Netflix Viewing Behavior Analysis Report

Begüm

May 30, 2025

Abstract

This report presents a comprehensive analysis of personal Netflix viewing activity from January 2019 through May 2025. We examine temporal patterns at daily, weekly, and hourly resolutions; compare series versus movies consumption; quantify binge-watching behaviors and seasonal effects; perform correlation and hypothesis testing; and develop predictive models for day-of-week and time-of-day classification. All findings are supported by detailed visualizations and statistical inference.

1 Introduction

Understanding the "when" and "what" of streaming consumption can inform both content strategy and personal viewing habits. This study aims to:

1. Characterize temporal viewing patterns (daily, weekly, hourly).
2. Compare consumption of series versus movies.
3. Quantify binge-watching frequency, seasonality, and favorite titles.
4. Assess seasonal influences on viewing volume and session length.
5. Explore correlations among key metrics and perform hypothesis tests.
6. Build and evaluate classifiers for viewing day-of-week and time-bins.

2 Data and Preprocessing

The raw data (`ViewingActivity.csv`) includes:

- `Start Time` (timestamp),
- `Duration` (HH:MM:SS),
- `Title` (series/episode or movie),
- `Device Type`, `Country`, etc.

Key cleaning steps:

1. Parse `Start Time` into `timestamp`, derive `date`, `hour`, `day_of_week`, `DayNum`.
2. Convert `Duration` to `duration_min`.
3. Split `Title` into `Series`, `Season`, `Episode` via regex; assign `ContentType` = Series or Movie.
4. Derive `IsWeekend`, `ViewsPerDay`, `IsBingeDay` (3 views/day).
5. Map month to `SeasonBin` = {Winter, Spring, Summer, Fall}.

The cleaned DataFrame is referred to as `df_clean` throughout.

