ANALYSIS OF TOOLWINDOW USAGE
ACTION PLAN

DATA EXPLORATION

- Data covers the activity of 205 users between 2025-07-03 and 2025-07-23 (20 days and 1 hour).
- Number of events captured: 1865 openings and 1638 closings of the tool window.
- In the “open” type, all of the closed events have NaN, and none of the opened events have NaN (as expected).

ASSUMPTIONS

- Each log represents **only one** event happening - either opening or closing of the window.
- There is no “in between” logs - user’s open corresponds to their next close and make up **complete episode**.
- Users may have incomplete or overlapping sessions - the code deals with these.
- Timestamp is an **epoch milliseconds**.

DATA CLEANING

- Orphan close events - ignored as we can’t determine their duration
- Multiple openings - the second one is treated as closing the first event and. opening the next
- If open doesn't have its close before the end it is ignored
- Timestamp transformed into pd.to_datetime().
- The “nan", "na", "none" replaced with np.nan.
- Filtering the dataset so that if duration is negative or zero the episode is ignored.

STRATEGY FOR MATCHING EVENTS AND CALCULATING THE DURATIONS

- We need to create a new data frame containing user_id, opening timestamp, closing timestamp and duration.
- First thing first, I created a group_by user_id as every user will have their own sequence of episodes. This data needs to be sorted so that chronology is preserved.
- Then it’s time to check the open_type and follow the following instructions:
    - iIf it is "opened", record the timestamp as t_open and the type as otype (manual or auto).
    - f it’s "opened" but t_open is not None (two consecutive openings without a closing), treat this timestamp as the closing time for the first session and the opening time for the next.
    - If it’s "closed" and t_open was recorded, then add the record containing what we stated at the beginning.
    - If it’s "closed" and t_open is None, then the session wasn’t started in our dataset and we ignore it.

VALIDATION

- Filtered so that duration is **positive**.
- Visualized distributions. → next page.

	count	mean	std	min	25%	50%	75%	max
open_type								
auto	1180.0	17464.653237	61726.337743	0.154	47.1090	285.4165	2382.6035	728695.209
manual	651.0	4640.452693	32058.624031	0.015	2.3935	13.9980	161.9940	447314.548

