

I have sessionized clickstream data for user interaction on website and mobile app with target variable satisfaction (satisfied , dissatisfied). There is a page sequence list with all page names in the order in each session with information like search_start, chat_start ect. When I did a summery statistics , it turned out that people who are starting a chat are more dissatisfied than satisfied. Also sequences with more error pages are more dissatisfied.Using a sequence analysis, how can I identify some customer pain points and possible reasons for dissatisfaction. Do your research and show results based on use cases and best practices

Here's a structured approach—grounded in real-world use cases and best practices—for mining your sessionized clickstream sequences to pinpoint customer pain points and root causes of dissatisfaction:

1. Prepare and Segment Your Sequences

1. Clean & normalize events

- Group synonymous page labels (e.g. “chat_start” vs. “conversation_init”).
- Remove bot/trial sessions.

2. Label by outcome

- Split into **Dissatisfied** vs. **Satisfied** cohorts.
 - You've already noted that **chat_start** and **error**-heavy sequences correlate with dissatisfaction—now drill deeper.
-

2. Mine Differential Sequential Patterns

Use **frequent sequence mining** algorithms (e.g. PrefixSpan or SPADE) separately on each cohort, then look for patterns that are **significantly more frequent** in the Dissatisfied group.

Example: you might discover that the subsequence

nginx

CopyEdit

search_start → **search_results** → **error_page** → **chat_start**

- occurs in 15% of Dissatisfied sessions but only 2% of Satisfied ones.
- **Tools & libraries:**
 - Python's `mlxtend.preprocessing.TransactionEncoder` + `PrefixSpan` implementations.
 - R's TraMineR package for sequence mining. dl.acm.org

By contrasting frequent subsequences, you surface **pain-point journeys** rather than isolated events.

3. Build & Compare Transition Models

1. **First-order Markov chains**
 - Estimate transition matrices for each cohort (i.e. probability of going from page A to B).
 2. **Compare matrices** to highlight **high-risk transitions** in Dissatisfied users (e.g. A→error_page jumps from 5% to 20%).
 3. **Visualize** with **Sankey diagrams** or **state-transition graphs** to make differences tangible for stakeholders. helio.app
-

4. Cluster Similar Session Paths

1. **Represent each session** as a feature vector (e.g., counts of n-grams, dwell times).
2. **Cluster** (e.g., k-means on sequence embeddings from Word2Vec/UX2Vec) to group common journey archetypes within the Dissatisfied cohort.
3. **Profile clusters** by their dominant patterns:
 - Cluster A: “search → filter → no results → chat”

- Cluster B: “browse → product_view → add_to_cart → payment_error”

This surfacing of archetypes helps tailor targeted fixes for each pain-point segment.

quantzig.com

5. Drill into Timing & Context

- **Dwell-time analysis:** identify pages where Dissatisfied users spend *too little* (give up) or *too long* (struggle).
 - **Contextual metadata:** device type, time of day, acquisition channel—overlay these to see if, for example, mobile users hit the error page more often.
-

6. Validate & Quantify Impact

1. **A/B test** redesigns or UI fixes for the top pain points (e.g. improved search suggestions, error-handling flows).
 2. **Measure lift** in satisfaction metrics post-remediation (e.g. a 10% drop in error-page encounters yields a 5-point NPS gain). netguru.com
-

7. Iterate with Continuous Monitoring

- Set up **automated dashboards** tracking the incidence of your identified “pain sequences.”
 - Trigger alerts when their frequency spikes—so you can catch regressions early.
-

Real-World Example: E-Commerce CTR Boost

A sportswear retailer used this very approach—mining differential patterns between high- and low-performing sessions, visualizing transition graphs, then clustering user journeys. They uncovered a “filter → no results → exit” pain loop, optimized their filtering UI, and saw a **120% uplift in click-through rates** within weeks. quantzig.com

By combining **differential sequence mining**, **Markov transition comparisons**, and **session clustering**, you’ll not only pinpoint *where* users stumble but also *why*—equipping you to prioritize fixes that move the needle on customer satisfaction.



Sources