3 Temporal Viewing Patterns

3.1 Monthly and Weekly Trends

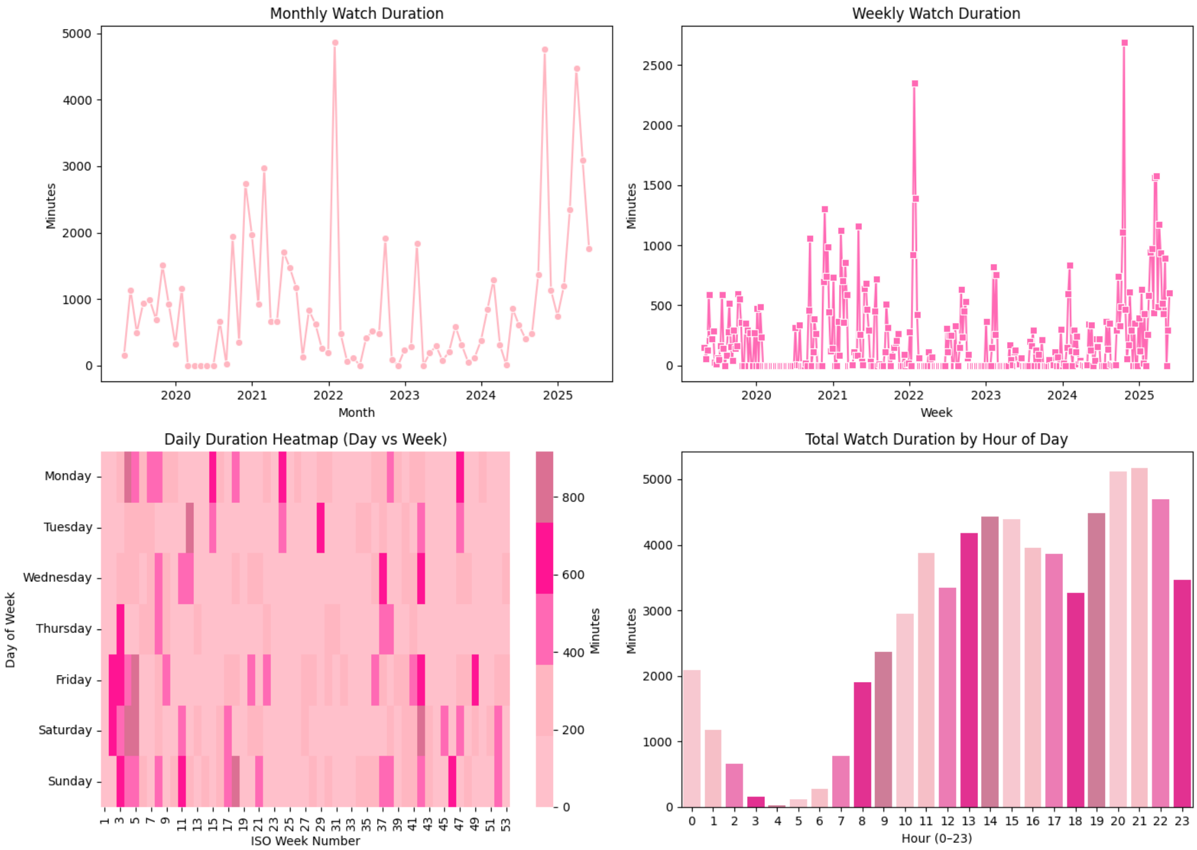

Figure 1: (a) Monthly total watch duration; (b) Weekly total watch duration; (c) Heatmap of daily duration by ISO week & weekday; (d) Hourly distribution of total duration.

- **Monthly peaks**: January 2022 (4,900 min), November 2024 (4,800 min), March 2025 (4,500 min).
- **Weekly volatility**: some weeks >2,700 min; many weeks <500 min.
- **Daily heatmap**: Fridays/Saturdays in early 2025 stand out.
- **Hourly profile**: peak viewing at 20:00–22:00 (5,200 min at 21:00).

3.2 Series vs. Movie Share

Series comprise **72.6%** of sessions; movies account for **27.4%**.

4 Content Breakdown

4.1 Top Series and Episode Durations

- **Top series**: Friends (359 views), HIMYM (195), B99 (170), The Blacklist (168).
- **Season distribution**: Season 1 (850 eps), Season 2 (520), down to 40 at Season 10.

Figure 2: Series vs. Movie share of all sessions.

- **Episode durations**: bimodal peaks at 22–25min and 45–50min.
- **Cumulative growth**: Friends shows steady consumption; HIMYM/B99 plateau then surge in late 2024.

5 Binge-Watching Patterns

- "Binge day" = 3 views/day.
- **Avg binge size**: peak in March (12.6 eps), April (12.5); lowest in August (5.6).
- **Binge days**: September (47), October (45); July (15).
- **Top binge series**: Friends (359), HIMYM (195), B99 (170).
- **Size distribution**: most binge days involve 3–8 eps; rare days exceed 20.

6 Seasonal Viewing Patterns

- **Total sessions**: Fall (1100), Winter (1097), Spring (986), Summer (498).
- **Total minutes**: Winter (22,018min), Fall (19,589min), Spring (15,488min), Summer (9,631min).
- **Movie vs. Series**: Series dominate all seasons (>19,000min in Winter/Fall).
- **Avg duration**: Winter (20.1min), Summer (19.3min), Fall (17.8min), Spring (15.7min).

7 Correlation & Trend Analysis

- ViewsPerDay vs. IsBingeDay: $r = 0.32$.
- DayNum vs. IsWeekend: $r = 0.82$.
- duration_min nearly uncorrelated with other metrics.
- Scatter shows no clear weekend/weekday separation in session lengths.

