Federato Report and Visuals

Shreshth Sharma

Krish Khasnobish

Anusheh Atif

Anwesha Bali

2024 - 02 - 24

| Table of Contents | |
|---|---|
| Topic | Pg Number |

Our group took a look and analyzed the Federato Riskop Data provided to us.
We used a number of techniques and data models to help us out. We have provided Appendices throughout this report which report to the Jupyter Notebook, which contains **both the visuals and the code.**

**<u>Initial Set Up:</u>**

Our first challenge was to access the data in a meaningful way. We created a script that cleaned up the data for us, and loaded approximately 250 CSVs with 100 000 to 200 000 rows each into centralized database as DuckDB. The script converts multiple CSV files, originally from JSON format, into a structured DuckDB database for efficient analysis. It reads all CSVs from a folder, cleans column names to ensure SQL compatibility, and creates a unified "events" table in DuckDB. Each CSV is processed, cleaned, and loaded into the database using bulk COPY commands. This process centralizes all user event data, enabling faster queries and supporting advanced analyses like retention tracking, user journey mapping, and causal inference.

This script can be founded in Appendix A in the Jupyter Notebook Attached.

In Appendix B, we can find the database scheme for the “events” table to help us write further queries.

**<u>Exploratory Analysis:</u>**

We then moved on to the Exploratory Analysis portion of our project to further understand the scope of the data we worked with.

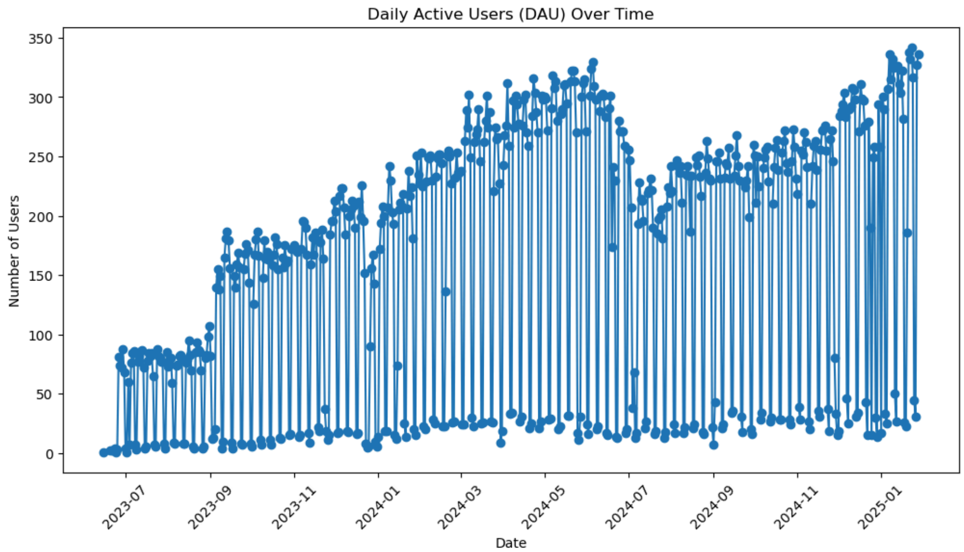
In Appendix C, we use the count function for events, as well distinct user IDs and Session IDs to find out total entries in the database, the amount of unique users, and amount of unique sessions that we have access to. After the data was cleaned, there were **15 409 437 total events, 2 243 unique users, and 607 879 unique sessions.**

In Appendix D, we find the total range of the data we are working with, which is the first and last event recorded. We found the earliest and latest event were on **2023-06-15 21:30:21.880 and 2025-01-28 21:00:00.162** respectively.

In Appendix E, we used aggregate methods to find the top 10 event types to see which events were the most commonly used for the Riskops Platform. The most common event type was account line rendering with 2.7 million out of the nearly 15.5 million being these events.

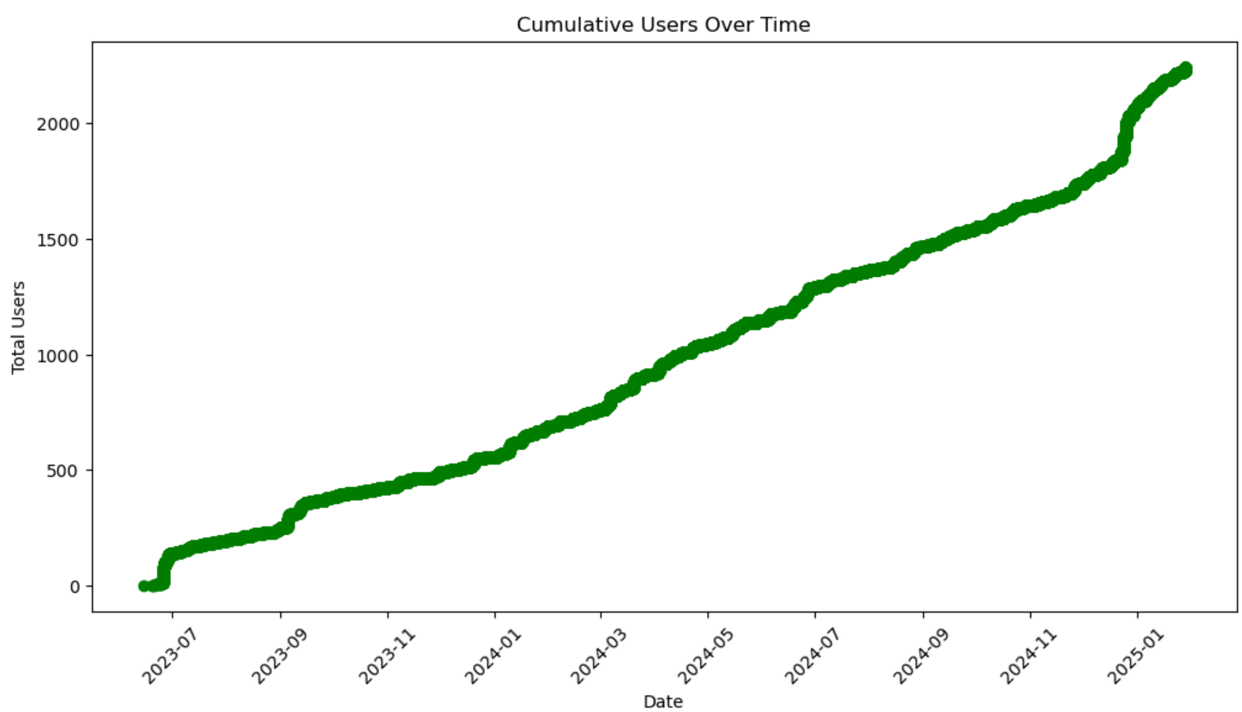
| | event_type | event_count |
|---|---|---|
| 0 | account-lines::widget:render | 2719869 |
| 1 | account-lines::configurable-table:render | 1524746 |
| 2 | account-lines:::view | 1011812 |
| 3 | dashboard:my-book:configurable-table:render | 785355 |
| 4 | dashboard:my-book:widget:render | 750288 |
| 5 | application-window-opened | 733715 |
| 6 | account:::view | 730780 |
| 7 | account-lines::layout:render | 688435 |
| 8 | dashboard:my-book::view | 525136 |
| 9 | action-center:action-details::view | 348010 |

In Appendix F, we found the daily active users. A SQL query was used to extract event dates and count unique user IDs for each day from the "events" table. The data was grouped by date and ordered chronologically. The resulting data was then visualized using a line plot, highlighting fluctuations in daily user activity. This approach helped identify trends, peaks, and drops in engagement, offering valuable insights into user behavior and platform performance.

