Analysis Summary

Elisei Shushpanov

Approach:

My approach followed a 5-part pipeline: (1) Error Diagnostics, (2) Baseline (LOW) Modeling, (3) LIFO Stack Modeling, (4) Timeout Capping Sensitivity Analysis, and (5) Final Reporting.

Assumptions

- **Data Integrity:** The analysis assumes the provided `toolwindow_data.csv` is the complete dataset for this task. Timestamps were converted from milliseconds to datetime objects, and all events were sorted chronologically by `user_id` and timestamp to build accurate user timelines.
- **Open-Open Events as Nested Sessions:** The initial diagnostics (Part 1) revealed 209 instances of consecutive "opened" events (O-O). O-O events were treated as valid nested sessions.
- **Modeling Strategy:** A simple Last-Open-Wins (LOW) model was established as a baseline (Part 2). A LIFO (Last-In-First-Out) Stack model was chosen as the primary, authoritative method (Part 3) because it is designed to correctly handle the nested sessions identified in the data.

How you handled errors or messy data

The data was first profiled for quality (Part 1) to inform the sessionization strategy.

- **Orphaned Events:** The diagnostics identified 8 close without open events and 235 open without close events. The LIFO stack model inherently handles both:
  - Orphaned closed events are discarded as there is no opened event on the stack to match with.
  - Orphaned opened events (trailing opens at the end of a user's timeline) are left unpaired on the stack and are not included in the final duration analysis, as their true duration is unknown.
- **Consecutive Events (O-O and C-C):**
  - O-O (209 events): As mentioned, these were assumed to be nested sessions, not errors. The LIFO model correctly pairs the first closed event with the last opened event.
  - C-C (16 events): Consecutive closed events are handled by the LIFO logic. The first closed event will pair with the last opened event; the second closed event will be discarded as an orphan because the stack is now empty (or its corresponding open event is deeper in the stack).

Strategy for matching open/close events and calculating durations

Two strategies were implemented to reconstruct sessions:

- **Baseline (LOW):** Implemented in Part 2, this model maintains a single "open" state. If a new opened event occurs before a closed event, it overwrites the previous opened event. This method is simple but incorrectly handles nested sessions.
- **Primary (LIFO Stack):** Implemented in Part 3, this model was chosen as the authoritative method.
  - When an opened event occurs, its data (timestamp, `open_type`) is pushed onto a stack.
  - When a closed event occurs, the most recent opened event is pop()-ed from the stack. This open-close pair forms a complete session.

  This LIFO approach correctly resolves nested O-O-C-C patterns into two distinct sessions.

For both models, `duration_seconds` was calculated by subtracting the `open_timestamp` from the `close_timestamp`. Only sessions with a duration greater than 0 were kept for analysis.

Findings:

All findings are based on the robust LIFO Stack model (Part 3).

Summary statistics for "manual" and "auto" opens

The LIFO model identified 1,630 valid sessions (1,005 'auto', 625 'manual'). The duration data is highly right-skewed (skewness: auto=8.0, manual=11.2), making the median the most reliable measure of central tendency.

Open Type	Session Count	Mean (sec)	Median (sec)	Std. Dev. (sec)	Min (sec)	Max (sec)
auto	1,005	6,952.1	185.8	31,280.6	0.15	409,873.7
manual	625	1,672.8	12.3	11,817.0	0.02	180,918.7

Table 1: Statistics derived from LIFO sessionization (Part 3)

Automated openings result in sessions that are, at the median, 15 times longer than manual openings (185.8s vs. 12.3s).

Measure of confidence

A two-part statistical analysis was performed:

- **Mann-Whitney U Test (Non-parametric):** Due to the high skew and non-normal distribution of the raw duration data, the Mann-Whitney U test was used.
  - P-value: 6.135e-63. This is far below the 0.05 threshold, indicating the difference between the two groups is statistically significant.
  - Effect Size (CLES): The Common Language Effect Size (CLES) was 74.6%. This means there is a 74.6% probability that a randomly selected 'auto' session will have a longer duration than a randomly selected 'manual' session.
- **Sensitivity Analysis (Timeout Capping):** To ensure the result was not driven by extreme outliers (e.g., sessions lasting days), a sensitivity test was performed (Part 4).
  - All durations were capped at the 99th percentile (99,948.95 seconds, or ~27.76 hours). This affected 17 (1.04%) of the sessions.
  - The capped data was log-transformed (`np.log1p`) to normalize it.

– A Welch's t-test (which does not assume equal variance) was performed on this capped, log-transformed data.

