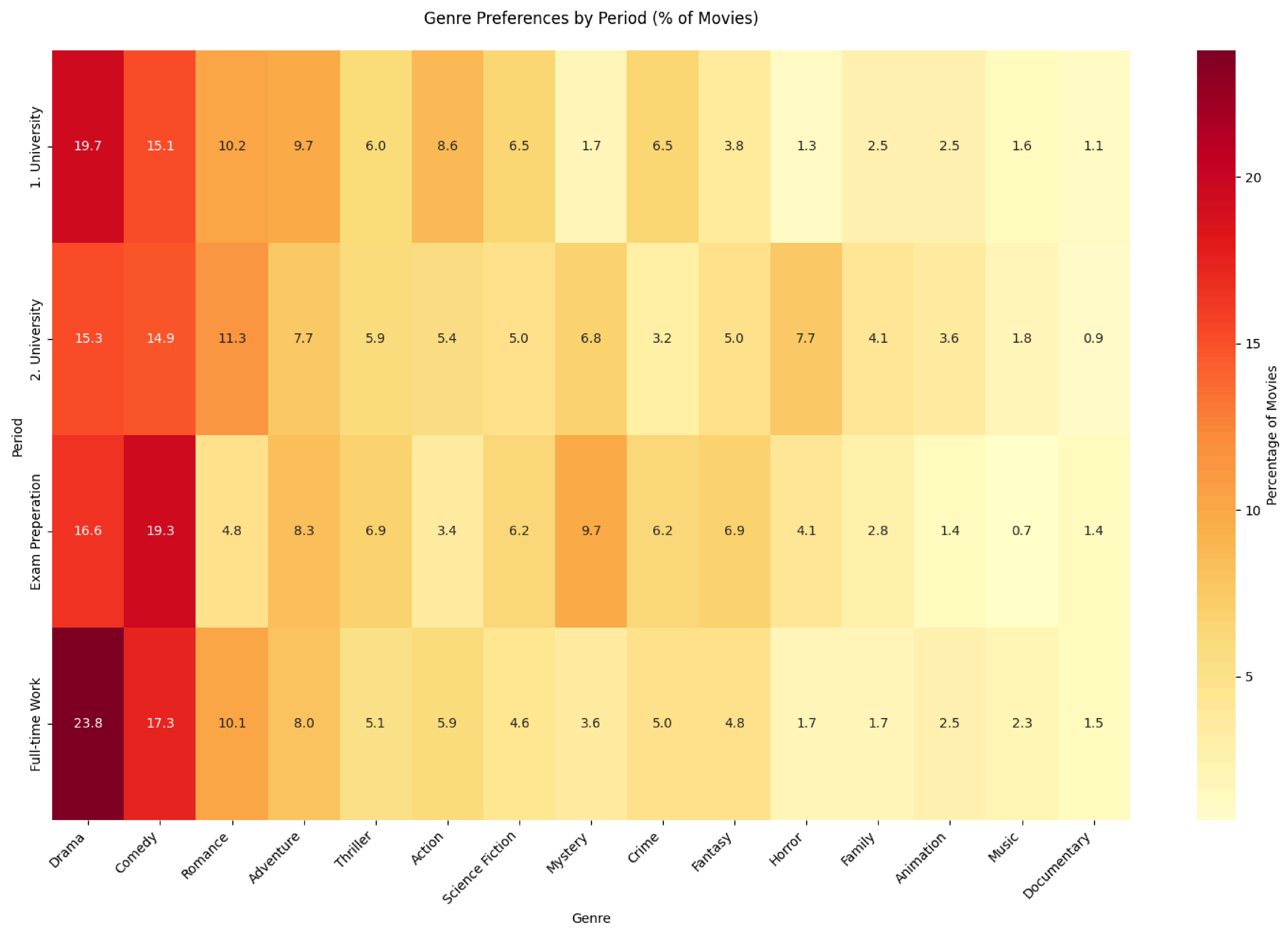
VISUALIZE MOVIE WATCHING PATTERNS



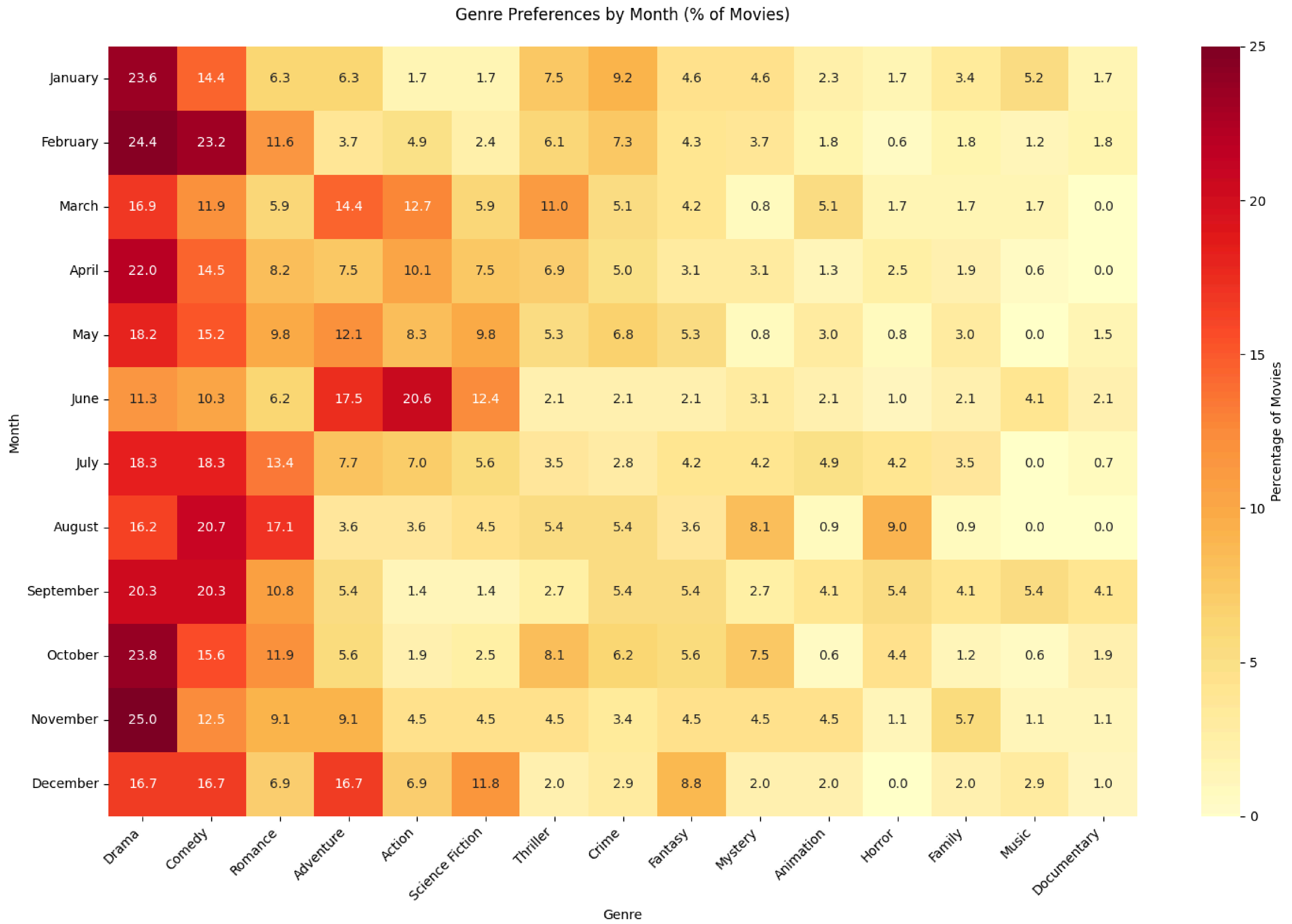
Minutes Analysis by Period:

custom_period	duration count	sum	mean	median	Watched Date min	max	duration daily_avg
1. University	244	27865.0	114.20	114.0	2017-12-01	2020-06-18	29.96
2. University	88	11280.0	128.18	116.0	2023-07-11	2024-11-18	22.74
Exam Preperation	58	7146.0	123.21	113.5	2022-09-30	2023-05-23	30.41
Full-time Work	237	31284.0	132.00	112.0	2020-06-25	2022-09-28	37.92

