Learning and Habits: The Borrowing Behavior of Library Users

Nico Rinck [*1] **Adriano Polzer** [*2] **Robin Allgeier** [*3] **Jonas Mahr** [*4] **Jannik Rombach** [*5]

Abstract

Late returns of library materials disrupt circulation efficiency and strain administrative resources. This study analyzes 2.4 million loan transactions from approximately 21,000 users at the Tübingen City Library (2019–2025) to understand borrowing behavior patterns and learning effects. We develop a session-based analytical framework that aggregates item-level records into user visits. Our analysis reveals that late return probability declines sharply during early sessions, while extension requests increase from 50% to 65%, indicating proactive learning rather than simple compliance. Borrowing activity exhibits strong temporal patterns aligned with operating hours, and media-type preferences remain influential but become less predictive with experience. These findings suggest that targeted communication systems for new users may be more effective than uniform enforcement mechanisms.

1. Introduction

Public libraries serve a critical role in democratic societies by providing equal access to knowledge, education, and cultural resources. Understanding user borrowing behavior enables libraries to improve service quality and ensure materials remain available to all. This study analyzes borrowing records from the Tübingen City Library, addressing key operational questions identified in collaboration with library staff to support evidence-based decision-making. Specifically, we investigate late return patterns, user learning effects, and temporal borrowing regularities.

A central operational challenge is the timely return of borrowed materials. Late returns disrupt circulation efficiency: when items remain overdue, they are marked as on-loan in the system but are neither available for other patrons nor counted as part of the accessible collection. In severe cases, overdue items can escalate to multiple reminders and eventually legal action to recover replacement costs, straining administrative resources. Studies of academic library circulation demonstrate that return behavior is shaped by multiple factors, including user characteristics, reminders, and institutional policies (Phelps & Campbell, 2015).

Beyond late returns, understanding broader borrowing patterns such as temporal regularities, learning effects, and media preferences enables data-driven policy decisions. We develop a session-based analytical framework (Section 2) that aggregates item-level records from a large-scale dataset into user visits, recognizing that patrons borrow multiple items during a single library visit. Our analysis (Section 3) reveals distinct temporal usage patterns tied to operating hours and shows that late returns decline over initial sessions while extension use increases. These findings can be interpreted in terms of user learning, behavioral adaptation, and their implications for library operations (Section 4).

2. Data and Methods

We base our analysis on a non-public export from the Tübingen City Library's circulation system covering the years 2019 to 2025. The dataset comprises over 2.4 million completed loan transactions from approximately 21 000 users. Each record corresponds to a single borrowed item and includes basic transaction information: borrowing and return timestamps, the number of loan extensions, and a flag marking whether the item was returned late. Additionally, each record contains anonymized user identifiers, user categories (distinguishing adults, children, and institutional accounts), and item-specific information such as media type indicators for 21 distinct categories. All analyses can be reproduced using the accompanying code repository.

Given the complexity of library circulation systems and internal data structures, we supplemented the quantitative dataset with qualitative domain expertise. We gathered domain knowledge through a interview with Mr. Feldhoff-Lange, a staff member of the Tübingen City Library, which provided contextual insights into data collection procedures, internal system conventions, and data quality practices.

[*]Equal contribution [1]Matrikelnummer 7314094, MSc Computer Science [2]Matrikelnummer 7366709, MSc Computer Science [3]Matrikelnummer 6288873, MSc Computer Science [4]Matrikelnummer 7322288, MSc Computer Science [5]Matrikelnummer 7317181, MSc Computer Science. Correspondence to: JR <jannik.rombach@uni-tuebingen.de>.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2025/26 (Module ML4201). Style template based on the ICML style files 2025.

Follow-up questions were clarified via email communication, which helped resolve inconsistencies in the raw export and inform several preprocessing decisions.

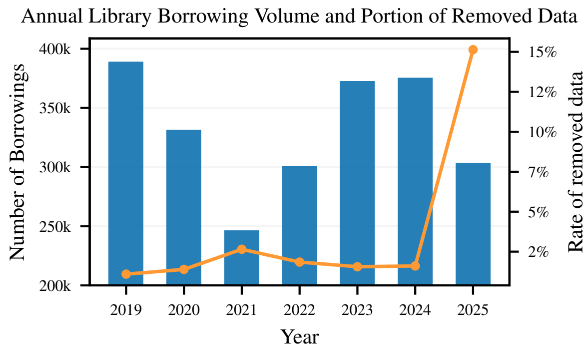

Figure 1. Library borrowing statistics by year. Blue bars show the total number of borrowings per year (left axis). The orange line indicates the share of transactions removed during data cleaning (right axis).