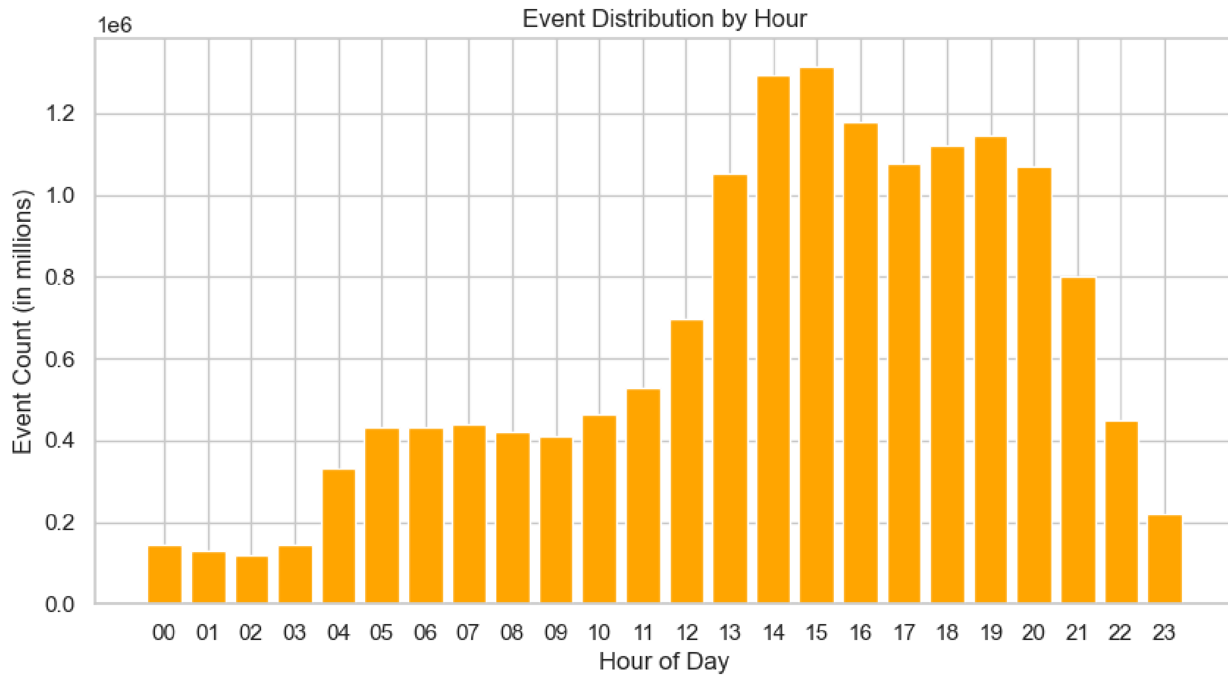



We notice that there were some time periods with steep declines, such as 2024-06, and 2023-11. Federato can and should further investigate these time periods to see the root reason for this.

Also in Appendix F, we found the cumulative user count, which shows the progression of the users' first login within the database.



In Appendix G, We found the event time by hour, by extracting the Hour Value from event time, to explore any patterns in what time of the day, most users are committing actions.




We found that most events are taking place from 1200h to 200h hours, with the peak being 1500h hours.

In Appendix H, we aggregated the sum of the event count, and ordered it by user. This provided insights into the platform's most engaged users, helping to highlight power users and their activity levels.

| | user_id | total_events |
|---|---|---|
| **0** | 0c4896b7-08fa-4cb4-852f-27ad584f0864 | 343369 |
| **1** | 96521245-b6a9-4393-a42e-3e8e470f5e5f | 241697 |
| **2** | 8dbcce7d-02e8-4556-b917-f5d3393bf859 | 215716 |
| **3** | 6066f244-0707-4e85-9fc5-70e6fefdea0b | 200510 |
| **4** | 82b1d870-9646-4b85-898f-3754cc30538e | 192109 |
| **5** | b6071cc4-c091-4de4-82b1-722eff62b91c | 186868 |
| **6** | b2c2fd37-7866-4136-bce0-6f453452836b | 172634 |
| **7** | a883990d-2aa3-4644-9511-5f1c035717f2 | 168171 |
| **8** | 1b992a40-b2c3-4b69-a5bc-05711eeddd80 | 164436 |
| **9** | b575b199-1108-46d1-b452-455d922394dd | 154645 |

| | user_id | session_id | event_count |
|---|---|---|---|
| 0 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1726222859279 | 14854 |
| 1 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1726122006359 | 10637 |
| 2 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1725907971092 | 9202 |
| 3 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1726475713708 | 6378 |
| 4 | bb8fa1af-a540-4d0f-a633-569c61f01281 | 1722862609519 | 5903 |
| 5 | 57ac3fc1-363b-4f0d-8802-e9cc93b47610 | 1721574311964 | 5803 |
| 6 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1725889254062 | 5680 |
| 7 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1725964537186 | 5671 |
| 8 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1732607054975 | 3793 |
| 9 | 3473675c-4986-4549-b988-ccb4ad18f605 | 1726158010936 | 3676 |

In Appendix I, we did the a similar methodology from Appendix H, to find top user sessions by event, to find information, and further explore which sessions were the most active and the reasons causing it.

**Interesting Takeaway**: 8 of the top 10 most active sessions were all by the same user. This raises the questions of bots or automation on the platform?

In Appendix J , the analysis calculated the 28-day user retention rate by identifying users who returned to the platform within 28 days after their initial session. The process began by determining each user's first session start time. It then tracked subsequent sessions for the same users, selecting those that occurred within the 28-day window following their first session. The final calculation compared the number of users who returned within this period to the total number of users, providing a percentage-based retention rate. This metric offered valuable insights into user engagement and long-term platform stickiness.
We found that **1434/2242 or ~64%** of the users returned for another session within 28 days of their first session.

We also found that since the events were largely unstandardized, it was difficult to create big picture overviews from them.

Therefore, in Appendix K, we refined our event types into event categories, based on a semantic model, to see a broade view on what users were using this platform for. We found that Render, and View events accounted for vast majority of events.

| Enhanced Category | Event Count |
|---|---|
| Render | 6,609,405 |
| View | 5,206,193 |
| Navigation | 1,317,858 |
| Click | 1,177,594 |
| Submission | 327,001 |
| Account Action | 117,366 |
| Edit/Update | 87,485 |
| Filter/Config | 55,835 |
| Other | 39,997 |
| Document | 20,888 |
| Submission Workflow | 263 |

## Events before Ending Session:

The main purpose of this analysis is to maximize the amount of time and engagement on the platform.
We need to ensure that we have a deeper analysis on what event is causing clients to end the session.

In Appendix L, we found that the top 10 events that were happening immediately before the session_end event were the following. his query analyzed the user events leading directly to a session end, excluding instances where 'session_end' or 'session_start' were the preceding events themselves. It used a window function (LEAD) to identify the immediate next event within each session, enabling the identification of actions users most commonly take right before ending a session. By filtering out redundant transitions and focusing on meaningful preceding events, the query provided insights into potential drop-off points in user sessions. The top 20 events preceding session ends were extracted, helping to pinpoint behaviors that may signal disengagement, thereby guiding optimization efforts to reduce session terminations.

| | preceding_event | count_before_session_end |
|---|---|---|
| 0 | account-lines::configurable-table:render | 29907 |
| 1 | account-lines::widget:render | 23016 |
| 2 | dashboard:my-book:configurable-table:render | 17804 |
| 3 | action-center:::close-click | 8767 |
| 4 | dashboard:my-book::view | 5383 |
| 5 | :all-accounts::view | 5240 |
| 6 | action-center:action-details::view | 5105 |
| 7 | account-lines:::change-rating-click | 5079 |
| 8 | account:::view | 3874 |
| 9 | ::configurable-table:render | 3358 |
| 10 | account-property-rating:perils:configurable-ta... | 2663 |
| 11 | submissions:all-policy::view | 2304 |
| 12 | :all-accounts:configurable-table:render | 1903 |
| 13 | account-property-rating:perils::view | 1882 |
| 14 | application-window-opened | 1798 |
| 15 | action-center:action-details:response-form:sub... | 1754 |
| 16 | :all-accounts:widget:render | 1563 |
| 17 | account-lines:::view | 1502 |
| 18 | submissions:all-exposures::view | 1400 |
| 19 | action-center:::view | 1277 |