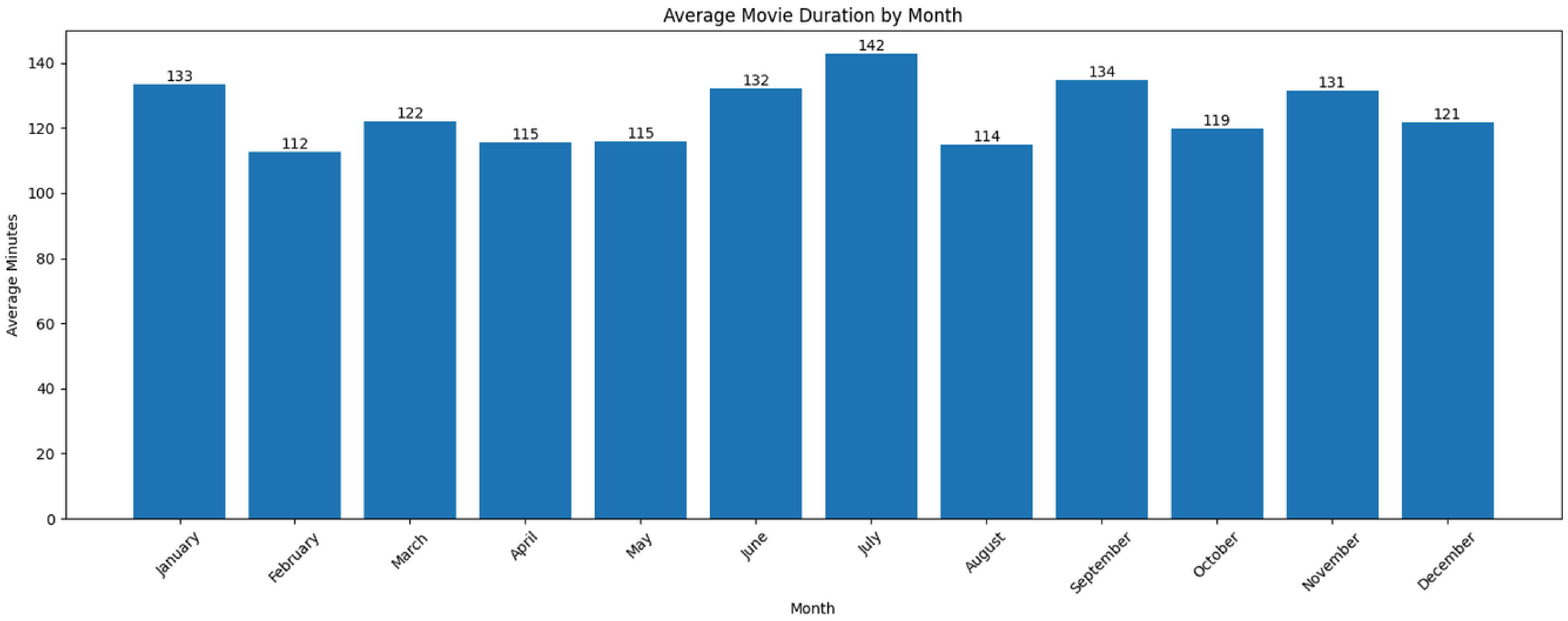
VISUALIZE MOVIE WATCHING PATTERNS

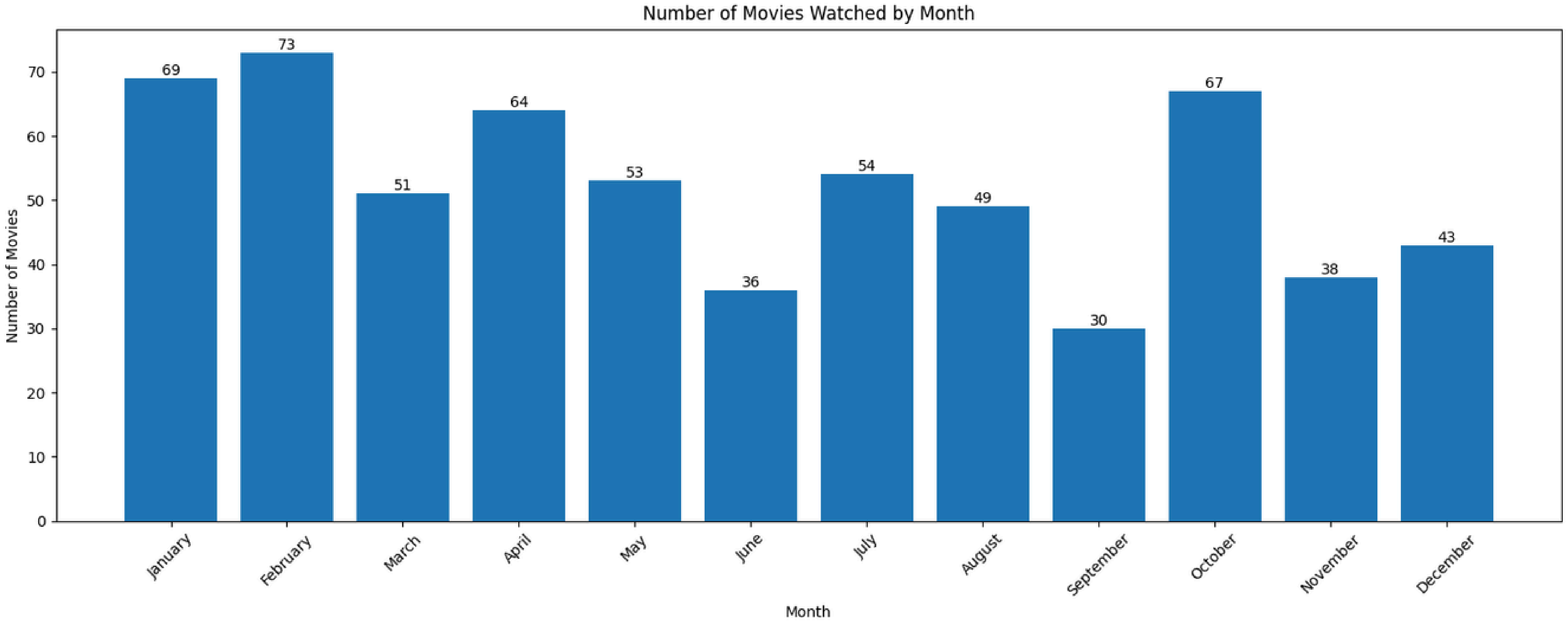


VISUALIZE MOVIE WATCHING PATTERNS



VISUALIZE MOVIE WATCHING PATTERNS





STATISTICAL ANALYSIS OF ENTERTAINMENT HABITS

To answer the following questions, busy periods were defined as