ToolWindow Usage Analysis Report

Viktor Shishmarev

October 24, 2025

1 Objective

Research Question: Is there a statistically significant difference in tool window session duration between manual and automatic opening methods?
Dataset: Event logs with 3,503 raw events tracking tool window activity (user ID, timestamp, event type, opening method).
Hypothesis:

- H_0: No difference in duration between manual and auto groups
- H_1: Significant difference exists between groups

2 Methodology: Data Cleaning & Anomaly Detection

2.1 The Challenge: Messy Real-World Data

Raw event logs contained significant data quality issues:

- Orphaned events (close without open, open without close)
- Duplicate opens, out-of-order timestamps
- Missing or invalid type values
- Unrealistic durations ($>$ 12 hour threshold)

2.2 Our Approach: Smart Episode Reconstruction

Key Innovation: User-level state machine for open/close event matching.
Algorithm:

1. **Load & Validate**: Multithreaded CSV import to SQLite (producer-consumer pattern)
2. **Sort by User**: Group events by `user_id` and order chronologically
3. **State Machine Matching**:
   - Track current "open" state per user
   - Match each "close" with corresponding "open"
   - Detect anomalies: duplicate opens, missing pairs, timestamp violations
4. **Bifurcation**:
   - **Clean episodes** → `clear` table (valid open-close pairs)
   - **Anomalies** → `anomaly` table (with reason: `missing_open`, `duplicate_open`, `>duration`, etc.)
5. **Duration Calculation**: duration $= \frac{\text{close\_ts} - \text{open\_ts}}{60000}$ (minutes)

Result: From 3,503 raw events → 1,574 clean episodes (617 manual, 957 auto)

2.3 Statistical Testing

Mann-Whitney U Test: Non-parametric comparison (no normality assumption, robust to outliers)
Cliff's Delta: Effect size measurement

$$\delta = \frac{\#\{x > y\} - \#\{x < y\}}{n_1 \cdot n_2}$$

where $\#\{x > y\}$ = number of pairs where manual > auto, n_1, n_2 = sample sizes
Interpretation: $|\delta| \geq 0.47$ = large effect

3 Results

3.1 Data Processing Summary

- Raw events: 3,503 → Clean episodes: 1,574
- Anomalies filtered: 352 — demonstrates importance of cleaning methodology
- Manual episodes: 617 — Auto episodes: 957

3.2 Descriptive Statistics

Metric	**Manual**	**Auto**
Count	617	957
Median (min)	**0.20**	**2.67**
Mean (min)	10.52	27.85
Q1 - Q3 (min)	0.04 - 2.22	0.53 - 15.23
Std Dev (min)	50.17	79.18

Table 1: Duration Statistics by Opening Type

3.3 Statistical Significance

Test	**Result**
Mann-Whitney U	151,898.50
p-value	< 0.001 **(highly significant)**
Cliff's Delta	-0.4855 **(large effect)**
Conclusion	**Reject** H_0

Table 2: Statistical Test Results

Interpretation: Manual durations are significantly and substantially lower than auto durations.