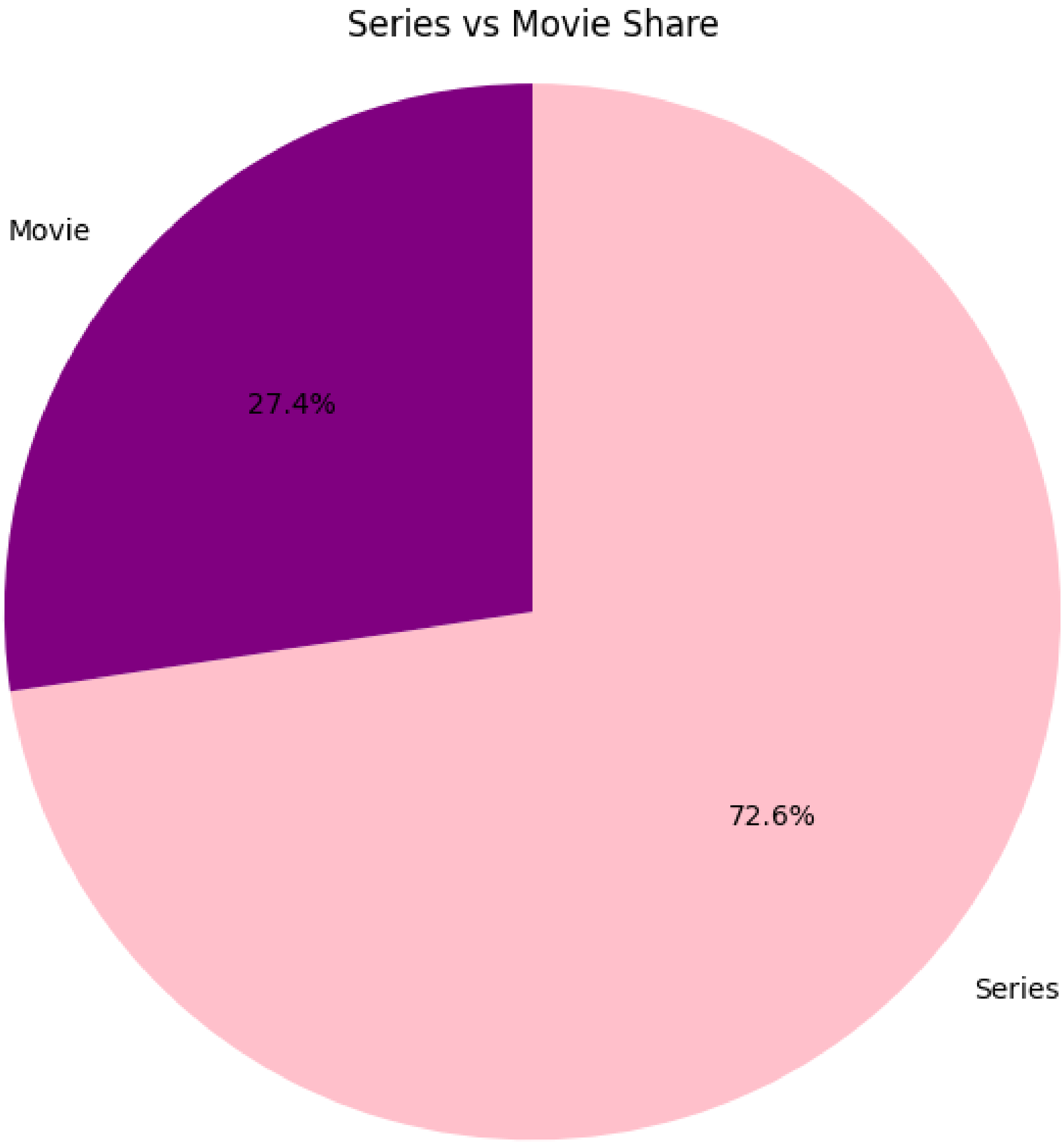

Figure 3: (a) Top 10 series by view count; (b) Season number distribution; (c) Episode duration histogram; (d) Cumulative monthly views for top 3 series.