exam preparation and activity patterns between busy and less busy times were compared. Mann-Whitnet U test was used to verify if any differences are significant and the differences were visualized to better understand patterns.

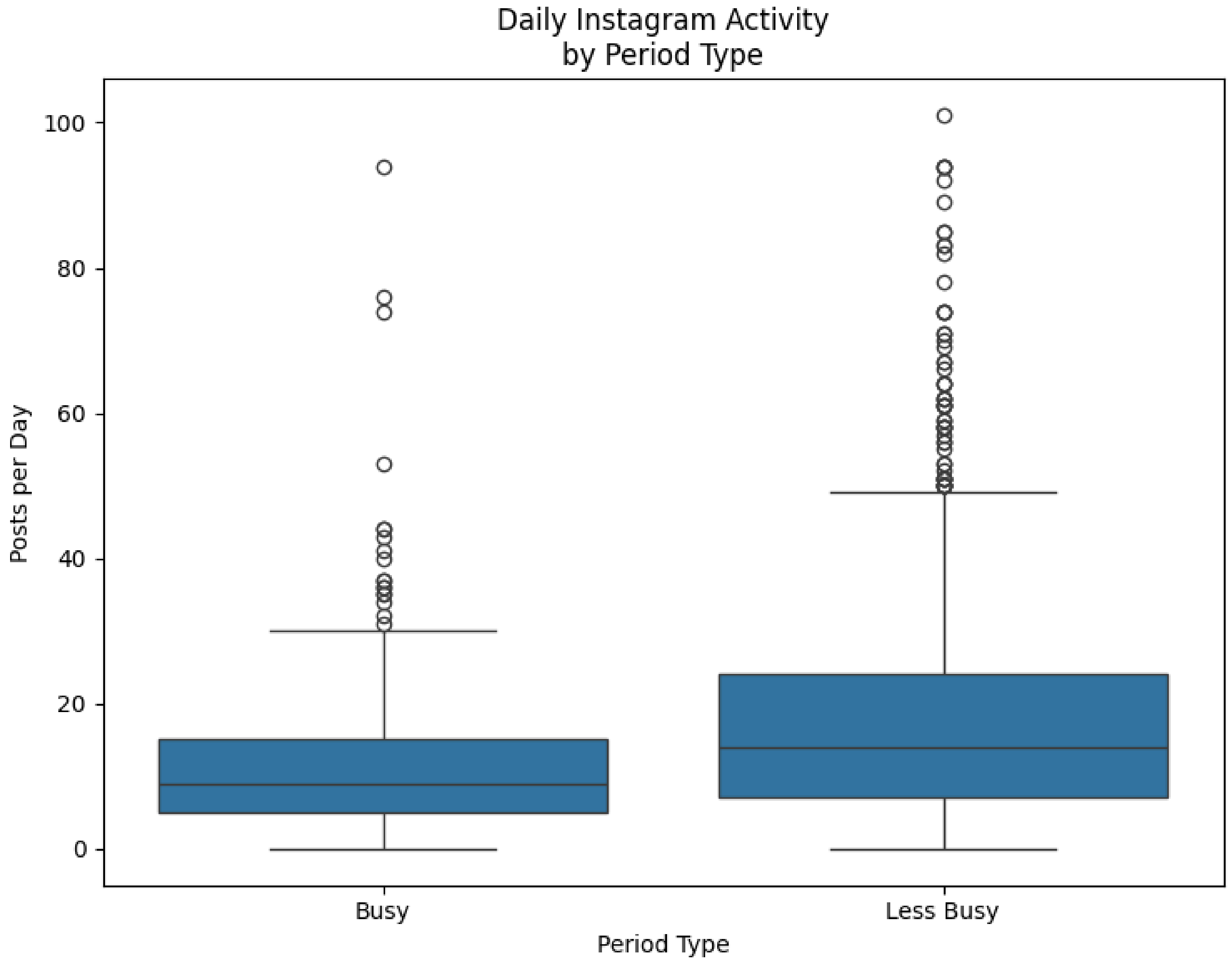
Short-form Content (Instagram):