Figure 1 summarizes the dataset at the annual level. The blue bars show the total number of borrowings per calendar year, with a pronounced drop in 2021 and a subsequent recovery. By 2024, borrowing volume has almost returned to pre-pandemic levels, indicating a near restoration of overall library usage.

2.1. Data Quality and Cleaning

The raw export from the circulation system reflects operational logging rather than a pre-cleaned analytical dataset, requiring extensive quality checks before analysis. We assessed missing values and found that return timestamps were absent in some transactions, corresponding to items still on loan and not yet returned to the library, while user identifiers were missing in 6.7%. We then verified temporal consistency and plausibility by checking that return dates never preceded borrowing dates, derived loan durations and days-late values were non-negative, and that almost all loans complied with the library's maximum borrowing period. Under library rules, the maximum borrowing period is 28 days with up to six extensions of 28 days each, yielding a maximum loan period of 196 days. To accurately assess compliance with this limit, we incorporated a dataset documenting the library's opening days, counting loan durations only on days when the library was operational.

A small number of extreme outliers exceeding this threshold were identified as erroneous and excluded. Genuinely overdue loans beyond the standard period were retained as valid instances of late returns.

Based on these checks, we applied a set of cleaning rules, of which the most influential were:

- Removal of transactions with missing return timestamps, which represent items still on loan at the time of data export and cannot be used for analyses of completed borrowing cycles. This primarily affected data from 2025 (nearly 50 000 transactions still active), while 2024 had only about 90 such cases.
- Exclusion of transactions linked to library staff, institutional accounts, or system users, as these do not reflect typical user borrowing behavior.
- Exclusion of loans with temporal inconsistencies or implausibly long borrowing periods, while retaining genuinely overdue loans.

The orange line in Figure 1 shows that the share of removed records remained stable across most years at around 1.5%, with a slight peak in 2021 due to irregular loan patterns during the pandemic and a pronounced increase to 15.14% in 2025 driven primarily by the large number of loans still active at export. Our cleaning approach prioritized retaining as much valid data as possible while removing only erroneous or incomplete entries.

2.2. Analysis Methods

To shift from item-level borrowing records to user-level behavioral analysis, we aggregated individual borrowings into user-specific sessions. Each record represents a single item, whereas users typically borrow multiple items in a single library visit. We thus define a session as all borrowings by the same user on the same calendar day Sessions were ordered chronologically per user, and each was assigned a session index as a proxy for accumulating experience. Session-level indicators were then derived via aggregation. A session was classified as *late* if *any* item was returned after its due date, and as *extended* if *any* item received a loan extension. This conservative threshold marks the session according to the user's behavior at that experience level. Even one late return or extension shows their actions linked to that session. This session-based representation provides the foundation for all subsequent analyses.

We examined behavioral adaptation over time using the session index as a proxy for user experience. For each experience level k, we computed the proportion of late sessions and sessions with extensions across all users who reached at least k sessions, ensuring each user contributes at most one observation per level. Let $L_{u,k}$ indicate whether the k-th session of user u contains at least one late item. The late-return learning curve is then

$$\hat{p}_L(k) = \frac{1}{|U_k|} \sum_{u \in U_k} L_{u,k}$$

where U_k is the set of users with at least k sessions. We computed the extension curve analogously.

To analyze media-type preferences across user sessions, we define each session's dominant media type as its most frequent category. Sessions with ties were excluded from this analysis. A user's early preferred type is the most frequent media type across combined borrowings in the first k_0 sessions, where we consider $k_0 \in \{1, 5, 10\}$ to assess how the choice of this baseline window affects preference stability. We then compute curves over the session index k showing how often the k-th session matches this early preferred type for each value of k_0.

We quantified uncertainty for learning curves and media stickiness via user-level bootstrap resampling. Given the dependence of observations within users, parametric methods are unsuitable. We resampled users with replacement, reconstructed their session sequences, and recomputed curves for each bootstrap sample. We derived 95% confidence intervals from the empirical distributions across 1 000 bootstrap iterations (Davison & Hinkley, 1997).