JETBRAINS Internships

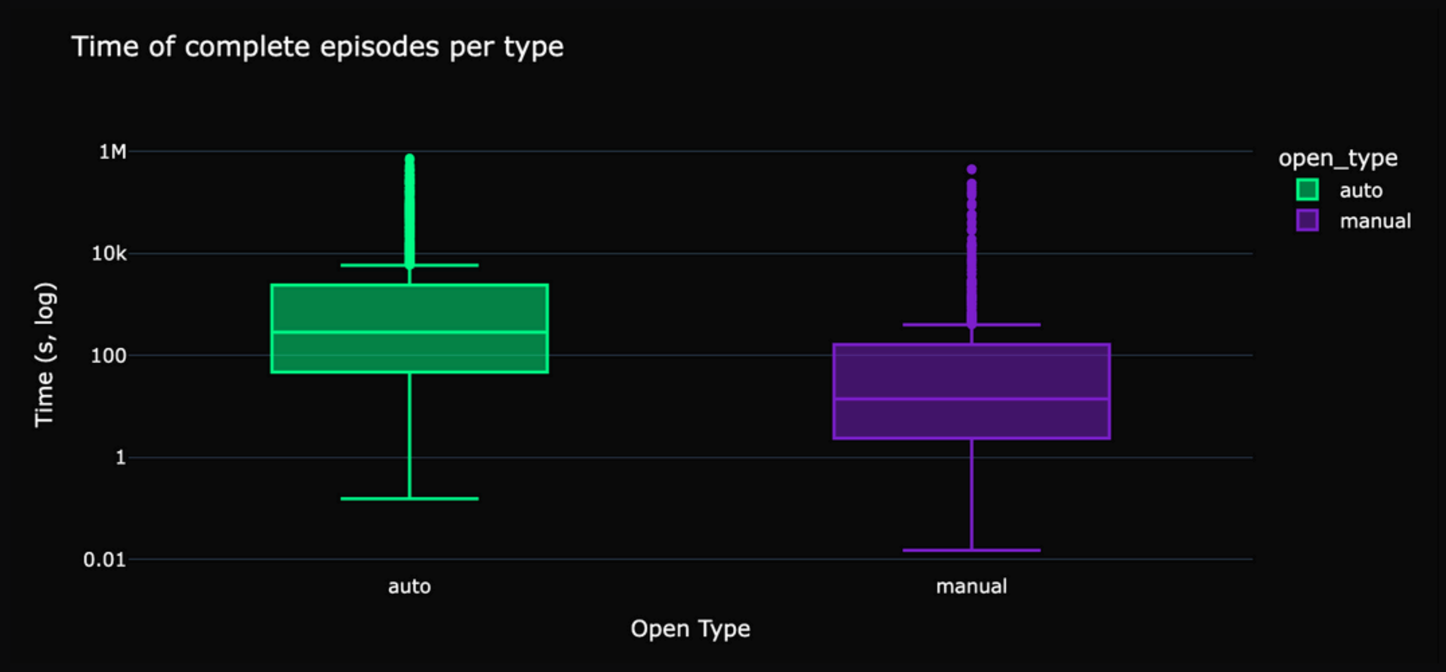
Author:
Aleksandra Mulewicz

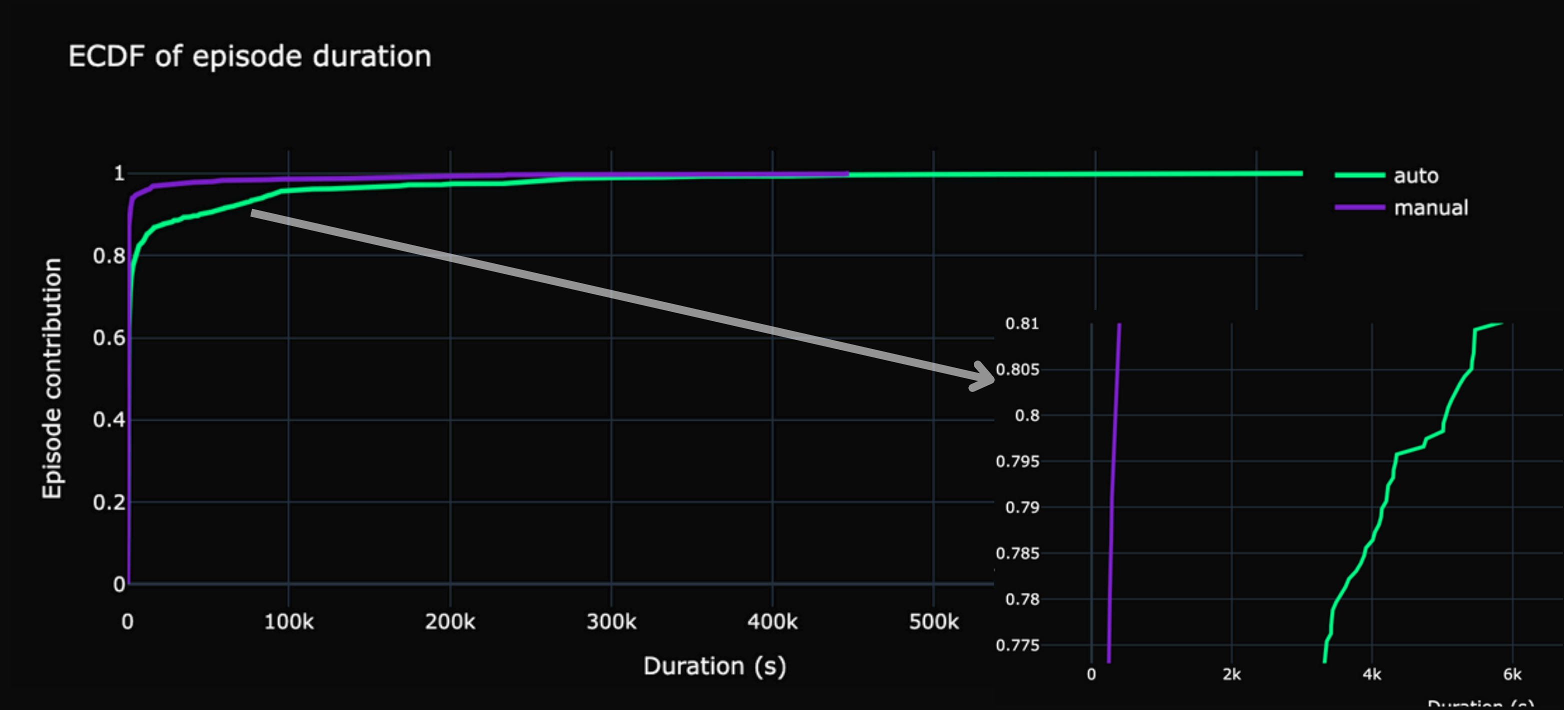
ANALYSIS OF TOOLWINDOW USAGE

VISUALISTION

DURATION FOR EACH TYPE

The plot illustrates the duration of the tool window episodes depending on how window was opened - either manually or automatically. The first thing we can notice is that the median of the auto is much higher than that of the manual. (285 and 14 seconds). On average the automatically opened window tends to stay open longer and the distribution of the automatic process is much wider which indicates more variation.



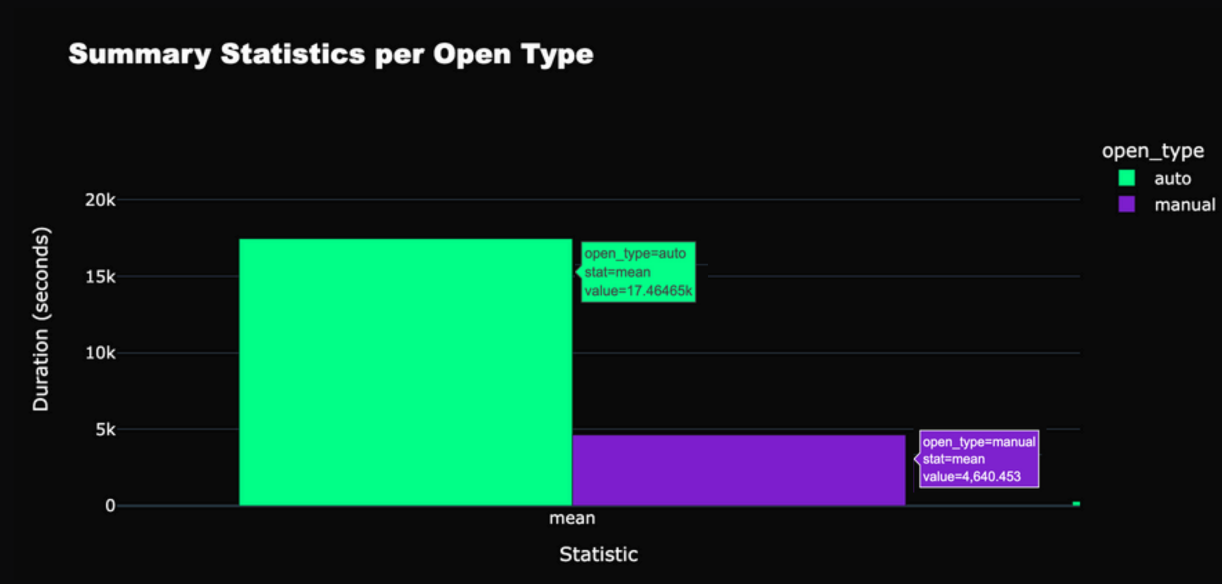


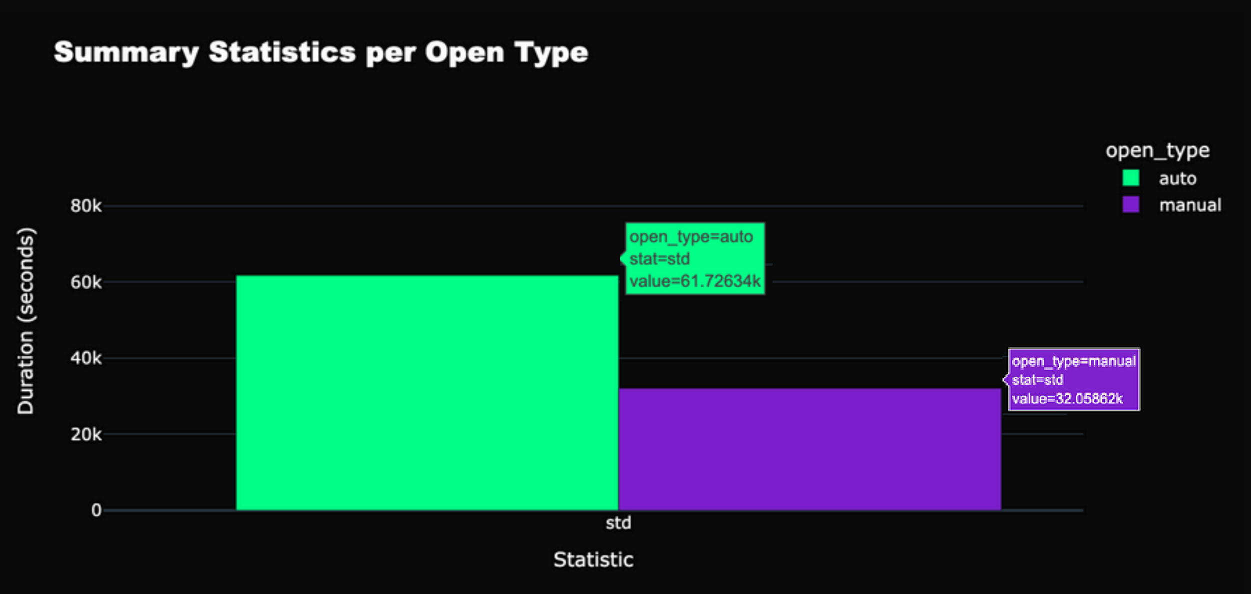
ECDF (empirical cumulative distribution function)

Both variables rise very quickly and reach 1 relatively early which indicates that most sessions are short with around 80% being shorter than 5k for auto and 1 k for manual. Manual at the beginnig grows slightly faster which illustrates shorter mean duration. The green line (automatic sessions) reaches further so the variation is much wider — including both very short and very long sessions.