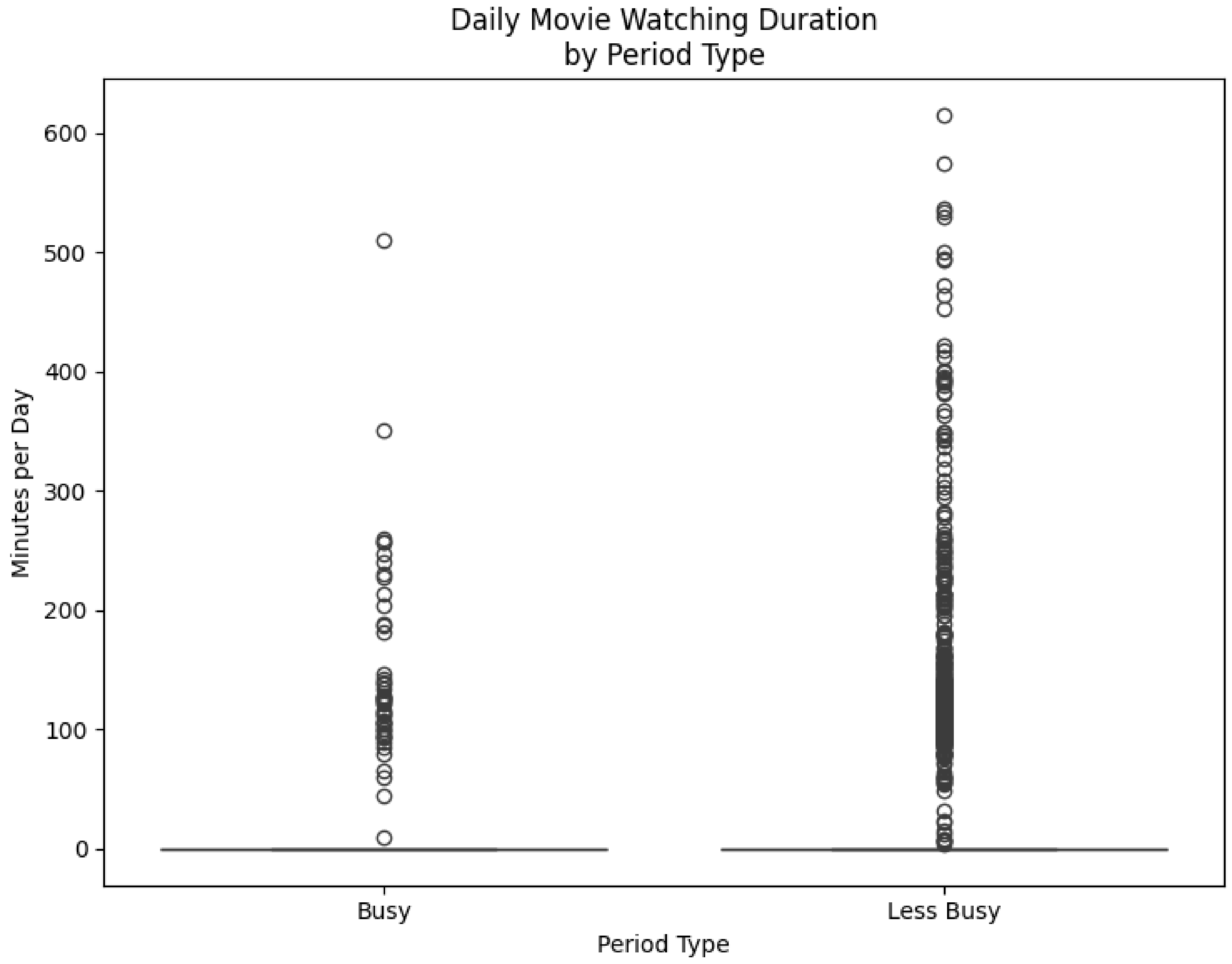
- Did I engage more with Instagram during busy periods as a quick stress relief?
- Was my Instagram usage more sporadic during exam preparation?

Long-form Content (Movies):

- Did I watch fewer movies during busy periods?
- Did I choose shorter movies when I was busier?

STATISTICAL ANALYSIS OF ENTERTAINMENT HABITS





Hypothesis Test Results:

1. Instagram Activity (shorter content):
Busy periods (Exam Prep) mean: 12.28 posts/day
Less busy periods (Work & University) mean: 17.41 posts/day
Mann-Whitney U test p-value: 1.0000

2. Movie Watching (minutes per day):
Busy periods (Exam Prep) mean: 30.28 minutes/day
Less busy periods (Work & University) mean: 31.25 minutes/day
Mann-Whitney U test p-value: 0.4395

STATISTICAL ANALYSIS OF ENTERTAINMENT HABITS

The statistical tests revealed the following patterns in how my entertainment consumption changed during busy vs. less busy periods:

Instagram Activity

- During busy periods (Exam Prep), I averaged 12.28 posts/day.
- During less busy periods (Work & University), I averaged 17.41 posts/day.
- The Mann-Whitney U test p-value of 1.0000 suggests this difference is not statistically significant.
- However, there is a moderate effect size of 0.224, indicating that while not statistically significant, there might be a practical difference in Instagram usage between periods.