8 Hypothesis Testing

We test at $\alpha = 0.05$:

1. **Weekend vs. Weekday views/day**: $t = 1.68,\ p = 0.0943$. No significant difference.
2. **Winter vs. Summer binge rates**: $\chi^2 = 8.41,\ p = 0.0037$. Higher binge rates in Winter.
3. **Weekend vs. Weekday total duration**: $t = 1.17,\ p = 0.2447$. No significant difference.

9 Predictive Modeling

9.1 Day-of-Week Classification

Overall accuracy 40%; best recall on Sunday (69%) & Saturday (63%); poor separation for Monday/Friday/Tuesday.

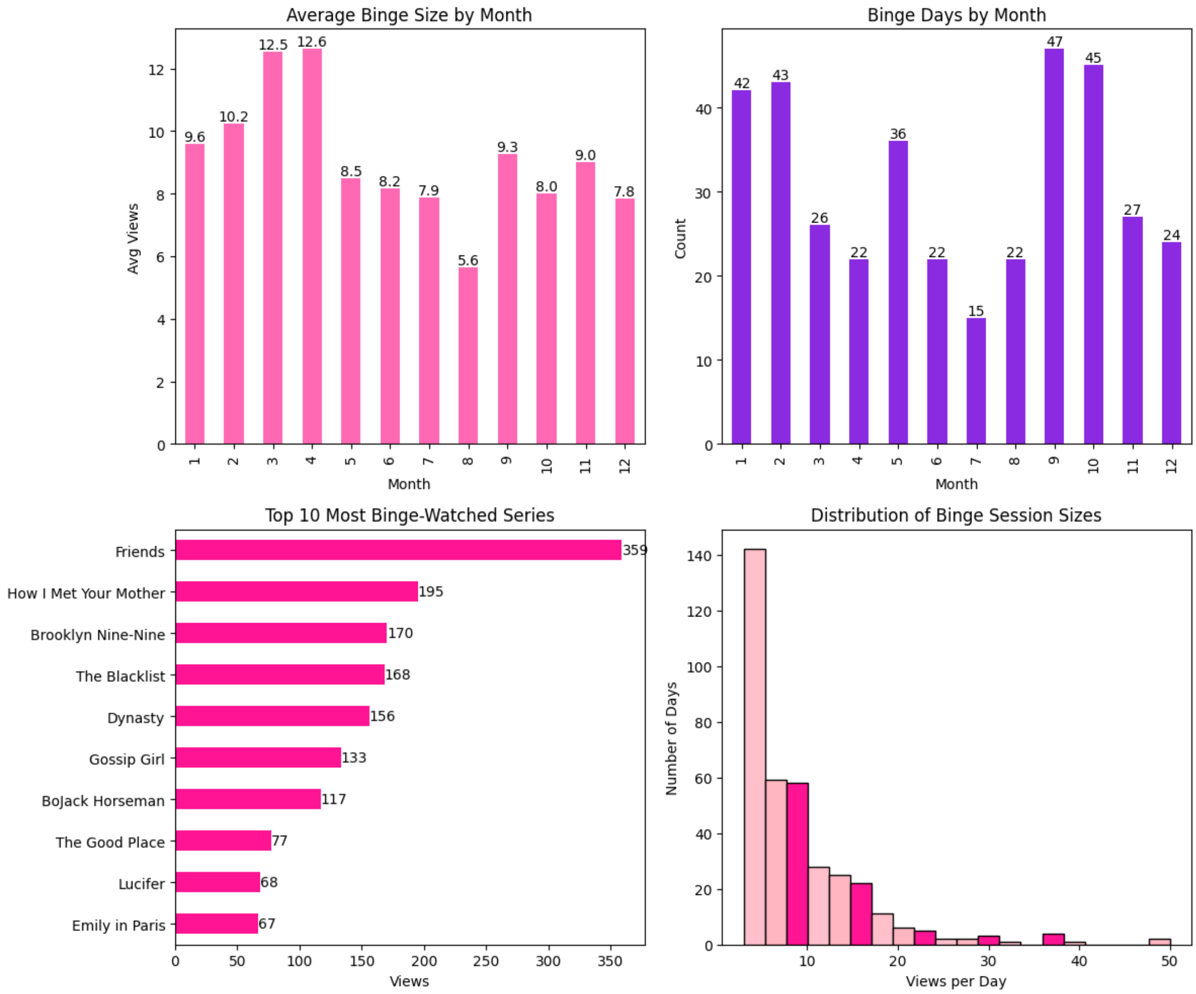

Figure 4: Binge-watching analysis: (a) Avg binge size by month; (b) Binge days by month; (c) Top binge series; (d) Distribution of binge session sizes.