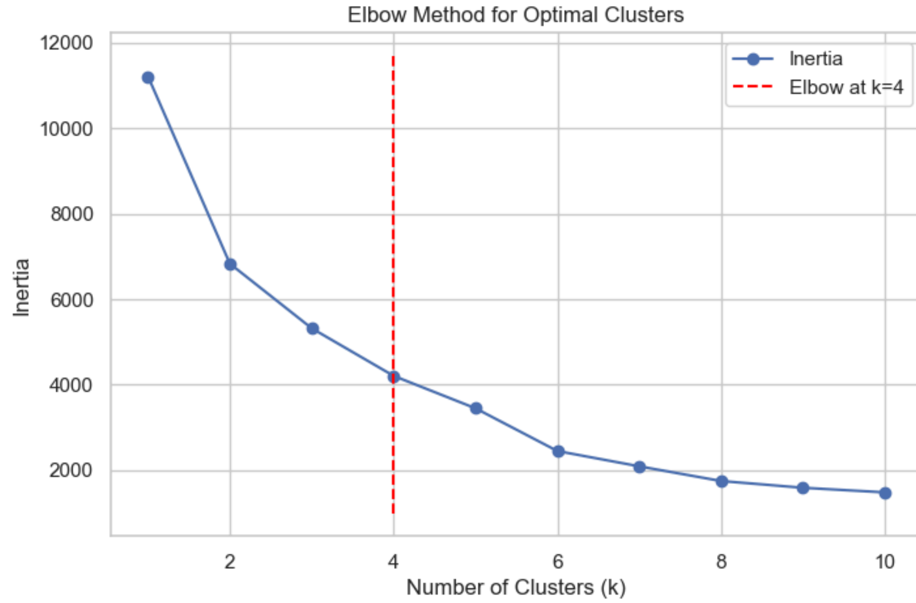
In our Insights Section, we give insights and recommendations on how we can improve this.

**<u>Segmentation Analysis, K - Means Clustering:</u>**

In Appendix M, we designed a query to extract comprehensive user engagement metrics from the DuckDB database, focusing on session-level and user-level behaviors. By utilizing a Common Table Expression (CTE) named session_durations, it first calculates the duration of each session by determining the time difference between the earliest and latest events within the same session. This avoids nested aggregates and simplifies the computation. This is a crucial step and function for K - Means Clustering we do to perform cluster analysis.

In Appendix N, We found out how many distinct groups we have, which will tell us the optimal number of k, for our k - means clustering that we use. K-Means Clustering was applied to segment users based on engagement metrics like total events, sessions, average session duration, session ends, and unique events. The data was standardized using StandardScaler to ensure balanced clustering. The Elbow Method, combined with KneeLocator, identified the optimal number of clusters by analyzing where adding more clusters no longer significantly reduced inertia. An Elbow Curve visualized this optimal k value. This clustering helps profile user behaviors, enabling targeted strategies to boost engagement and retention. We found the optimal number of clusters to be 4.

```
Optimal number of clusters (k): 4
```

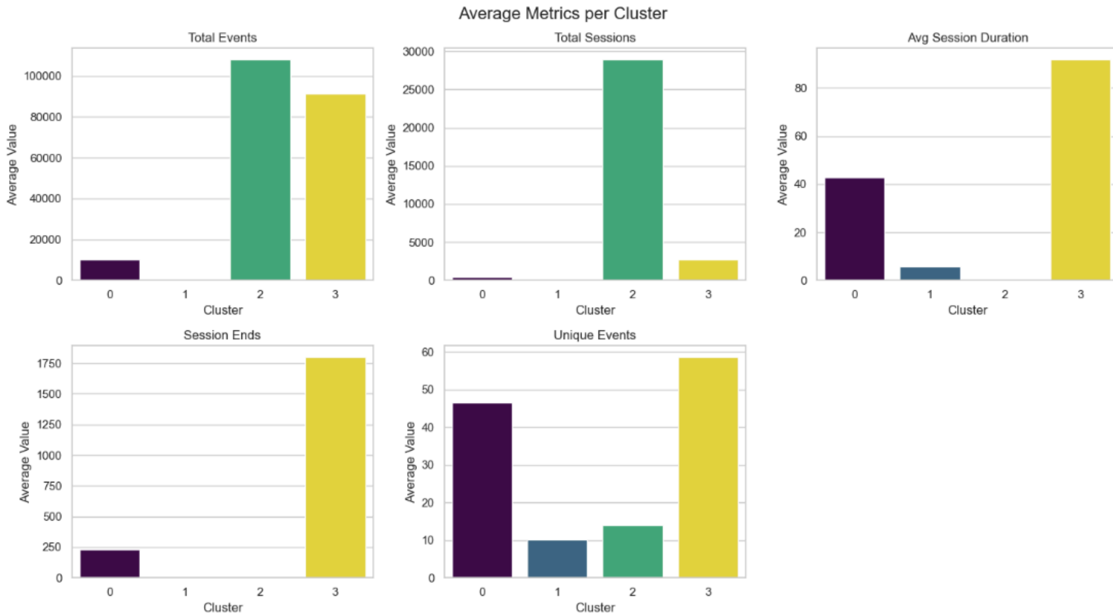



In Appendix O, we applied the clusterization logic based on our metrics to our users.
This analysis profiled user clusters by calculating summary statistics, including mean, median, minimum, and maximum values for engagement metrics such as total events, sessions, average session duration, session ends, and unique events.

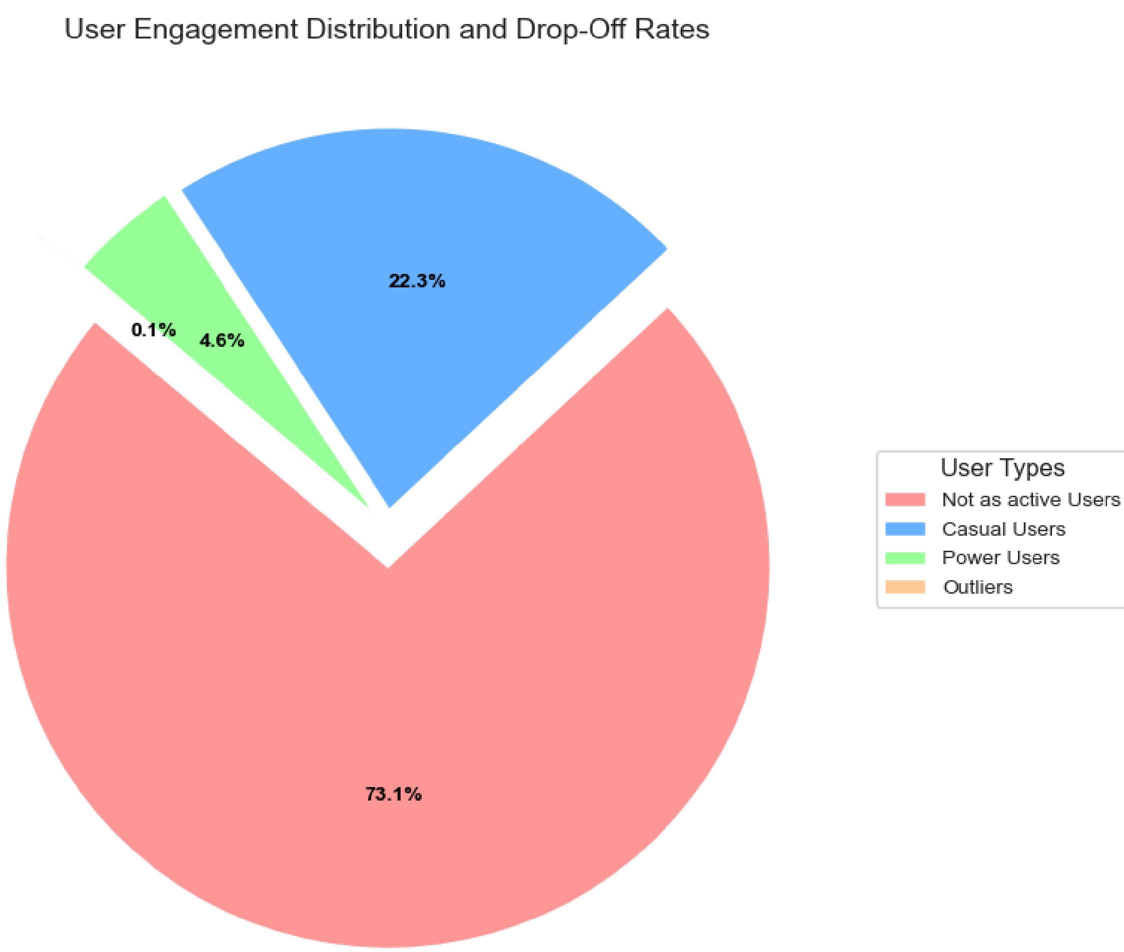
In Appendix P, Users were grouped based on their cluster assignments, and the distribution of users across clusters was visualized. Bar plots highlighted the average behavior within each cluster, while a count plot showcased the number of users in each group. These visualizations provided insights into user behavior patterns, helping identify high-engagement segments and potential areas for optimization.