MEAN DURATION TIME PER OPEN TYPE

Just as we expected the mean of the duration of the tool window automatic episode is much higher than that of the manual. Automatic type on average was opened 4 times longer than manual. This difference makes intuitive sense: automatic windows might remain open in the background, while manual engagements are intentional and task-specific, typically closed once the user finishes the action.



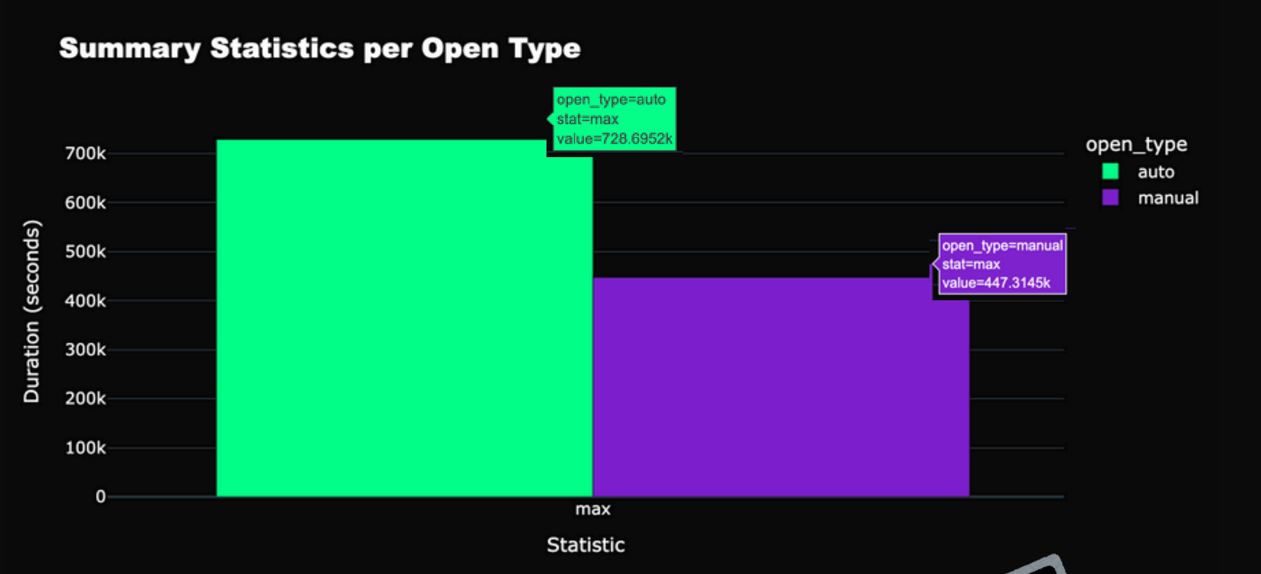


THE VARIATION OF THE EPISODES

The variation (spread) in episode duration is also significantly higher for automatic. This can be explained by the dual nature of automatic processes — some are brief, triggered momentarily, while others last for extended periods due to background monitoring or idle waiting states. The distribution for automatic openings is also much wider, suggesting greater variability in system behavior, possibly due to background tasks that vary in complexity or duration.

MEAN DURATION TIME PER OPEN TYPE

Just as we expected the max duration of the tool window automatic episode is much higher than that of the manual. Such extreme values likely occur when background processes stay open indefinitely — for example, a tool window automatically triggered by the system but never explicitly closed. Manual sessions rarely reach such extremes because user attention and action naturally limit their duration.