9.2 Time-Bin Classification

Perfect accuracy (100%) achieved—model trivially learns deterministic mapping from hour to time-bin.

10 Conclusion

This multi-method analysis reveals:

- A pronounced preference for series over movies (72.6% vs. 27.4%).
- Viewing peaks in winter and fall, lulls in summer.
- Binge-watching most frequent in September–October and March–April.
- Session length independent of time/day; binge classification weakly correlated with daily view count.
- Predictive modeling easily classifies time-bins but struggles with day-of-week.

These insights can inform personalized recommendations and schedule planning for optimized streaming experiences.

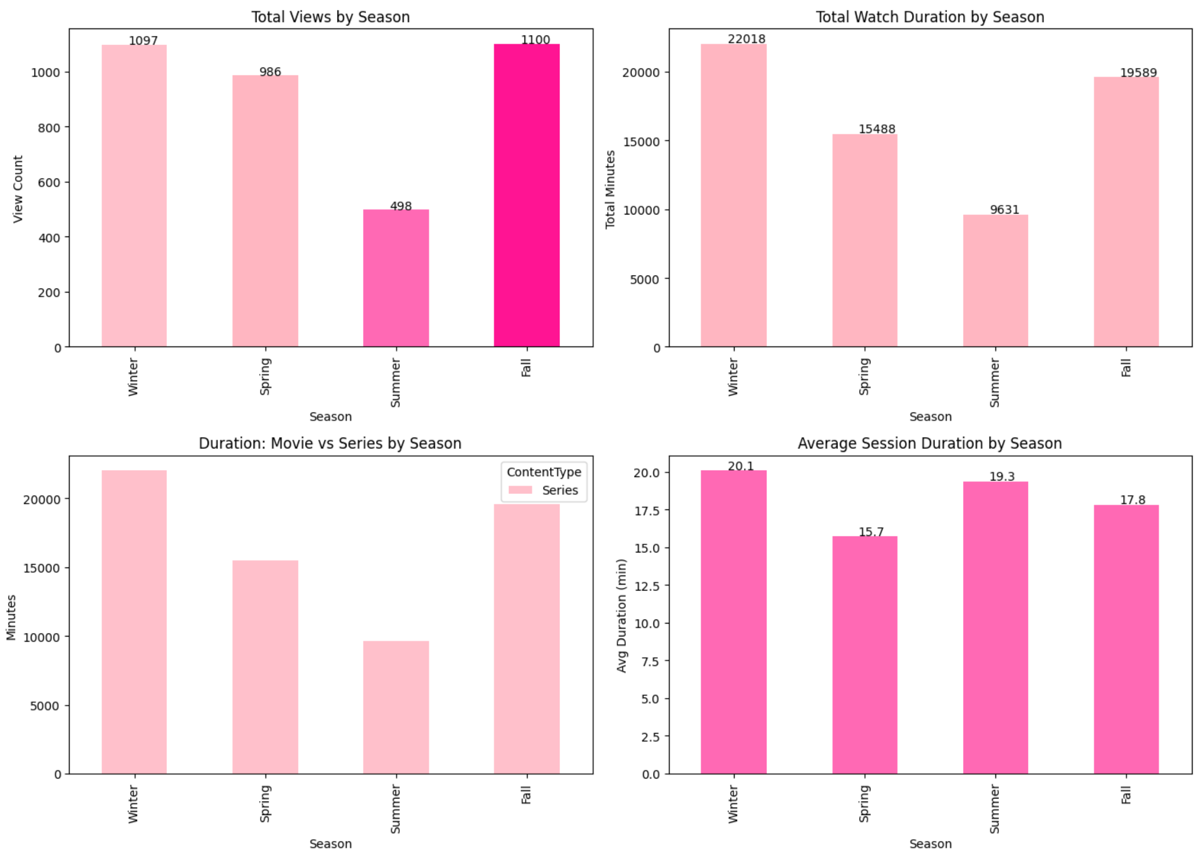

Figure 5: Seasonal patterns: (a) Total sessions; (b) Total duration; (c) Movie vs. Series duration; (d) Avg session duration by season.