In Our key takeaways for clustarization were that nearly 73% of users fall into the cluster where the average event per session, and total events is incredibly low, and average time is also low. Meaning many users are just passively viewing and not using the platform to it's full extent. This was Cluster 1. Cluster 0 was of casual users, who also had low averages, but significantly higher than Cluster 1. Cluster 3 were our power users, users who leveraged many events, and also had a high number of time spent, and average events per session. Cluster 2 was two users, likely bots, as they had practically 0 time spent, but the highest total events, and events per session.

```
🔹 Cluster Profile Summary:
   cluster  total_events_mean  total_events_median  total_events_min  \
0        0       10107.310621               5146.0                12   
1        1         171.295482                 31.0                 1   
2        2      108134.000000             108134.0             51832   
3        3       91451.330097              82356.0             37774   

   total_events_max  total_sessions_mean  total_sessions_median  \
0             59508           469.929860                  296.0   
1              9595            13.727717                    3.0   
2            164436         28931.000000                28931.0   
3            343369          2671.417476                 2462.0   

   total_sessions_min  total_sessions_max  avg_session_duration_mean  ...  \
0                   1                3174                  42.762630  ...   
1                   1                 573                   5.813301  ...   
2               21336               36526                   0.055502  ...   
3                 301               18870                  91.850622  ...   

   avg_session_duration_max  session_ends_mean  session_ends_median  \
0                265.242150         232.807615                 53.0   
1                 79.821712           5.791209                  0.0   
2                  0.084105           4.500000                  4.5   
3                343.165110        1800.582524               1630.0   

   session_ends_min  session_ends_max  unique_events_mean  \
0               0.0            1569.0           46.565130   
1               0.0             445.0           10.179487   
2               0.0               9.0           14.000000   
3             299.0           15773.0           58.669903   

   unique_events_median  unique_events_min  unique_events_max  User Count  
0                  42.0                  7                142         499  
1                   9.0                  1                 36        1638  
2                  14.0                  5                 23           2  
3                  58.0                 14                146         103  
```



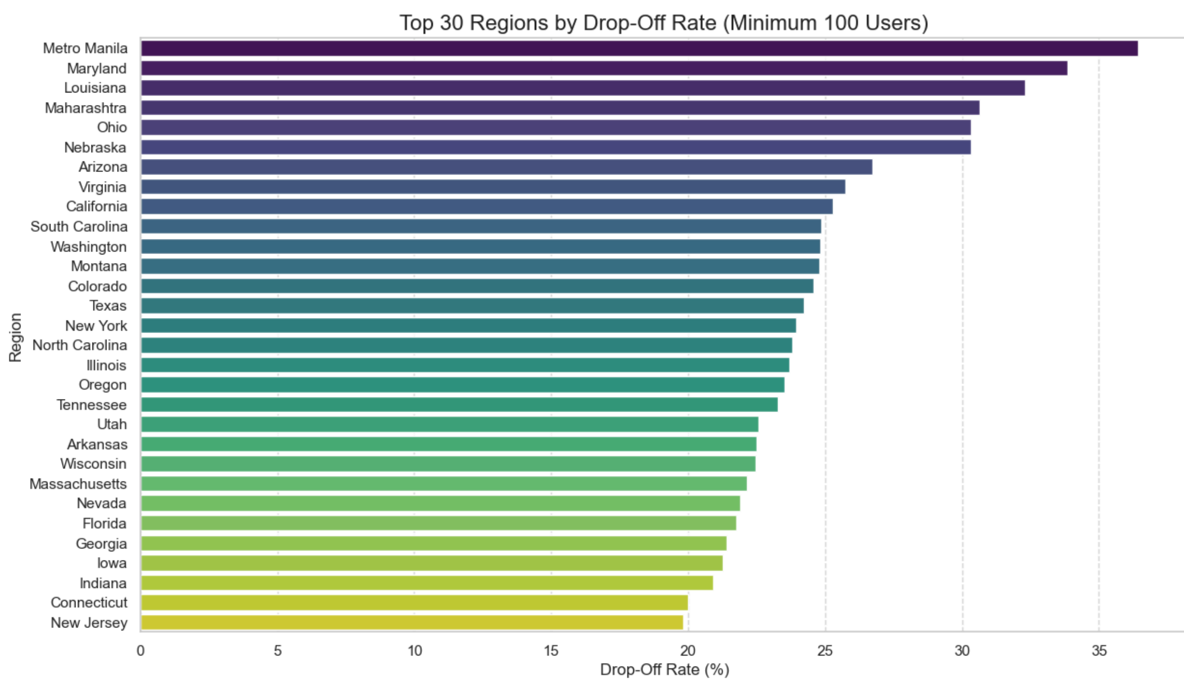
Average Metrics per Cluster
Total Events
Total Sessions
Avg Session Duration
Session Ends
Unique Events
Average Value
Cluster

User Engagement Distribution and Drop-Off Rates
22.3%
0.1%
4.6%
73.1%
User Types
Not as active Users
Casual Users
Power Users
Outliers

## **Regional DropOffs:**

In Appendix Q, the analysis identified regions with the highest user drop-off rates while ensuring that only regions with a significant user base were considered for meaningful insights. Using DuckDB, the query tracked user sessions and pinpointed events where users dropped off (defined as the last recorded event in a session) across various regions and device types. To refine the analysis, only regions with at least 100 users were included, filtering out low-traffic areas that could skew results. The top 30 regions with the highest average drop-off rates were then visualized using a bar plot, highlighting potential areas for improvement in user engagement. This approach allows stakeholders to focus optimization efforts on regions with substantial user activity and high disengagement, ensuring targeted strategies for reducing drop-offs and enhancing user retention.