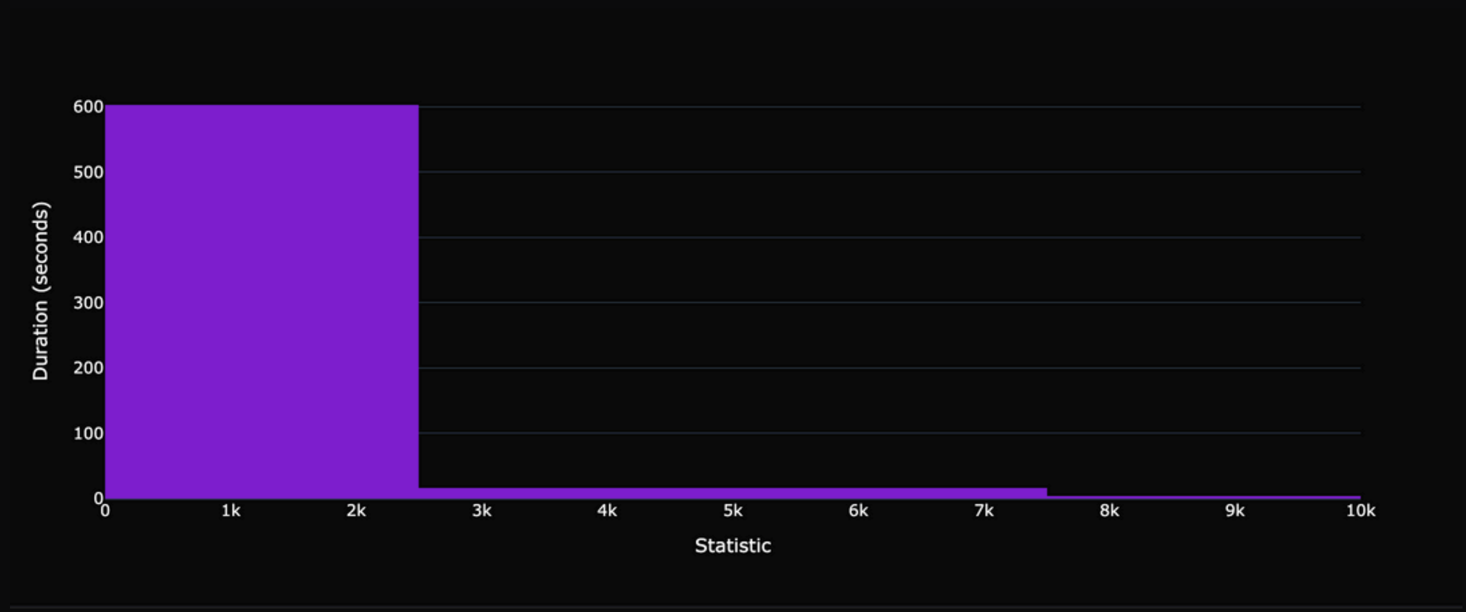
JETBRAINS Internships

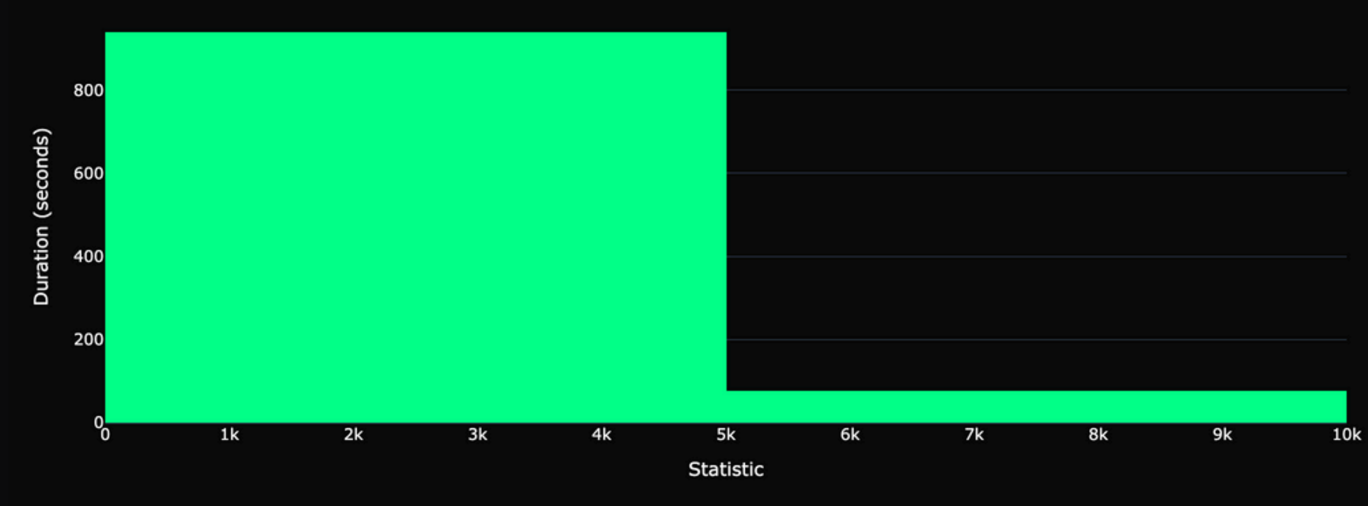
Author:
Aleksandra Mulewicz

ANALYSIS OF TOOLWINDOW USAGE

STATISTICALLY SIGNIFICANT?

DISTRIBUTION SHAPE





NORMAL DISTRIBUTION?

- To verify whether the distributions of episode durations follow a normal distribution, I created histograms for both open_type. As we can see on the plots, the data is heavily skewed to the right (meaning that most episodes are short, with a few very long outliers), and therefore we have no reason to suspect an underlying normal distribution. That limits our test choices, and parametric tests that assume normality (like the standard t-test) are not appropriate for these data.
- I chose the Mann–Whitney U-test to conduct my research because:
  - It does not assume normality.
  - Observations are independent - a key requirement of the test (opening-type episodes come from different opening mechanisms).
  - duration_sec is continuous and can be ranked - the basis of this test.
  - It is stable even with large outliers - instead of mean, it operates on the median.
  - Does not require equal variances.
- Firstly, I created my null and alternative hypothesis.

H0 (Null Hypothesis): There is no significant difference between the duration of manual and auto.

H1 (Alternative Hypothesis): There is a significant difference between the duration of manual and auto.

LET'S SET SIGNIFICANCE LEVEL TO 0.05

- I calculated two values
  - U = 579923.50, it is a sum of ranks and states how different the two groups' rankings are
  - p = 0.00000 (so small, it's rounded to 0), it is a probability that such a U occurs by chance
- My p value is extremely small therefore there is no doubt about the outcome....

With the significance level set to 0.05, we reject H_0 and accept H_1.

The differences found between manual and auto opening durations are statistically significant with a 95% confidence level.

JETBRAINS Internships

Author:
Aleksandra Mulewicz

ANALYSIS OF TOOLWINDOW USAGE

NEXT STEPS

POSSIBLE IIMPROVEMENTS OF MY APPROACH

PROBLEM Currently, durations are treated uniformly, regardless of when they occur.

POSSIBLE SOLUTION
- Adding time-of-day and day-of-week analysis: are manual openings more frequent during active hours, while auto processes happen overnight?
- Could show whether auto sessions dominate during idle times or system maintenance.

PROBLEM All users are analyzed together.

POSSIBLE SOLUTION
- Clustering users by behavior such as average behaviour or number of manual vs auto-sessions
- Identifying outliers who keep tool windows open unusually long and power users (the ones that use it to its fullest)

PROBLEM Only Mann–Whitney U-test used for duration differences.

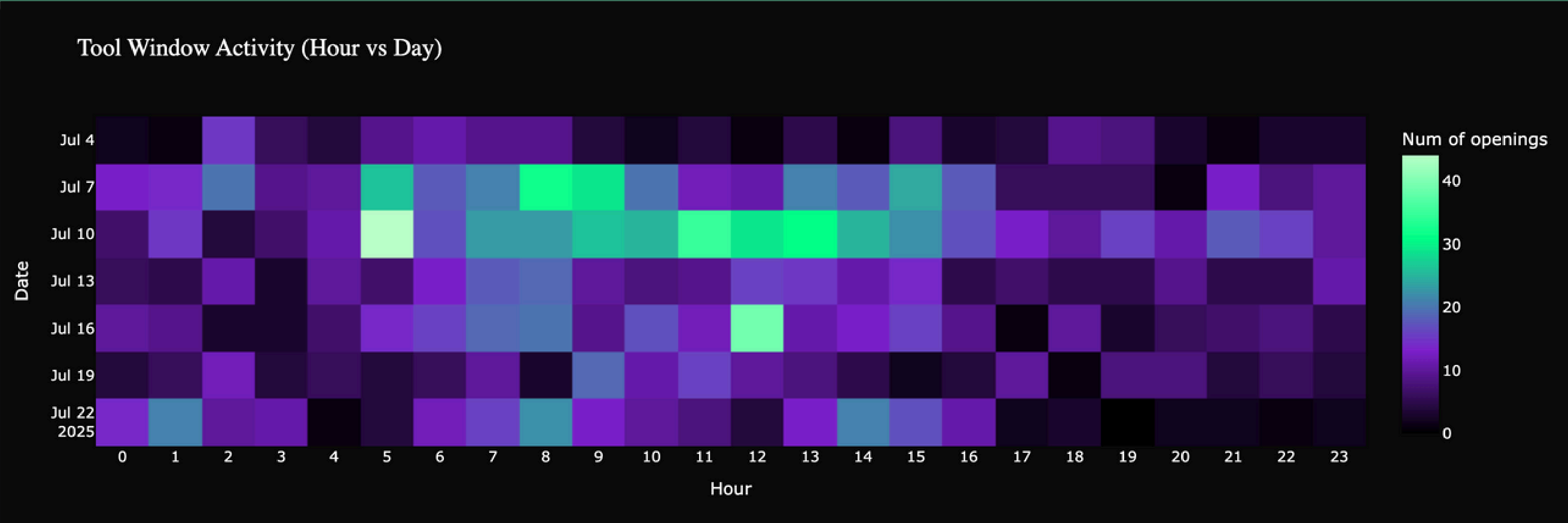
POSSIBLE SOLUTION
- Adding effect size
- Considering survival analysis to model how long windows tend to stay open before closing.