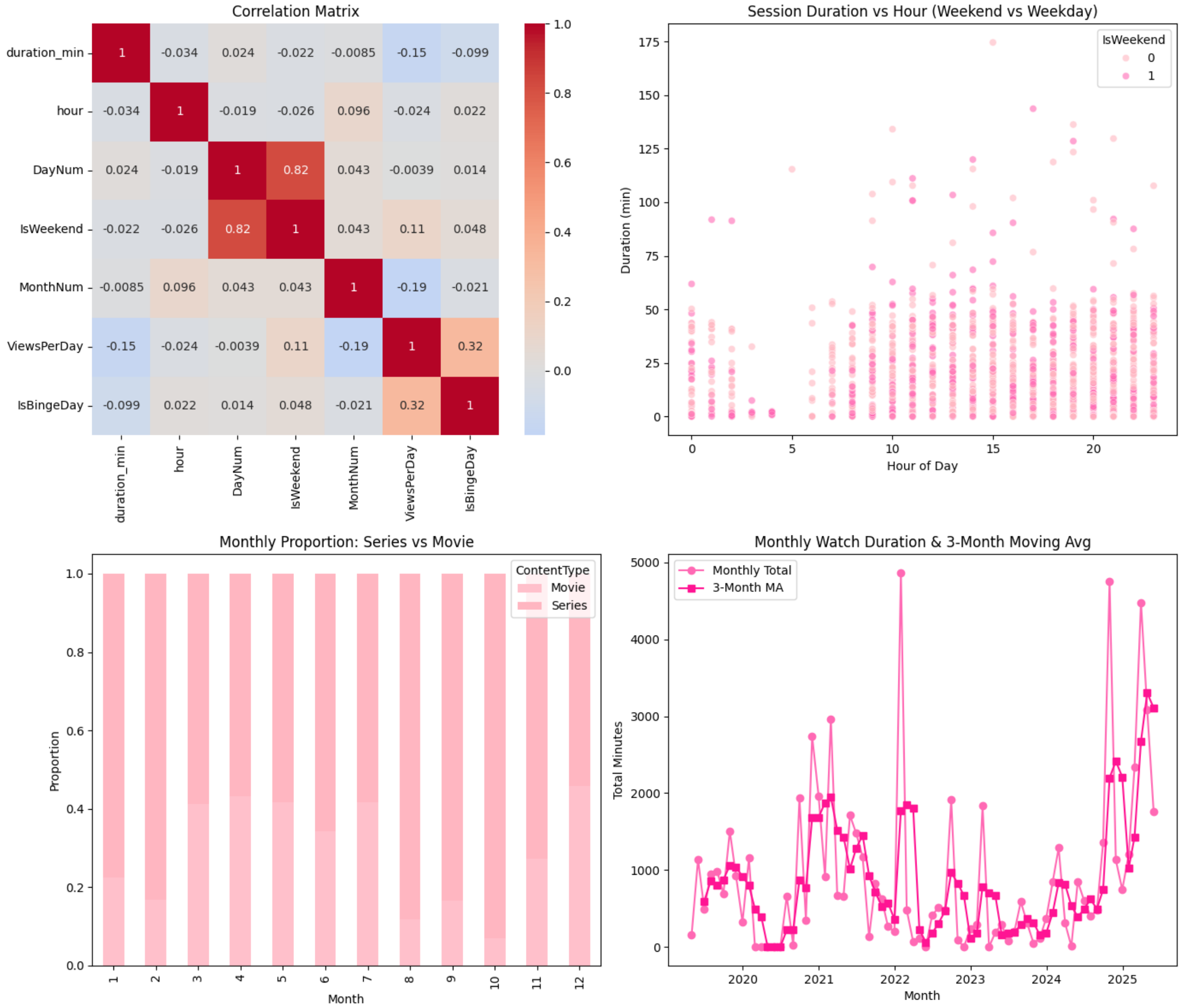

Figure 6: (a) Correlation matrix of key features; (b) Session duration vs. hour, colored by weekend flag.

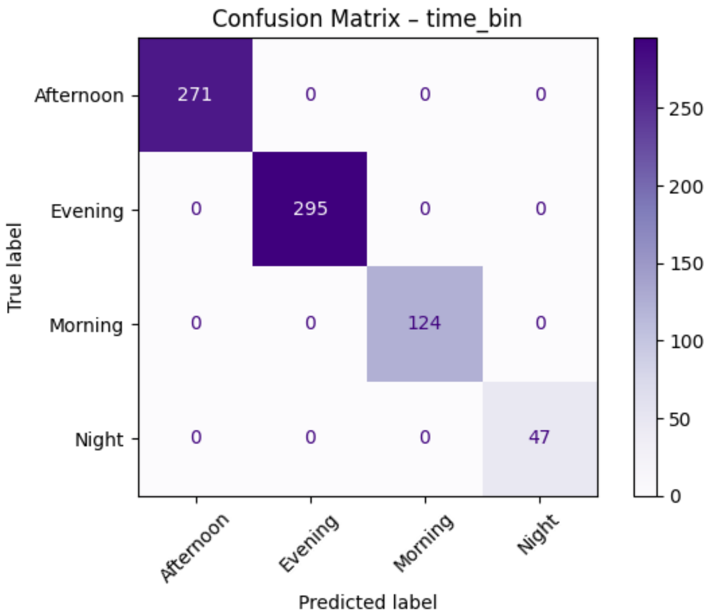