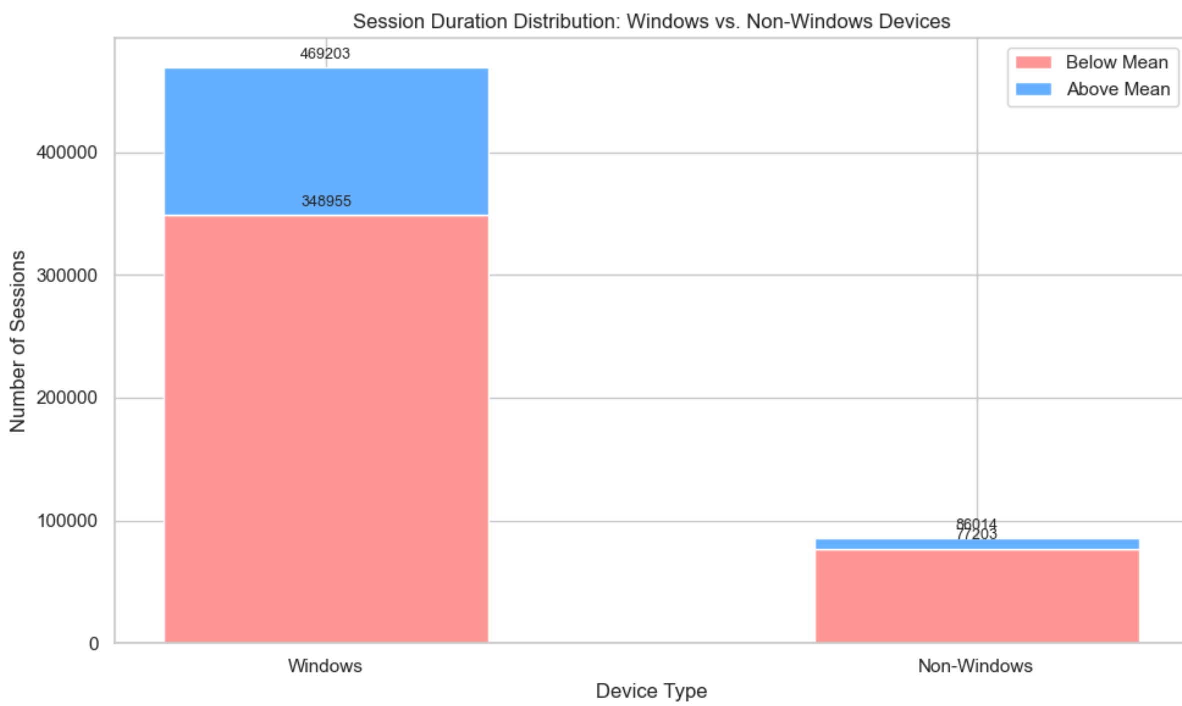



## **Windows vs Non-Windows:**

In Appendix R, we compared windows vs non-windows devices. This analysis examined user session durations to explore potential differences in engagement between Windows and non-Windows devices. The first step calculated the average session duration by identifying the start and end times for each session and computing the mean session length across all users. Next, sessions were classified as either above or below this average duration. The relative risk of longer engagement for Windows users compared to non-Windows users was then calculated, providing insight into how device type influences user behavior. A stacked bar chart was created to visualize the distribution of sessions across the two device categories, highlighting differences in user engagement patterns. This approach helped determine whether specific devices are associated with longer sessions, informing optimization strategies for user retention.

Surprisingly, the relative risk of window users being above the mean was near 2.5 meaning, they were 2.5x more likely to actively use the platform. Recommendations and Insights will be provided in the Insights Sections.

```
Mean Session Time: 13.91 minutes

Relative Risk Analysis: Windows vs. Non-Windows Devices
   windows_above  windows_below  non_windows_above  non_windows_below  \
0       120248.0       348955.0             8811.0            77203.0

   relative_risk
0       2.501849
```




## Sampling :

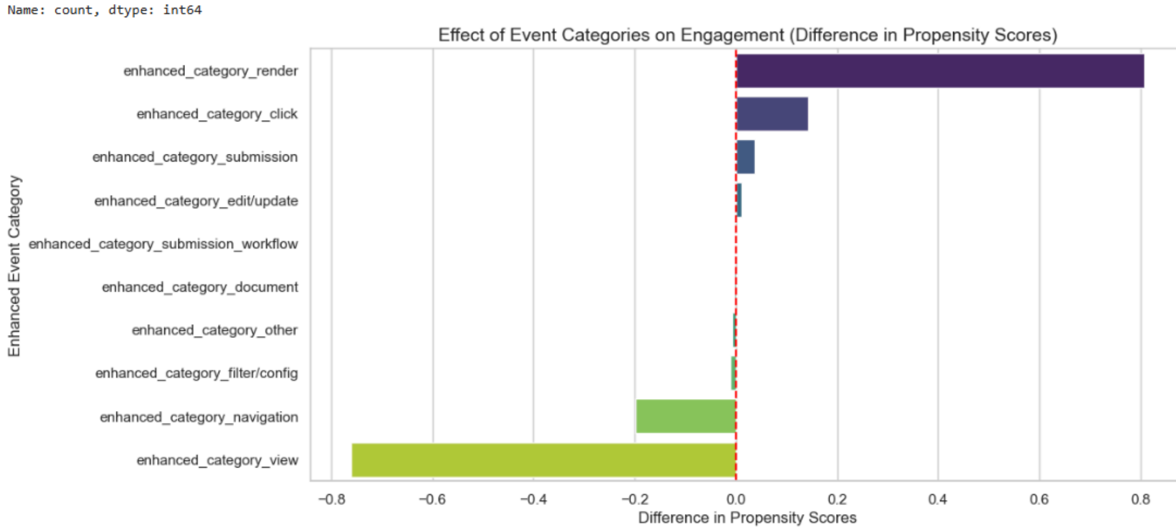
The following calculations and insights we found were quite heavy, to limit this, in Appendix S:

We set up a dataframe for further use which sampled a random 100 000 users.

**<u>Casual Interference / Propensity Score :</u>**

In Appendix T, we employed Propensity Score Matching (PSM) to assess the causal impact and interference of different event categories on user engagement. The process began by converting event timestamps into datetime format to accurately calculate session durations, which were then used to classify sessions as "engaged" or "not engaged" based on whether they exceeded the mean session duration. Events were categorized into broader types, such as navigation, views, or submissions, to simplify analysis. Logistic regression was applied to estimate propensity scores, representing the likelihood of user engagement based on event types. Nearest Neighbor matching was then used to balance engaged and non-engaged samples, ensuring a fair comparison. Finally, the effect of each event category on engagement was visualized through a bar plot, highlighting which actions most significantly influenced user retention and session length. This approach offered insights into user behavior and informed strategies to enhance platform engagement.

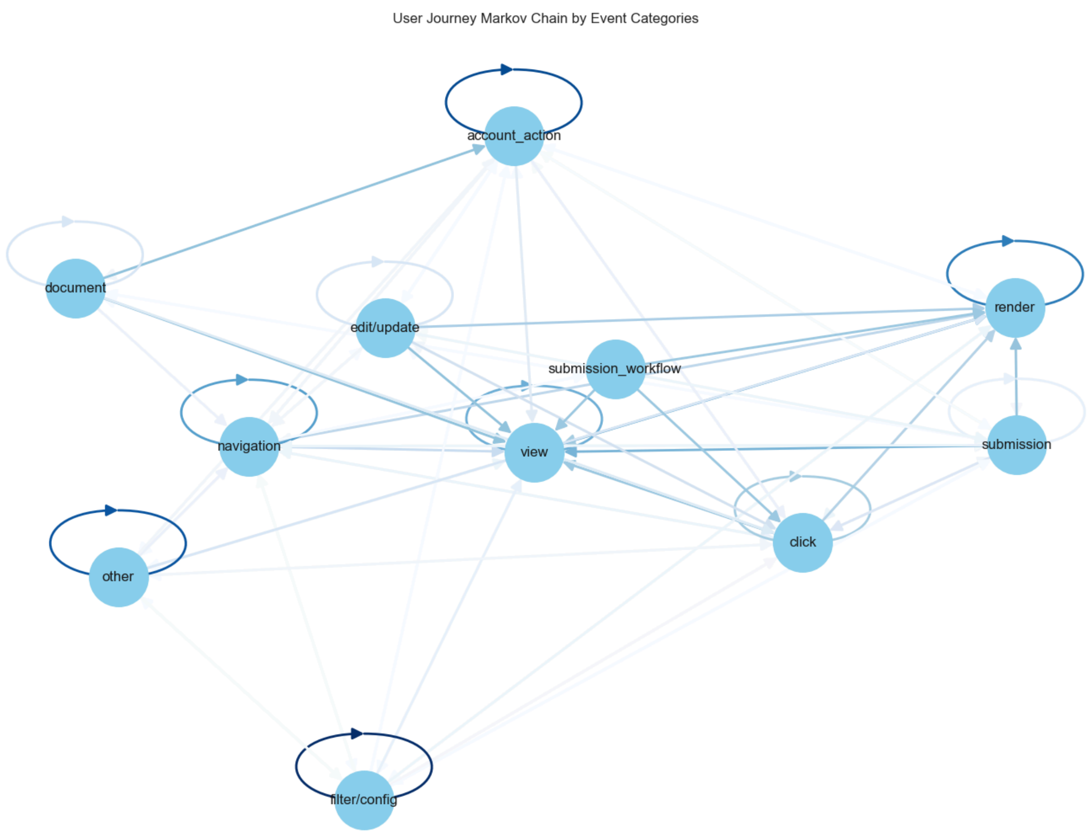
```
Mean Session Duration: 12.45 minutes
engaged
0    70077
1    30052
Name: count, dtype: int64
```