3. Results

Borrowing activity at the Tübingen City Library exhibits strong temporal patterns tied to operating hours and user schedules. Figure 2 shows average session counts aggregated into half-hour bins for weekdays (Tuesday–Friday) and Saturdays (Monday/Sunday closed). Each bin represents the mean number of sessions occurring within that time window, averaged across all recorded days.

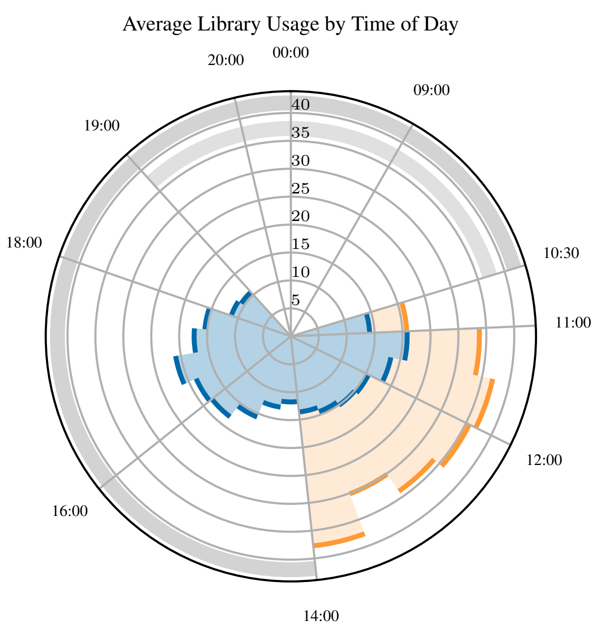

Figure 2. Average library usage by time of day on a 24-hour clock. Mean session counts per time bin are shown for Tue–Fri (■) and Saturday (■). Nighttime hours are visually compressed to reflect library closing times, which are indicated by grey radial bands.

Usage concentrates heavily during daytime hours, with clear peaks from late morning through afternoon. Weekdays show sustained high volume from opening through early afternoon, with a noticeable dip around midday consistent with lunch breaks. Activity rises again in the afternoon, aligning with work and school schedules. Saturday exhibits a later start and more compressed activity, suggesting leisure-driven rather than routine use. These aggregate patterns reveal shared daily rhythms shaped by operational hours and typical schedules, though individuals vary considerably in their regularity.

Among users with at least 10 sessions, the mean probability of returning on the same day of the week is 40.4%, indicating moderate day-of-week preferences, with some showing stronger consistency. Visit times within a day are more dispersed, with a mean standard deviation of 2.1 hours across users, suggesting most do not restrict themselves to narrow time windows. However, approximately 3% exhibit high precision, with standard deviations below 1 hour, indicating they return consistently at similar times of day.

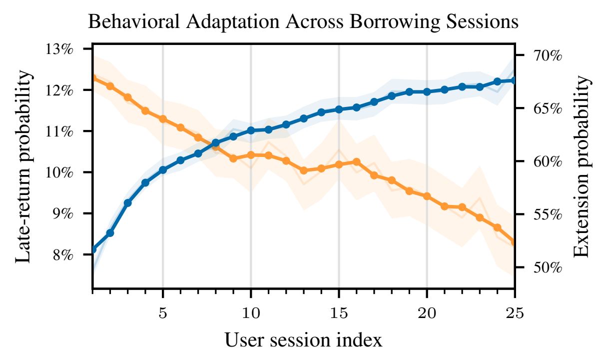

Figure 3. Session-based late-return (-●- left axis) and extension (-●- right axis) probabilities across the user session index. Shaded bands denote user-level bootstrap confidence intervals, while thin lines show raw estimates and thicker lines rolling averages.

Beyond these patterns, we observe learning effects in how users manage late returns and loan extensions. Figure 3 shows both probabilities against the session index, tracking users from their first visit through their 25th session. Late-return probability shows a sharp decline over the first nine sessions, dropping from approximately 13% to about 8%, indicating rapid acquisition of library procedures such as due dates and reminder systems. Beyond session 10, the decline slows, suggesting users have largely adapted to library procedures though learning effects continue.

Extension probability exhibits the inverse pattern, rising from 50% to over 65% during this window. This increase reveals a shift in user strategy: rather than becoming more punctual, experienced users proactively request extensions to avoid lateness. The diverging trajectories suggest that learning manifests as adaptive behavior within library constraints rather than perfect compliance. Late returns persist among experienced users, indicating structural constraints rather than procedural unfamiliarity.