Movie Watching

- During busy periods, I watched an average of 30.28 minutes of movies per day.
- During less busy periods, this slightly increased to 31.25 minutes per day.
- The Mann-Whitney U test p-value of 0.4395 indicates this difference is not statistically significant
- The effect size of 0.004 confirms that my movie watching habits were virtually identical between periods.

STATISTICAL ANALYSIS OF ENTERTAINMENT HABITS

Key Insights

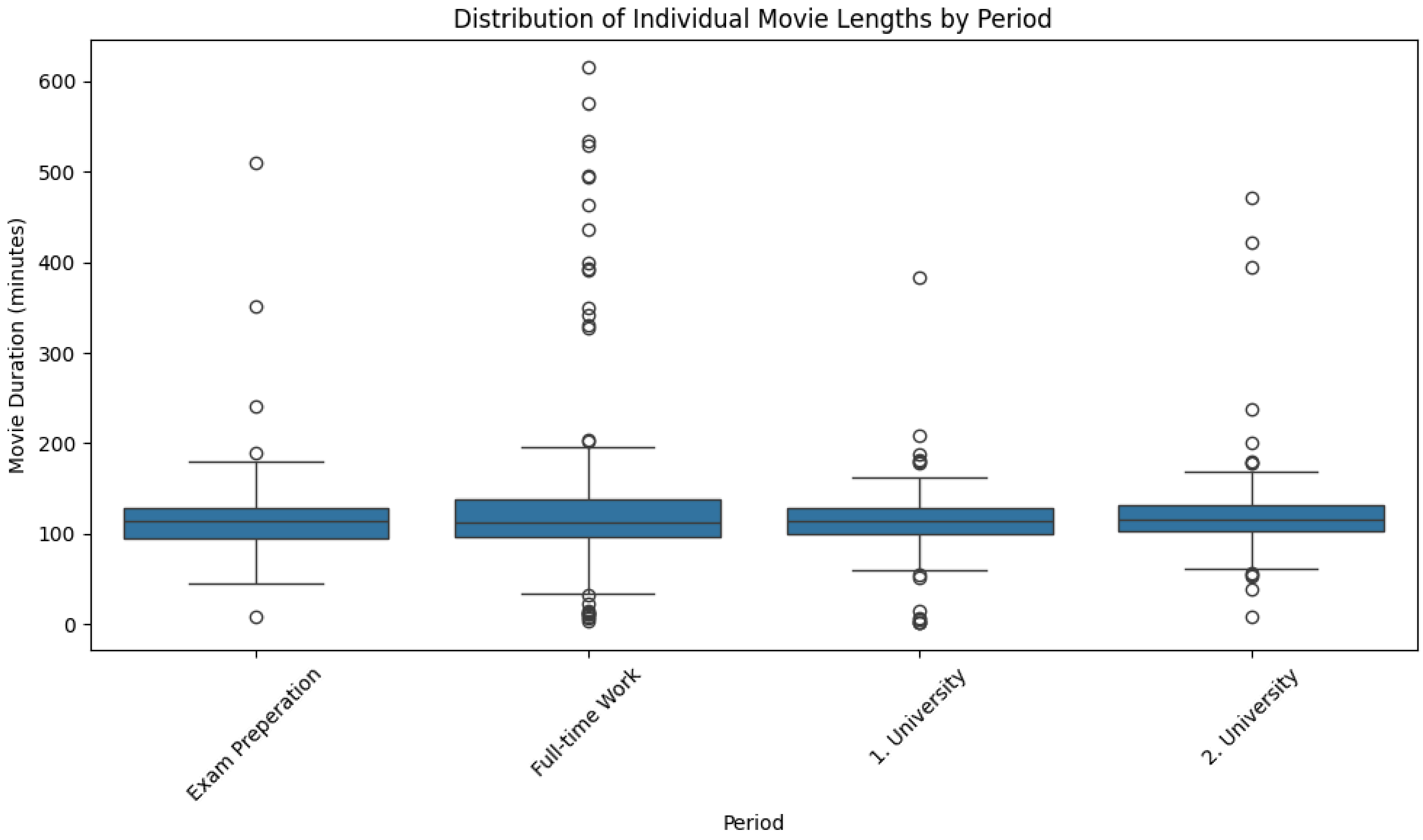
1. While statistical tests show no significant differences.
2. Instagram usage shows a moderate difference between periods (effect size 0.224), suggesting some change in behavior even if not statistically significant.
3. Movie watching remained remarkably consistent (effect size 0.004), indicating I maintained similar viewing habits regardless of how busy I was.
4. The contrast between Instagram's moderate effect size and movies' negligible effect size suggests that short-form content consumption was more affected by life circumstances than long-form content.
5. This might indicate that while I preserved my movie-watching routine, my Instagram habits were more susceptible to change based on life circumstances.
6. These findings provide a more detailed view of my entertainment consumption patterns, showing that while movie watching remained stable, Instagram usage showed practical (though not statistically significant) variations across different life periods.