Render, Click, and Submission category events  had positive propensity, where as, the view, and navigation had negative values. This suggests looking into the problems regarding the negative propensity categories, and addressing these issues.

**User Journey and Markov Chains:**

In Appendix U, the code builds and visualizes a Markov Chain to analyze user journey patterns based on event categories, ultimately helping to predict the most likely next action a user might take. The process begins with enhanced event categorization, where raw event types are grouped into broader, more intuitive categories such as view, click, navigation, and submission. This simplifies the analysis and provides clearer insights into user behaviors. The events are then chronologically ordered for each user and session, allowing for the construction of accurate event sequences that reflect real user flows. Using these sequences, a Markov Transition Matrix is created, capturing the probability of transitioning from one event category to another. This matrix forms the foundation for mapping user journeys and identifying common navigation patterns across the platform. The resulting transition probabilities are visualized using a Directed Graph, where nodes represent event categories and edges reflect the likelihood of moving between them. The thickness of each edge corresponds to the strength of the transition, making it easy to identify dominant user flows and potential bottlenecks. Beyond visualization, the code also predicts the next best action based on historical behavior. By identifying the most probable next step for a user currently in a specific event category, the system can offer dynamic suggestions to guide users toward deeper engagement. This predictive approach not only highlights areas where users tend to drop off but also empowers the platform to implement targeted interventions, such as personalized recommendations or optimized UI paths. Ultimately, this analysis helps refine the user experience, improve retention rates, and drive more meaningful user interactions.



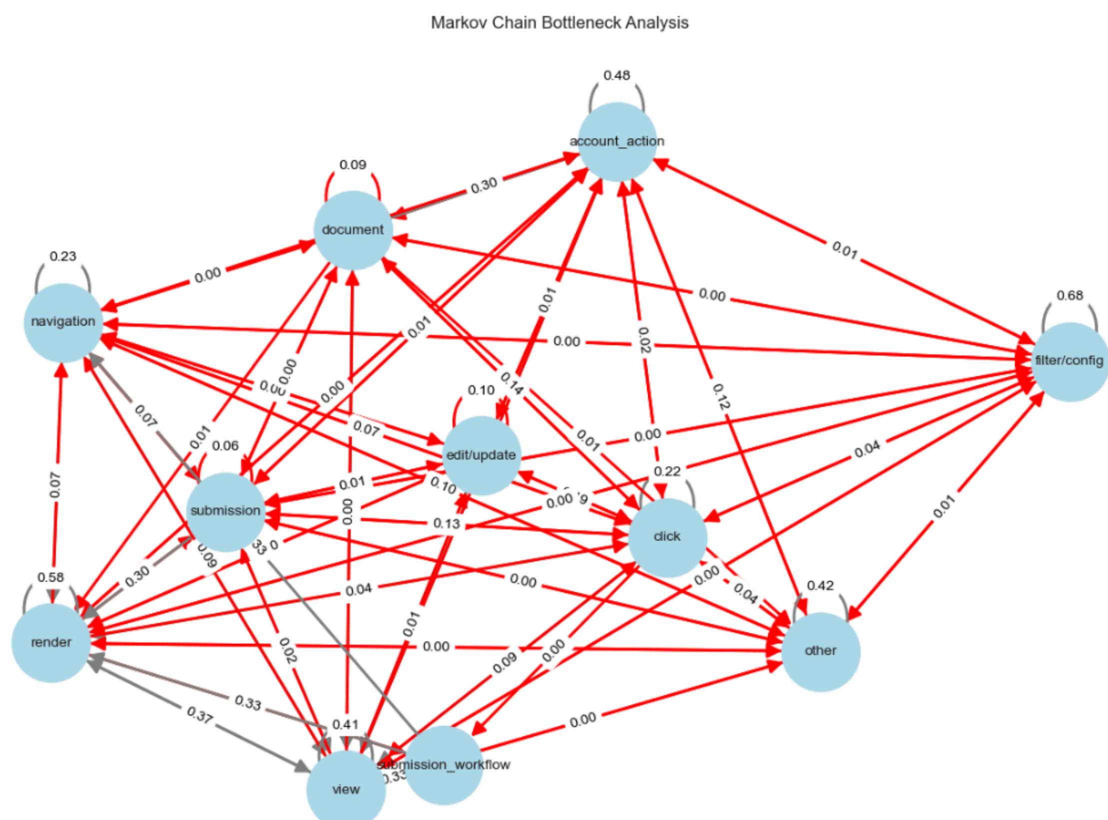

Furthermore, we can also generate the top 10 most common user paths to better make sense of what users are experiencing.

Top 10 User Journeys (Most Frequent Paths):

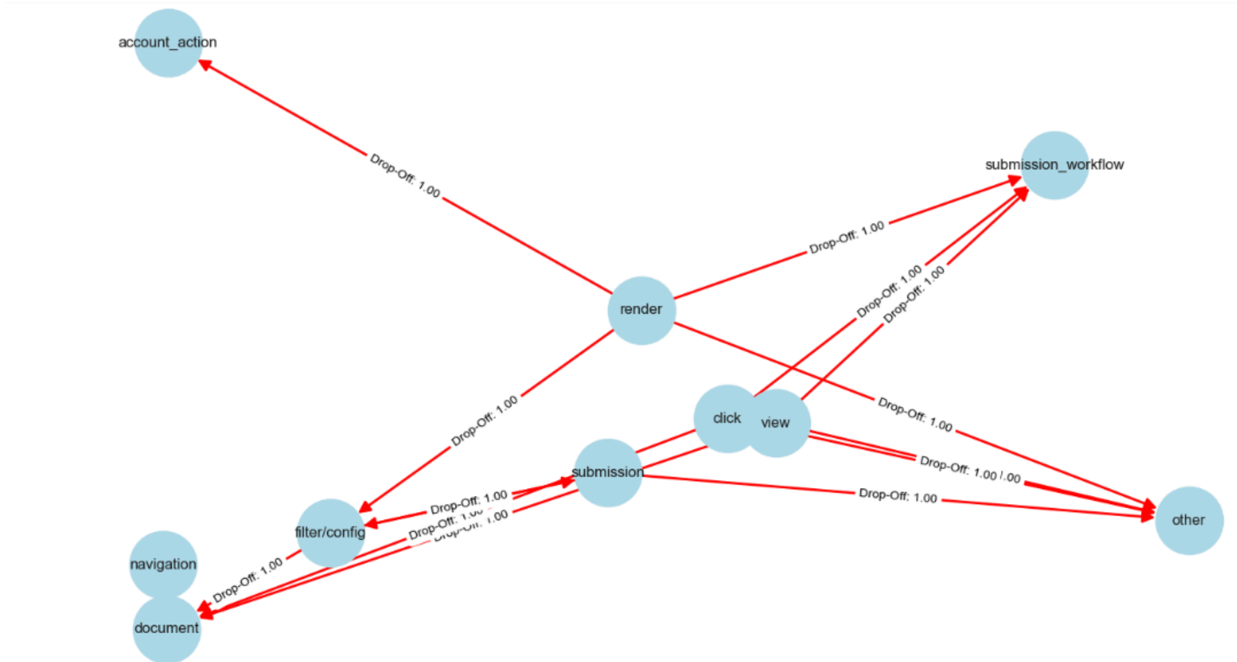
1. render — Count: 19157
2. view — Count: 17804
3. navigation — Count: 6793
4. click — Count: 4354
5. render -> render — Count: 2987
6. view -> render — Count: 1571
7. render -> view — Count: 1499
8. view -> view — Count: 1459
9. submission — Count: 1025
10. render -> render -> render — Count: 592