Finally, we examine how users' media-type preferences evolve across sessions. Figure 4 shows, for the early preference window $k_0 \in \{1, 5, 10\}$, the share of sessions whose dominant media type matches the user's early preferred type, plotted against the session index. Across all three curves, consistency decreases with increasing session index, with the strongest decline for smaller k_0.

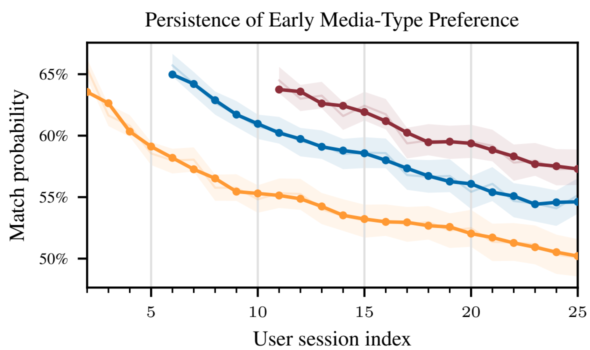

Figure 4. Consistency relative to the dominant media-type across the user session index for $k_0 = 1$, $k_0 = 5$ and $k_0 = 10$. For each k_0, the curve starts at session $k_0 + 1$. Shaded bands denote user-level bootstrap confidence intervals, while thin lines show raw estimates and thicker lines rolling averages.

For $k_0 = 1$, users match their first-session dominant media type in about 64% of second sessions. By session 25, this value drops to roughly 50%, a decrease of about 14 percentage points. For $k_0 = 10$, the decrease is about 7 percentage points, and the curve remains close to 60% in later sessions. This is consistent with the fact that choosing a larger k_0 defines the early preferred type from more observations, making it more robust and less sensitive to variability in the first sessions. Overall, the curves indicate a gradual shift away from the early preferred type, leveling off at higher session indices. Over time, the early dominant type becomes less predictive of later sessions, while still indicating a persistent preference.

4. Discussion & Conclusion

Our analysis of borrowing records from the Tübingen City Library reveals several behavioral patterns. Most notably, late return behavior is not primarily driven by persistent user unreliability, but rather by lack of experience during early library engagement.

Late return probability declines sharply from approximately 13% in first sessions to close to 8% over the initial sessions, while extension requests rise from 50% to over 65% during the same period. This diverging pattern indicates that users learn to navigate library procedures proactively rather than simply becoming more punctual. Even among experienced users, late returns persist due to structural factors beyond user knowledge or behavior.

These findings offer practical insights for the Tübingen City Library. Overdue risk is concentrated in early sessions, suggesting that targeted communication and reminder systems for new users may be more effective than uniform enforcement mechanisms. For instance, automated reminders or welcome materials explaining borrowing procedures could be specifically targeted at users during their first visits. The increasing use of extensions indicates that users value this system, and ensuring it remains accessible could further reduce involuntary late returns.

Beyond late returns, we observed that borrowing activity concentrates during daytime hours aligned with work and school schedules, while media-type preferences remain influential but become less predictive over sessions as experienced users expand their borrowing range. These patterns reveal how users adapt their behavior to library operations over time. Understanding these temporal patterns could inform staffing decisions and resource allocation during peak usage hours. Additional analyses available in the accompanying repository examine factors such as media type, user categories, and seasonal patterns that influence late return behavior.

Several limitations constrain our analysis. The dataset covers only 2019 onwards due to earlier system conventions, and user behavior can only be observed indirectly through transactions. Contextual factors such as whether items are borrowed for oneself or others remain unobserved.

Nevertheless, this work demonstrates the value of systematically analyzing library circulation data to support informed operational decisions. The session-based representation and preprocessing framework established here provides a foundation for further analysis, such as evaluating the effectiveness of reminder systems or examining policy changes over time. By revealing learning effects and behavioral adaptation patterns, libraries can move from reactive enforcement toward proactive support systems.

Contribution Statement

Nico Rinck analyzed media-type preferences and their relationship to user behavior. Jonas Mahr examined users' interest categories. Adriano Polzer investigated temporal patterns in user behavior. Robin Allgeier obtained the raw data and conducted data quality assessment and cleaning. Jannik Rombach developed the session-based framework and analyzed learning behavior. All authors contributed to the report preparation.

References

Davison, A. C. and Hinkley, D. V. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

Phelps, S. F. and Campbell, N. Library fines make a difference in academic library book return behaviour. *Evidence Based Library and Information Practice*, 10(3):96–98, 2015. doi: 10.18438/B8H89K.