ANALYZING INDIVIDUAL MOVIE LENGTH CHOICES

After looking at overall patterns, the specific movie choices were analyzed. Instead of just looking at total minutes per day, I want to understand:

1. Did I choose shorter movies during busy periods?
   - This might indicate strategic choices in entertainment consumption.
   - For example, picking a 90-minute movie instead of a 3-hour epic during exam preparation.
2. How consistent were my movie length preferences?
   - Did my preference for movie duration vary by period?
   - Were there any notable outliers (very long or short movies)?
3. What was the typical movie length in each period?
   - What are means, medians, and standard deviations?
   - Were my movie length choices more variable in certain periods?

ANALYZING INDIVIDUAL MOVIE LENGTH CHOICES




```
Movie Length Analysis:

Average Movie Length by Period:
                  mean  median    std  count
custom_period
1. University    114.20  114.0  32.26    244
2. University    128.18  116.0  66.37     88
Exam Preperation 123.21  113.5  69.83     58
Full-time Work   132.00  112.0  91.28    237

Mann-Whitney U test p-value: 0.3040

Effect size for movie length difference: -0.041
```

ANALYZING INDIVIDUAL MOVIE LENGTH CHOICES

Period-by-Period Analysis

1. 1. University: Average movie length of 114.20 minutes with relatively consistent choices (std dev: 32.26)
2. 2. University: Slightly longer average of 128.18 minutes but more variable choices (std dev: 66.37)
3. Exam Preparation: Mean length of 123.21 minutes with high variability (std dev: 69.83)
4. Full-time Work: Longest average at 132.00 minutes with the most variability (std dev: 91.28)

Busy vs. Less Busy Periods

1. Busy Periods: Average movie length of 123.21 minutes (median: 113.50)
2. Less Busy Periods: Very similar average of 123.78 minutes (median: 114.00)
3. The Mann-Whitney U test p-value of 0.3040 indicates no significant difference
4. The small negative effect size (-0.041) confirms that movie length choices were nearly identical between periods.

ANALYZING INDIVIDUAL MOVIE LENGTH CHOICES

Key Insights

1. Despite different life circumstances, I maintained remarkably consistent movie length preferences.
2. The median movie length stayed around 112-116 minutes across all periods, suggesting a "sweet spot" for my attention span.
3. Variability in choices increased over time (from std dev of 32.26 in early university to 91.28 during work).
4. Full-time work period showed the most diverse choices, reflecting more flexibility in entertainment scheduling.
5. The similarity between busy and less busy periods suggests that when I chose to watch a movie, the time commitment was npt a major factor in my selection.
6. This analysis suggests that while my overall movie watching frequency might have changed, when I did choose to watch a movie, I did not significantly alter my preferences for movie length based on how busy I was.