**<u>Bottlenecks in User Journeys:</u>**

In Appendix W,  We construct a Markov Chain to analyze user behavior and identify potential bottlenecks in the user journey. It begins by applying an enhanced event categorization to simplify raw events into broader categories like navigation, view, click, and submission. The user events are sorted chronologically, and transitions between event categories are mapped to build the Markov Chain, where each edge represents a user action and its associated probability. The bottleneck identification process highlights transitions with low probabilities (below a defined threshold, e.g., 20%), indicating points where users often drop off or disengage. These weak links in the user flow can signal usability issues or areas lacking clear calls to action. The visualization leverages a directed graph where nodes represent event categories, and edges illustrate the user flow between them. Bottleneck transitions are colored red, helping stakeholders quickly spot areas that may require optimization. This analysis is valuable for improving user retention, guiding UI enhancements, and streamlining the user journey.




Since this is very overwhelming to see and undertand, we find the top 15 more painful (bottleneck heavy) transitions.

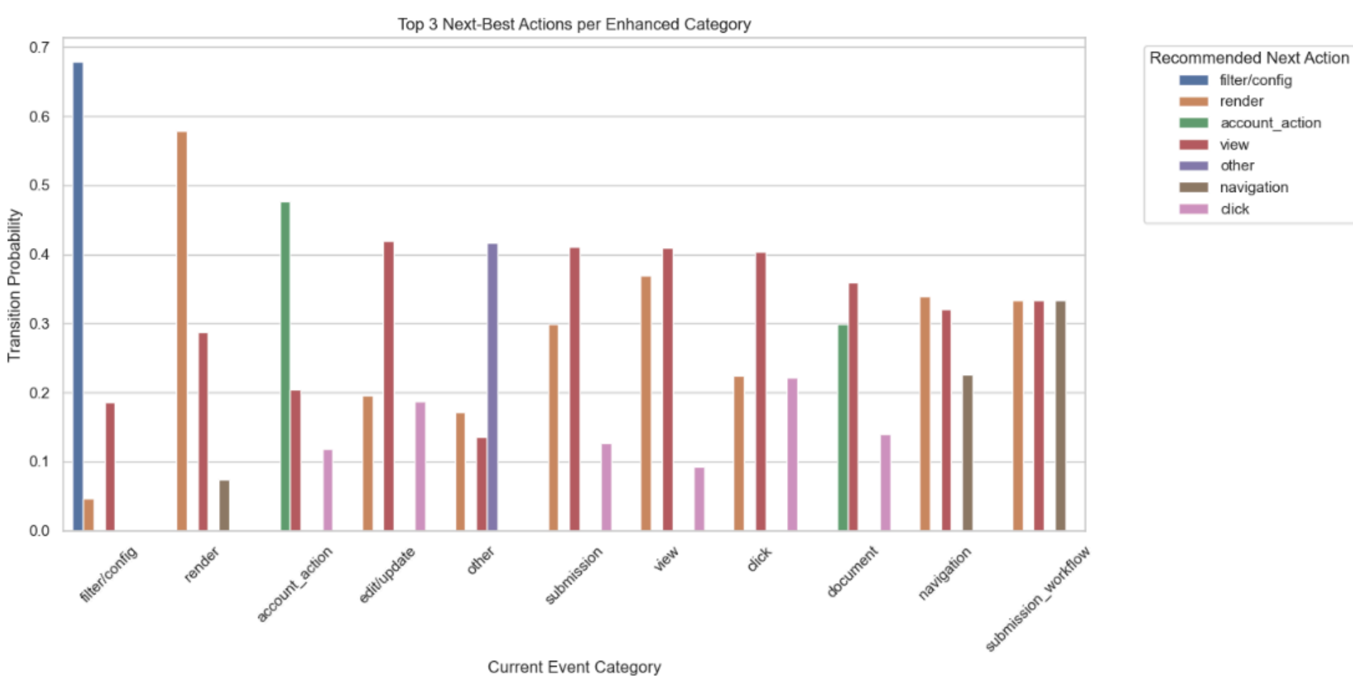



These can be read in the following table:

| enhanced_category | next_category | probability | drop_off_rate |
|---|---|---|---|
| render | submission_workflow | 0.000023 | 0.999977 |
| view | submission_workflow | 0.000028 | 0.999972 |
| click | submission_workflow | 0.000121 | 0.999879 |
| render | account_action | 0.000180 | 0.999820 |
| submission | other | 0.000482 | 0.999518 |
| render | filter/config | 0.000518 | 0.999482 |
| navigation | document | 0.000840 | 0.999160 |
| view | other | 0.000853 | 0.999147 |
| submission | filter/config | 0.000965 | 0.999035 |
| render | other | 0.001396 | 0.998604 |
| filter/config | submission | 0.001582 | 0.998418 |
| filter/config | document | 0.001582 | 0.998418 |
| view | document | 0.001962 | 0.998038 |
| click | other | 0.002053 | 0.997947 |
| click | document | 0.002173 | 0.997827 |

**Next Step Optimizations:**

In Appendix X, we conducted a top 3 next-best action analysis using a Markov chain transition matrix to identify the most probable user actions following each event category. It processed the Markov chain data (markov_df) to determine the top three next actions for each enhanced event category, ranking transitions based on their probabilities. This approach provided a clear understanding of user flows and highlighted the most common paths users took through the platform. By visualizing these transitions in a bar plot, the code made it easier to interpret user behavior patterns, where each bar represented the likelihood of moving from one event category to another. This analysis was particularly useful for optimizing user journeys, as it highlighted the most effective paths that drove engagement and helped in designing smoother workflows. It also supported personalized recommendations by suggesting the next logical actions based on historical data, enhancing user experience. Additionally, by identifying event categories with low transition probabilities, the analysis flagged potential bottlenecks that required attention. Overall, this method offered valuable insights for improving user retention, guiding users through desired actions, and enhancing overall platform engagement.



**Insights and Recommendations :**

Based our previous Analysis's we can come up with come insights and actionable recommendations:

1. Retention Rate: The average initial user retention rate is roughly 63% as discussed earlier. Set up a notification system, to alert user of benefits of the platform when they first start, as this will increase the likelihood of them staying longer, if they are retained.
2. Preceding event before session drop off:

   The drop-off analysis revealed critical user behaviors that often precede session terminations, offering valuable insights for platform optimization. High drop-off counts were observed after data-heavy views like account-lines::configurable-table:render and dashboard:my-book:configurable-table:render, suggesting that users may be passively browsing without taking further action. To reduce disengagement, implementing call-to-actions or next-step suggestions post-render could encourage deeper interaction.