Figure 7: Confusion matrix for day-of-week classifier (Random Forest, 200 trees).

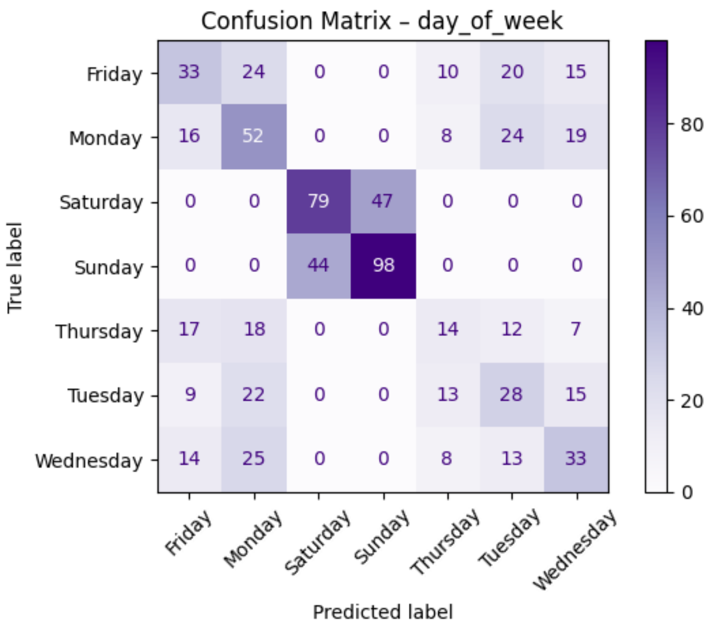

Figure 8: Confusion matrix for time-bin classifier.