JETBRAINS Internships

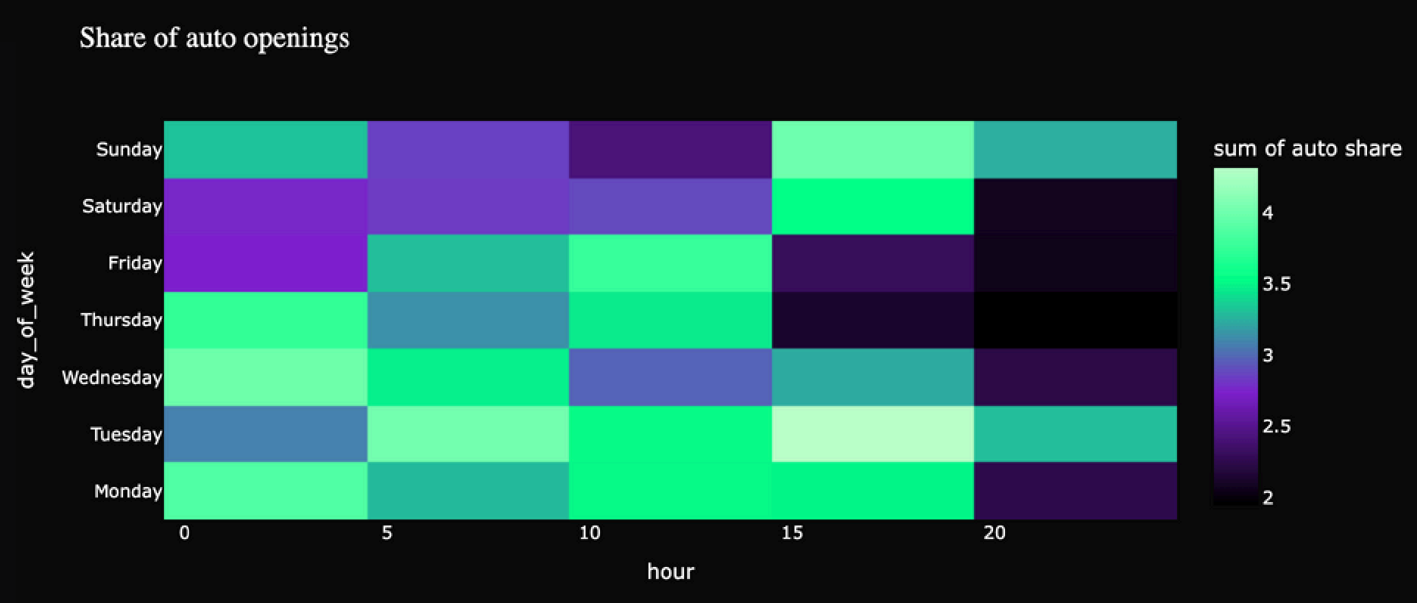
Author:
Aleksandra Mulewicz

ANALYSIS OF TOOLWINDOW USAGE

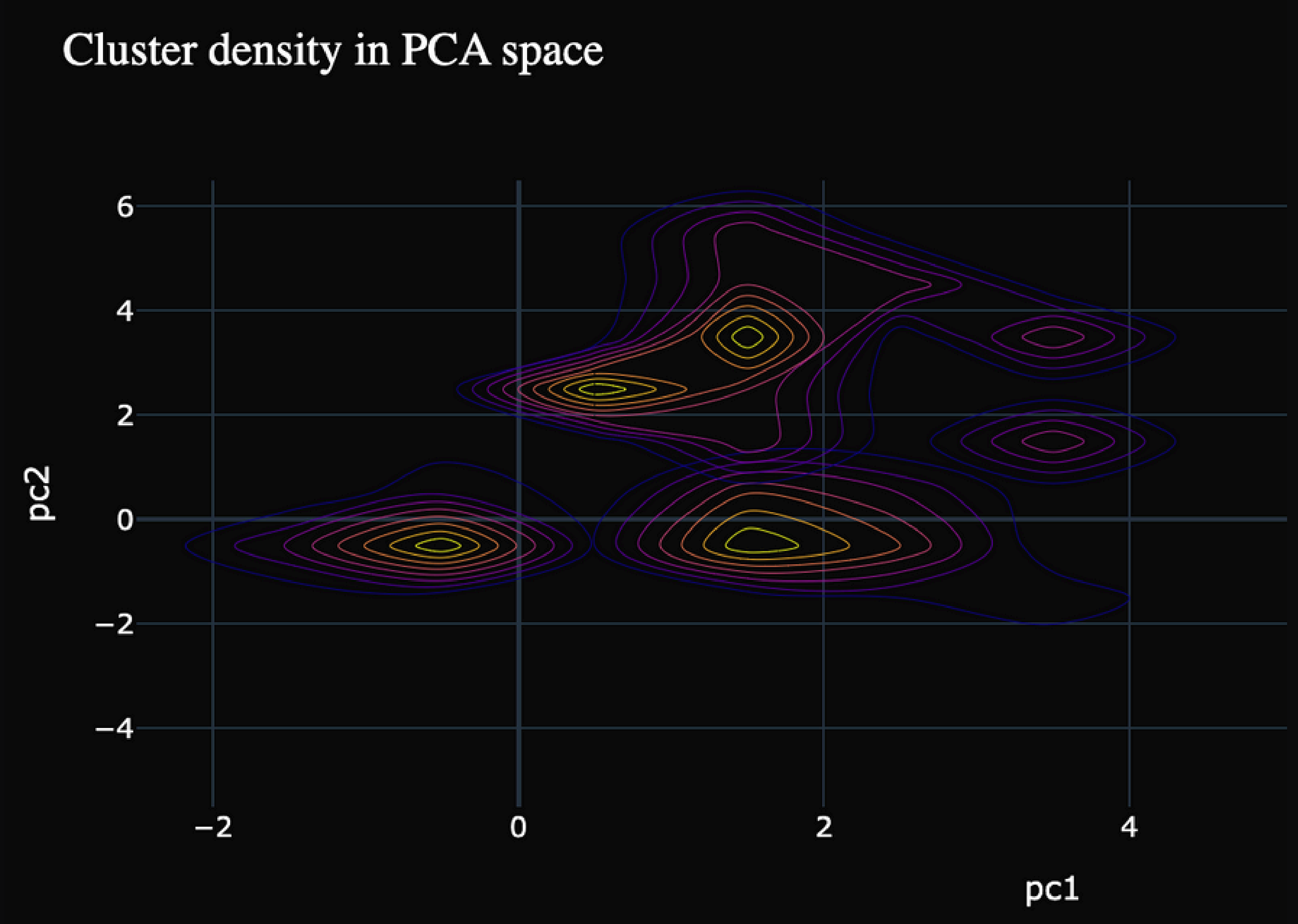
LET'S DO IT: User Behaviour Analysis



The brightest zones (green) cluster around 7 – 14, showing users are most active during standard work hours. Each day follows a similar pattern: chill early morning → high peak mid-morning → drop by evening. That indicates **habitual IDE usage** rather than random spikes.