   In cases where users actively close components, such as the action-center:::close-click, it indicates task completion or a lack of further guidance. Introducing prompts like "Before You Go" messages or highlighting related tasks could help retain users.

   Similarly, views like account:::view and submissions:all-policy::view often lead to exits, possibly signaling task completion or a lack of actionable follow-ups. In these instances, offering related accounts, policies, or comparison tools could extend user engagement.

   For instances where users interact directly, such as account-lines:::change-rating-click or action-center:action-details:response-form:submit-click, users likely completed specific tasks before exiting. Offering follow-up suggestions or similar tasks immediately after these actions can maintain momentum and encourage continued platform use.

   Finally, drop-offs following application-window openings and form completions suggest moments of confusion or natural end-points in the user journey. Implementing guided walkthroughs for new users and post-action recommendations for task-oriented users can help improve overall retention.

By identifying these patterns and integrating targeted optimizations, the platform can better guide users through their workflows, reduce drop-offs, and boost overall engagement.

Further Insights in Appendix L.

3. The clustering analysis revealed four distinct user groups with varying engagement patterns. **Cluster 1 — Inactive/Low Engagement Users** (1638 users, the largest cluster) exhibited the lowest total events and sessions, with an average session duration of about 6 minutes and minimal interaction with platform features. This group, while the largest, represents disengaged users who log in but don't interact deeply. To improve engagement, the platform could introduce onboarding improvements like feature walkthroughs or in-app tooltips, launch re-engagement campaigns using targeted emails with tips and success stories, and implement gamification strategies to encourage activity. **Cluster 0 :Casual Users** (499 users) displayed moderate activity with an average session duration of around 42 minutes and a fair number of unique events but relatively low total sessions. These users engage occasionally but aren't deeply invested. Strategies to enhance engagement include promoting underused features, using personalized in-app notifications to suggest next steps, and collecting feedback through surveys. **Cluster 3 : Power Users** (103 users) showed the highest levels of engagement with a large number of total events and sessions, an impressive average session duration of around 92 minutes, and a high diversity of events (~58 unique events). These highly engaged users likely see strong value in the platform. Retaining them could involve inviting them to beta-test new features, offering advanced analytics or premium services, and maintaining their loyalty through exclusive perks. **Cluster 2 : Outliers** (2 users) had extremely high event counts but very few users, suggesting the presence of system users, bots, or data anomalies. These users should be verified for legitimacy and excluded from performance analyses if found to be outliers. Overall platform trends indicate a high inactivity rate, with about 75% of users falling into low-engagement categories, but the strong potential of power users highlights the platform's value when fully utilized. User drop-off remains a concern, with many users failing to progress from casual to power users. To address

these gaps, key recommendations include enhancing onboarding to guide users from inactivity to higher engagement, nurturing power users as brand advocates through exclusive features, and increasing platform stickiness for casual users with tools like notifications and goal tracking.

Further Details in Appendix P

4. Windows vs Non-Windows:

   The analysis reveals that Windows users are approximately 2.5 times more likely to have above-average session durations compared to non-Windows users. This suggests that the Federato RiskOps platform may be more optimized or intuitive for Windows environments, leading to higher engagement. In contrast, non-Windows users, potentially including Mac, Linux, and mobile users, might face usability challenges or performance limitations that contribute to shorter sessions. This disparity highlights an opportunity to improve the user experience on non-Windows devices by addressing compatibility issues, optimizing UI/UX elements, and ensuring consistent feature accessibility across platforms. By focusing on these enhancements, Federato could boost overall user engagement, reduce drop-off rates, and expand its reach among a more diverse user base.

5. Propensity Score/Casual Interference:

   The propensity score analysis reveals that render events (enhanced_category_render) are the strongest positive drivers of user engagement, indicating that users interacting with rendered content are more likely to stay engaged. Optimizing the rendering experience or increasing its frequency could further boost engagement. Click and submission events also show moderate positive effects, suggesting that interactivity promotes retention, while passive viewing (enhanced_category_view) leads to disengagement, highlighting the need for calls-to-action or personalized prompts. Navigation and document events have mixed effects, with navigation slightly negative and document interactions neutral, indicating opportunities to improve guidance and engagement. Secondary actions like filters and workflows show minimal impact, suggesting that resources should focus on high-impact areas. Overall, enhancing rendered content, promoting interactivity, and mitigating passive behaviors are key strategies for improving user engagement.

6. High Drop Off Areas:

   The high drop-off analysis revealed several critical bottlenecks in user navigation that hinder engagement and conversion. A significant issue was observed in the **Submission Workflow**, where transitions from render, view, and click to the submission_workflow had drop-off rates exceeding 99.97%. This indicates that users either struggle to reach the submission process or find it too complex or inaccessible from current entry points. Similarly, the **Account Engagement** flow, particularly the render → account_action transition, showed a 99.98% drop-off rate, suggesting users rarely proceed to manage account settings after viewing a page. Another major concern was the **low conversion from viewing to document access**, with view → document and navigation → document transitions showing drop-off rates of 99.80% and 99.91%, respectively. This implies that users either cannot find relevant documents or lack the motivation to explore further resources.

Additionally, **filtering tools were found to be underused**, as paths leading to filter/config from render and submission had drop-off rates around 99.95%. This suggests that users find these features unintuitive or not valuable. **Click paths also proved unproductive**, with click → other and click → document showing over 99.7% drop-off rates, indicating that clicks often led to disengaging or irrelevant destinations. Lastly, **unclear navigation flows** caused users to disengage, as evidenced by the high drop-off rates from view and render to other events, suggesting users might be encountering dead ends or irrelevant content.

To address these issues, several recommendations emerged. **Optimizing submission workflow access** by improving visibility and adding clear calls to action could reduce user friction. **Simplifying account actions** and providing contextual prompts may encourage users to explore account-related settings. **Enhancing document discoverability** with in-page links, pop-ups, or tooltips explaining document relevance could boost deeper engagement. Similarly, **revamping filtering tools** with intuitive designs and onboarding guides can make them more appealing. **Reevaluating click-based navigation** to remove dead ends and guide users to high-value content is essential, while **exit-intent strategies**, such as pop-ups suggesting next steps or related features, can help retain users before they leave. Finally, **conducting usability testing** in high drop-off areas would provide deeper insights into user pain points, enabling iterative improvements to the overall user journey.

**Why this works for Federato:**

The proposed framework effectively guides users toward actions that maximize engagement by leveraging advanced techniques such as Markov Chains, propensity score modeling, and causal inference to map user journeys and predict optimal next steps. By identifying high drop-off points and bottleneck transitions, the framework pinpoints critical moments where real-time recommendations could prevent disengagement, such as optimizing submission workflows and simplifying complex user paths. Its design accommodates varying user behaviors, enabling tailored recommendations for different user types and insurance companies using Federato, enhancing its scalability and applicability.

From a business impact and feasibility perspective, the framework integrates predictive modeling, graph analysis, and causal inference to deliver data-driven insights that are directly implementable within Federato's dynamic platform. The ability to analyze user sessions in real-time and apply adaptive recommendations ensures the solution can actively improve user retention and engagement. The use of platform-agnostic models also supports integration across multiple user environments, including Windows and non-Windows devices, broadening its feasibility.

The solution demonstrates innovation and creativity by focusing on strategies that increase daily user engagement over a 28-day period. By targeting key engagement drivers—such as optimizing high-impact actions and reducing friction in user flows—the framework offers actionable recommendations that not only enhance user experience but also improve platform utility for underwriters. This approach helps underwriters stay engaged longer, guiding them toward high-value features more effectively, which ultimately aligns with Federato's business goals of maximizing platform usage and customer retention.

AI Disclaimer: LLMs were used to comment the code and for some qualitative descriptions, but no code was written with LLMs.