– P-value (Welch's t-test): 4.006e-61.

The statistical significance holds, confirming that the observed difference is a core pattern in the data and not just an artifact of a few extreme outliers.

Visualization

Session Duration Histogram (Part 3)

A histogram of the final `log_duration` for both 'auto' and 'manual' LIFO sessions, showing two clearly distinct distributions.

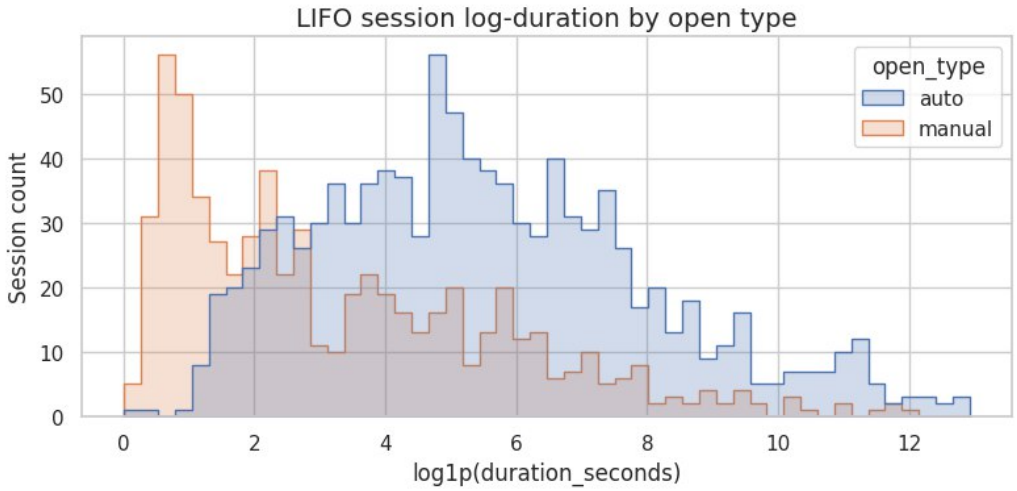

Figure 1: Session Duration Histogram (Part 3)

Session Duration Boxplot (Part 3)

A boxplot of the `log_duration` by `open_type` provided a clear comparison of the median and interquartile ranges (IQR) for both groups, visually confirming the statistical findings.

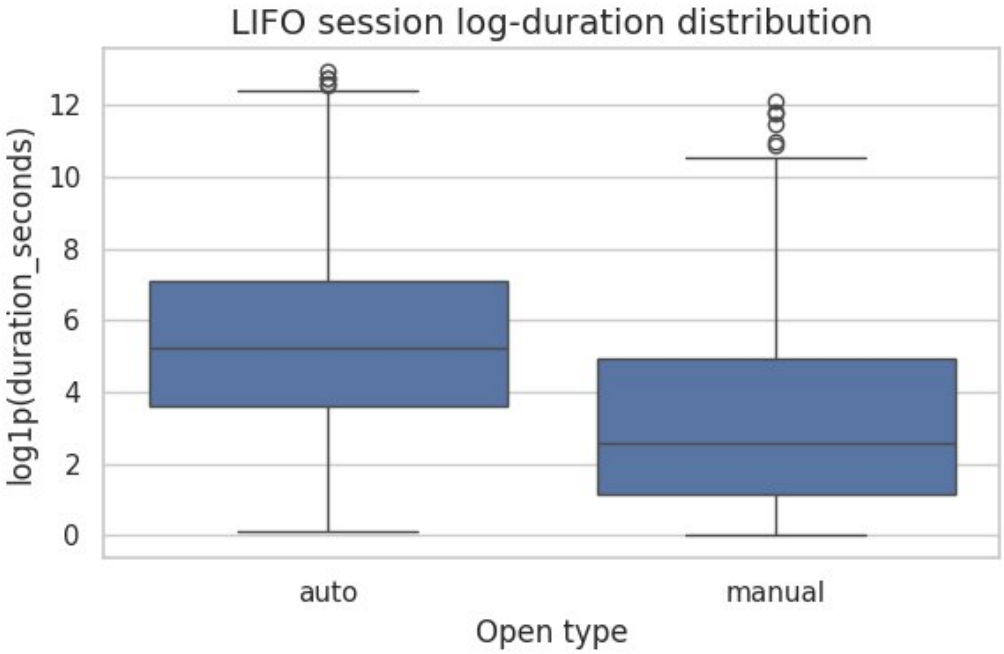

Figure 2: Session Duration Boxplot (Part 3)