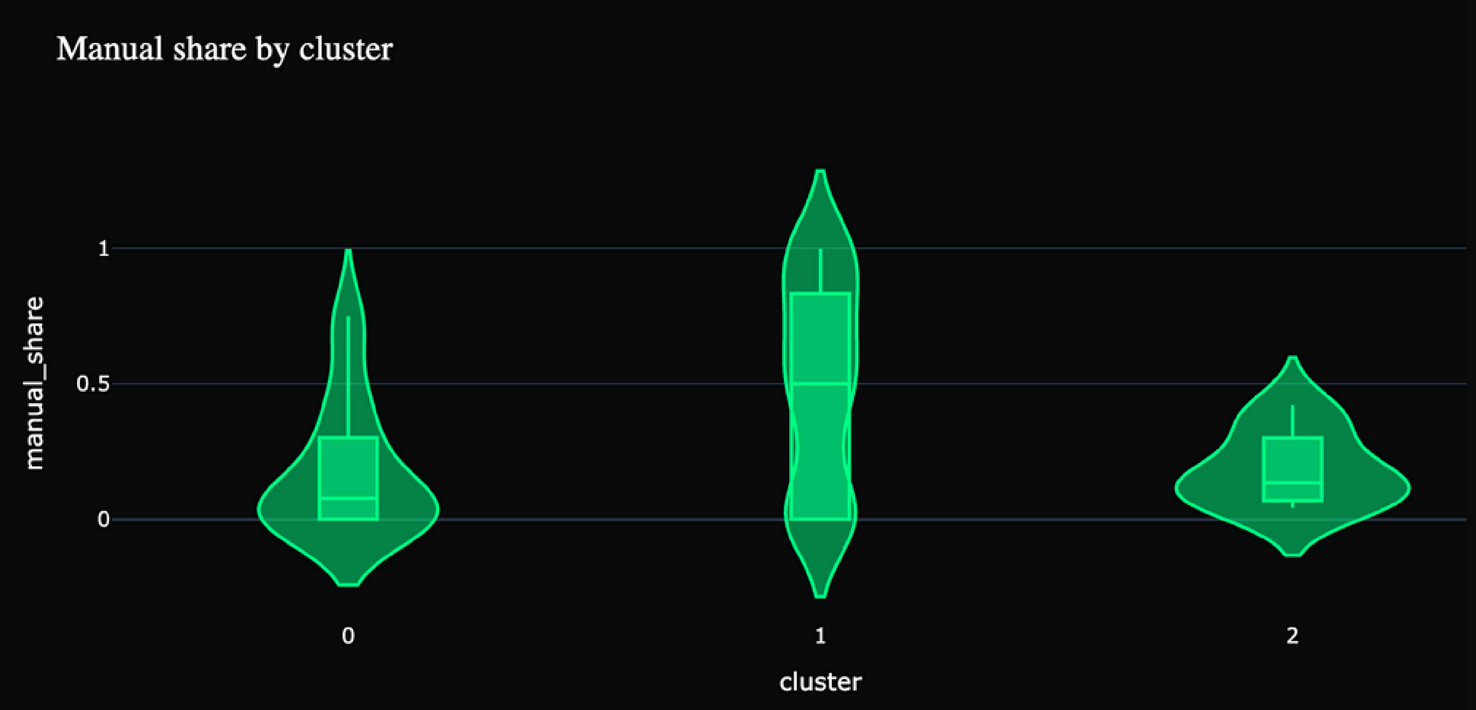


Across all days, the share of auto openings drops sharply after 18:00, confirming that automated events follow working-hours. Most green bands appear on Tuesday to Thursday, particularly around morning to early afternoon. This suggests users engage in more automated workflows midweek



Using PCA, we can identify three main groups of users with similar tool window usage patterns. The horizontal and vertical separation suggests that the main differences between users are caused by session duration, frequency of openings, and the balance between manual and automatic interactions. This confirms that the input data contains enough signal to distinguish user groups so the distribution isn't random.

This boxplot shows that Cluster 0 has the highest average session duration, with several extreme outliers reaching well beyond 300 000 seconds. Clusters 1 and 2 display considerably shorter sessions, most under 10 000 seconds. Cluster 0 likely represents users who keep tool windows open for extended periods while clusters 1 and 2 reflect short, focused, or task-specific sessions.



Cluster 1 stands out with a high median and wide spread of manual share, indicating that these users open most tool windows manually, reflecting interactive, hands-on work. In contrast, clusters 0 and 2 have much lower manual shares, showing a stronger dependence on automatically triggered windows. The shape indicates that cluster 1 demonstrates a clear preference for manual openings, separating it from the other groups as they show signs of mixed behaviour. Cluster 0 has concentration near zero therefore there is a high dominance of automated sessions