ANALYZING ENTERTAINMENT HABITS DURING HIGH VS. LOW STRESS PERIODS

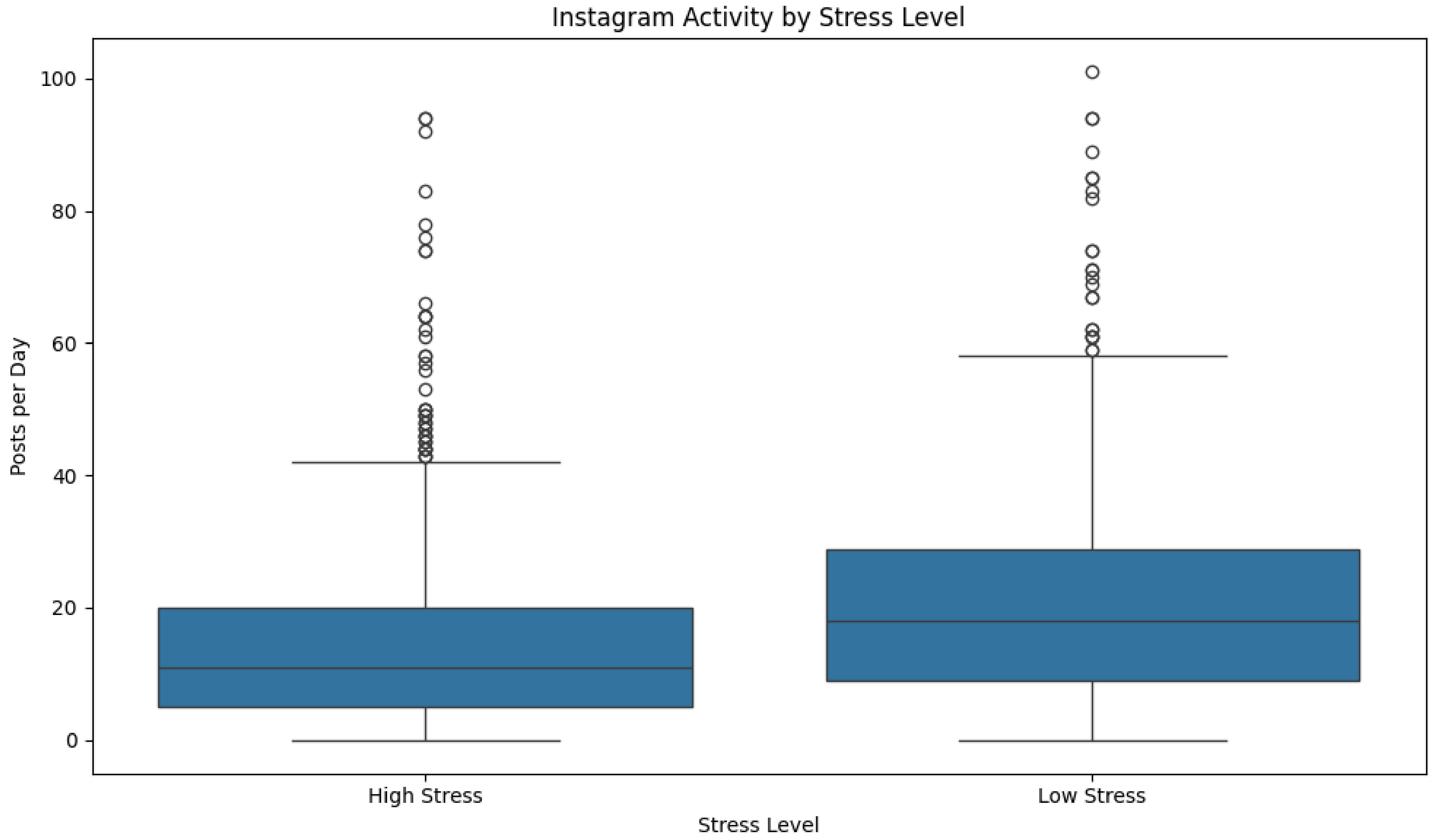
After examining busy vs. less busy periods, entertainment consumption was analyzed according to stress levels. The periods were categorized as:

1. **High Stress**
   a. Exam Preparation
   b. University
   c. University
2. **Low Stress**
   a. Full-time Work

The goal was to understand:

1. Did stress levels affect my entertainment choices differently than just being "busy"?
2. Was there a difference in how I consumed short vs. long-form content during stressful periods?
3. Did the predictability of work life (low stress) vs. university/exam life (high stress) impact my entertainment habits?

ANALYZING ENTERTAINMENT HABITS DURING HIGH VS. LOW STRESS PERIODS



Instagram Usage Analysis by Stress Level:

High Stress (University & Exam) average posts/day: 14.21 ± 13.03
Low Stress (Work) average posts/day: 21.28 ± 16.61

Mann-Whitney U test p-value: 1.0000
Effect size: 0.286

ANALYZING ENTERTAINMENT HABITS DURING HIGH VS. LOW STRESS PERIODS

Instagram Usage Analysis by Stress Level:

1. High Stress (University & Exam) average posts/day: 14.21 ± 13.03
2. Low Stress (Work) average posts/day: 21.28 ± 16.61
3. Mann-Whitney U test p-value: 1.0000 indicates no statistically significant difference
4. Effect size: 0.286 suggests a moderate practical difference

ANALYZING ENTERTAINMENT HABITS DURING HIGH VS. LOW STRESS PERIODS

Key Findings

1. While not statistically significant (p = 1.0000), there was actually higher Instagram usage during low-stress work periods.
2. The moderate effect size (0.286) indicates a meaningful difference in usage patterns.
3. The higher standard deviation during work periods (16.61 vs 13.03) suggests more variable usage patterns during low-stress times.
4. This contradicts the initial hypothesis that stress might lead to increased social media use.
5. The data suggests I might have had more consistent routines during high-stress periods, while allowing for more varied usage during more relaxed work life.
6. This analysis reveals that my Instagram usage patterns were different than expected, with higher and more variable usage during low-stress work periods rather than during high-stress academic times. This might indicate that during high-stress periods, I maintained more structured and controlled social media habits.

THANK YOU