3.4 Visualizations

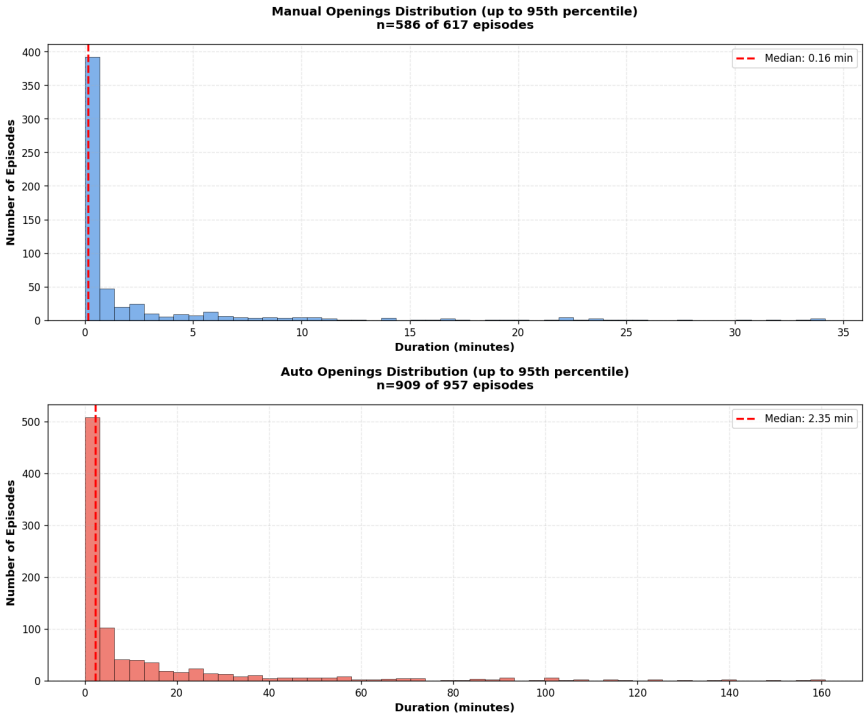

Figure 1: Duration Distribution Comparison

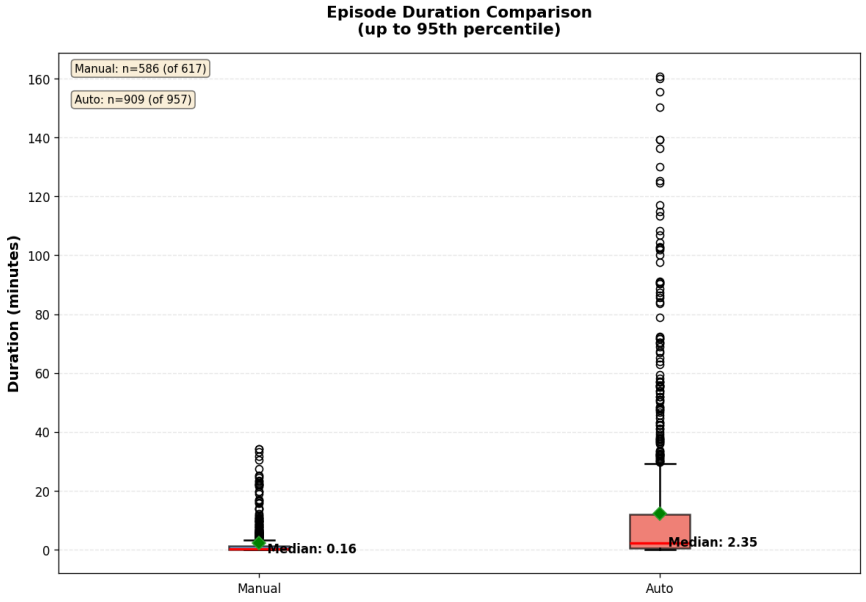

Figure 2: Box Plot: Quartiles and Outliers

4 Key Insights

4.1 Main Finding

Manual openings are 13× faster to close (median): 0.20 min vs 2.67 min
Why?

- **Manual**: Task-focused, quick information lookup → close immediately
- **Auto**: Workflow-integrated (debugging, testing) → stay open during entire task

4.2 Methodological Success

- Without state-machine matching, analysis would be impossible
- Bifurcation strategy (clean vs anomaly tables) ensures data integrity
- User-level processing prevents cross-contamination

5 Conclusions

Question	Answer
Significant difference?	**YES** — $p < 0.001$, large effect ($\delta = -0.4855$)
Practical impact?	**Manual 13× faster to close** (0.20 vs 2.67 min median)
Business insight?	Manual = quick lookup; Auto = workflow integration

---

Technology Stack: Python 3.12+ — SQLite3 — scipy — matplotlib — FastAPI — Docker
Repository: `https://github.com/shishmarevv/ToolWindowData` — **Reproducible:** `make all`