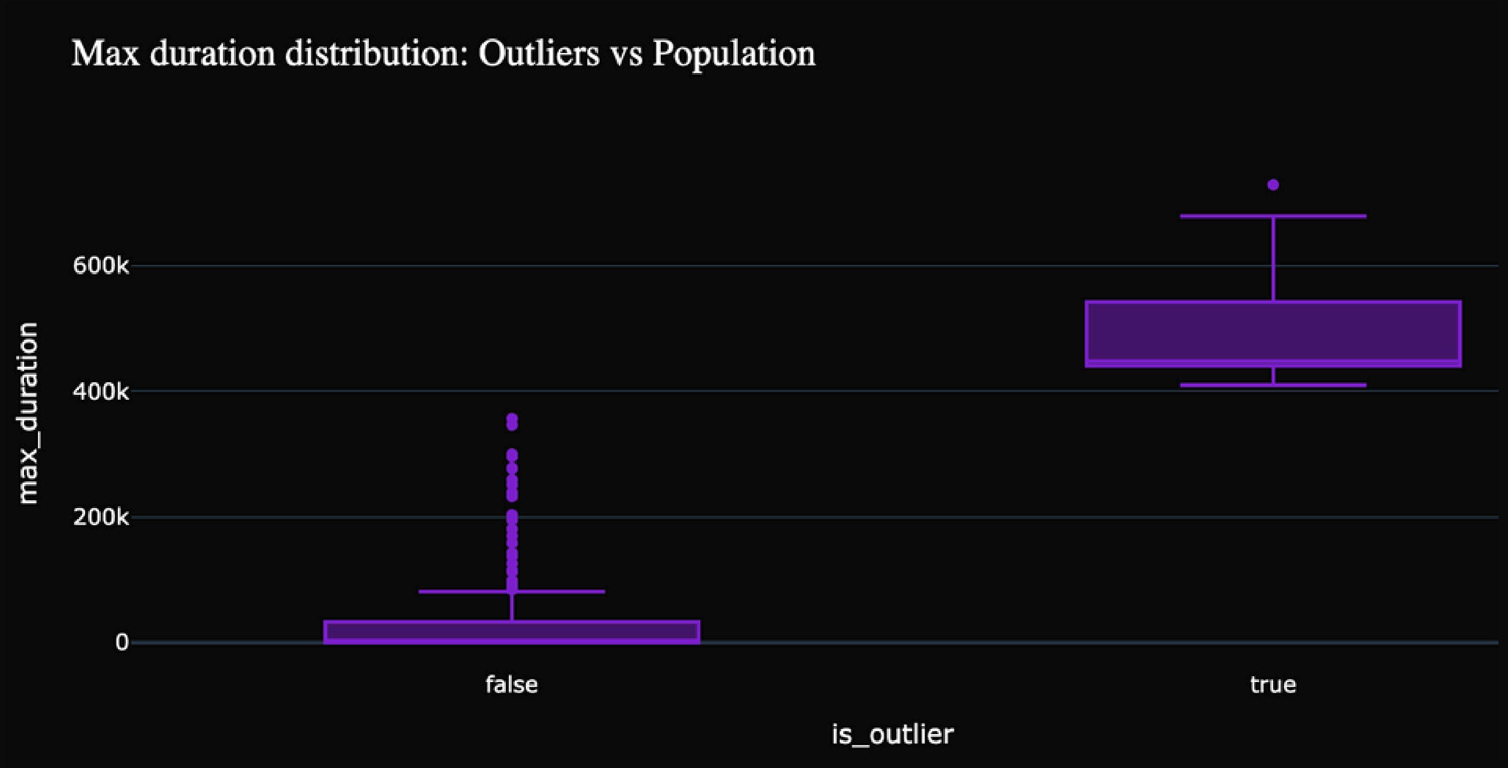
JETBRAINS Internships

Author:
Aleksandra Mulewicz

ANALYSIS OF TOOLWINDOW USAGE

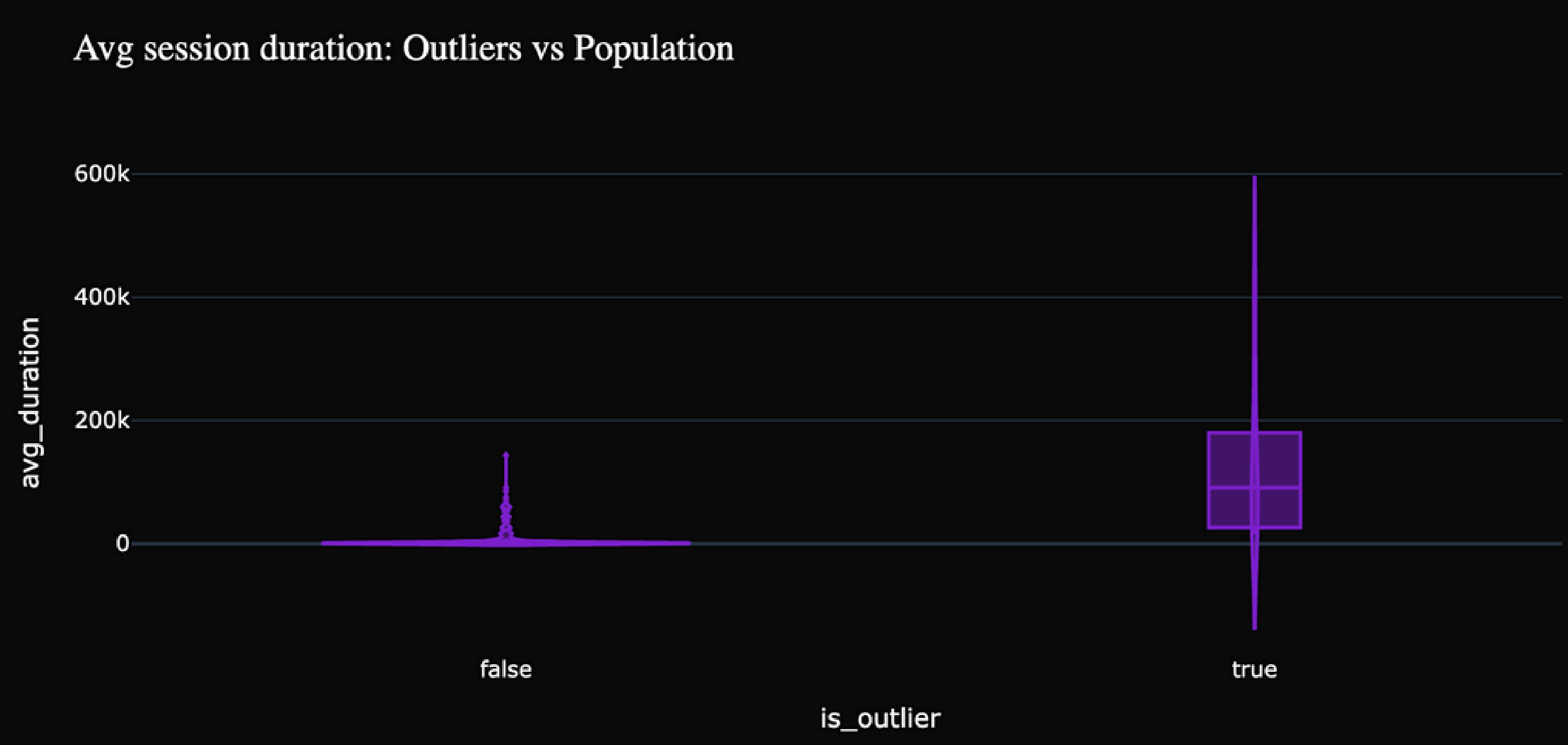
Outliers and power users identification

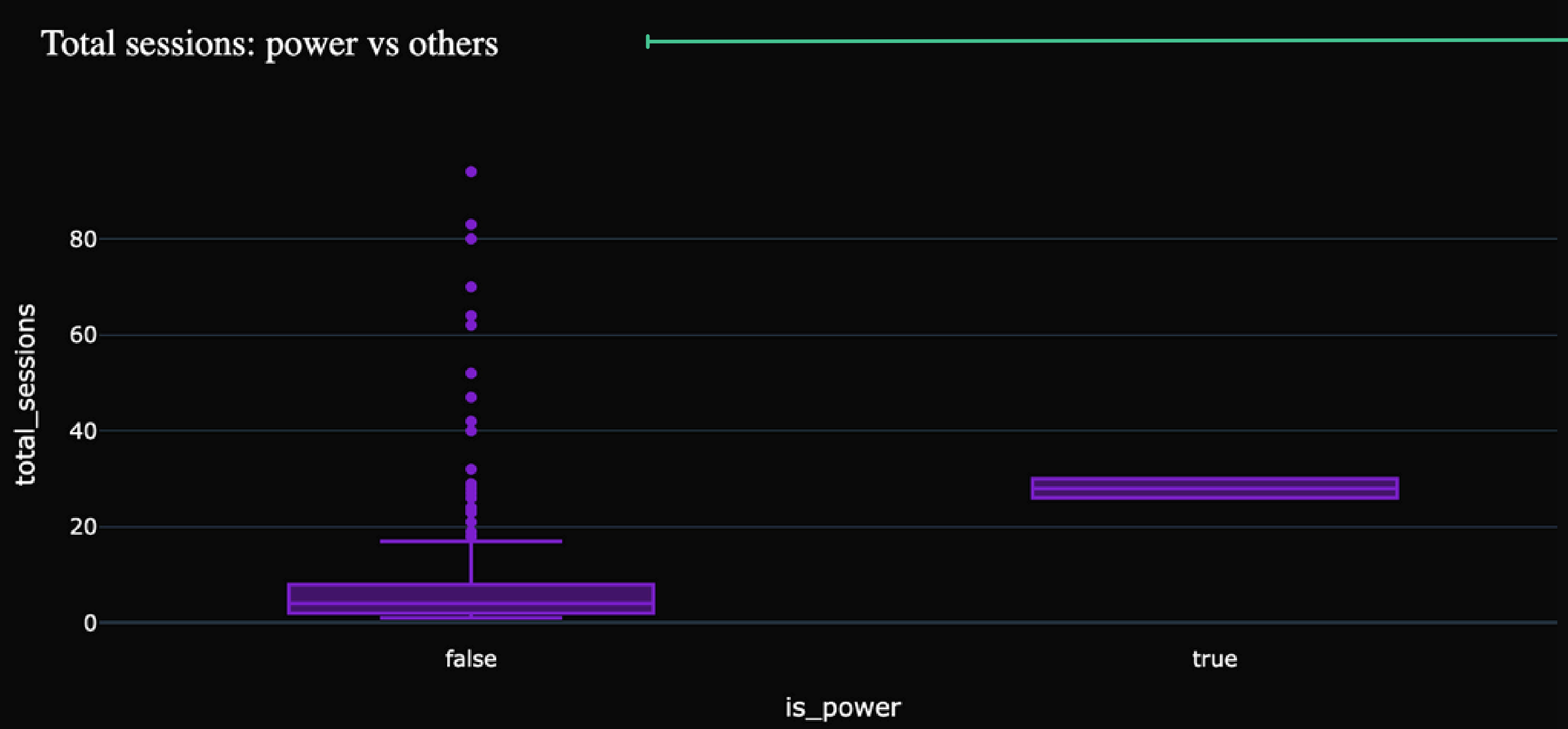
- **Outliers** (unusually long sessions)
- **Power users** (frequent, manual-heavy interaction)



Outlier users have extremely long maximum session durations, reaching several hundred thousand seconds, while the majority of users have much shorter session lengths clustered near the bottom. The outliers reach around 728 k seconds which is 202 h which is more than 8 days of activity! These values are not representative of typical user behaviour and may (will) skew global duration statistics.

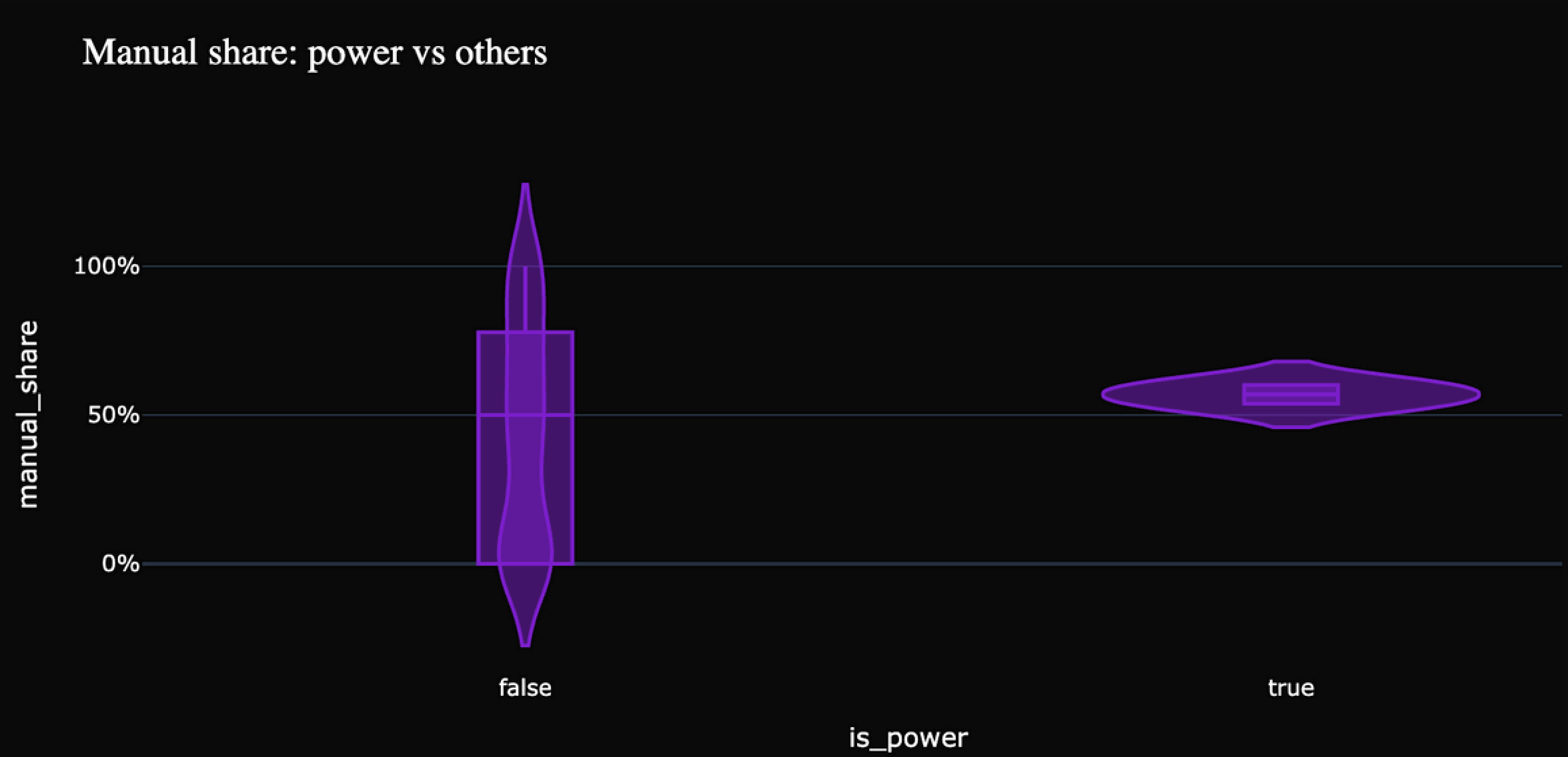
Outliers maintain consistently longer sessions, suggesting either automated workflows or extended development activities. These should be treated separately when analysing engagement, as they differ from the typical session duration distribution.





The distribution for the population is narrow and concentrated at the bottom, while power users cluster at a much higher median, with very few outliers. That highlights the consistent activity of power users as they open and interact with tool windows far more frequently. They likely represent core users who rely heavily on the tool for daily work.

Power users tend to manually trigger and control tool windows, suggesting intentional, hands-on interaction patterns. They were identified by high number of total sessions and manual share larger than 50%. This indicates deeper familiarity with IDE features and customization, likely reflecting experienced developers or testers who prefer direct control rather than automated workflows.



JETBRAINS Internships

Author:
Aleksandra Mulewicz