Patient Treatment in Emergency Departments

Iliano Fasolino

iliano.fasolino@studenti.unimi.it

49594A

Business Information Systems

Prof. Paolo Ceravolo

a.a. 2025 - 2026

Contents

1 Description of the Case Study (Code HERE)

In this report, we analyze a large volume of processes related to the treatment of patients in an Emergency Department (ED).

This analysis and study is based on four key points:

1. Filtering and cleaning of the event log by building a preprocessing pipeline
2. Performance analysis
3. Process discovery and Conformance Checking
4. Improvement suggestions

To analyze and visualize the dataset, the main libraries used are Pandas, NumPy, Matplotlib, Seaborn, and **PM4Py**. Additionally, I used **Disco** to obtain slightly different animations and visualizations to study the data and better understand certain situations (in fact, some analyses will be shown using both PM4Py and Disco).

2 Filtering and Cleaning

The main point I focused on initially was understanding the structure of the event log before applying any transformation, immediately noticing the presence of many missing values (approximately 70-90%), but justified as (also specified in the assignment) attributes specific to certain activities.

As the first of the three main doubts I noticed and needed to understand how to address, there was the presence of many simultaneous events with the same timestamp (for each *stay_id*). Rather than thinking they were incorrect logging errors, I considered them as a sort of "implicit feature" of the process, namely that two or more medications were simply administered, which can occur normally in many cases/diseases/situations of this kind.

2.1 Simultaneous Events with Same Timestamp

Given that in process mining the main objective is to study primarily a **macro** flow (Enter → Triage → Treatment → Discharge, therefore from A to Z without leaving behind analysis between intermediate processes), and that I perceived the details on individual medications more as "event-level attributes," i.e., useful for specific analyses but not useful for the main goal, namely process discovery.

Therefore, the choice that seemed more appropriate, in order to maintain a certain focus on the **end-to-end** flow of the process, was to **aggregate** these simultaneous events, making the process less granular and lighter.

2.2 Missing Values

Having established that rather than data quality errors, missing values represent attributes that apply only to certain activities (therefore a normal pattern), I analyzed that:

- *Acuity*, *Gender*, *Race* → recorded only during "Triage in the ED"

- *Vital signs* → recorded only during "Vital sign check"
- *Drug* info → recorded only during "Medicine reconciliation/dispensations"

Therefore, the first step I took was to distinguish between case-level attributes (constant for each patient) and event-level attributes (specific to each activity), and subsequently for case-level attributes I applied forward/backward fill[1] within each case; while for event-level attributes simply leaving "NaN".

Main Columns (Essential for process mining)

- `stay_id` → Case ID
- `time` → Timestamp
- `activity` → Activity name

Case-level Columns (Useful for analysis)

- `gender`, `race`, `arrival_transport` → Patient characteristics
- `acuity` → **Very important** (urgency level 1-5)
- `chiefcomplaint` → Reason for access
- `disposition` → Outcome (HOME, ADMITTED, etc.)

2.3 Filter Noise Events

Typically in process mining, noise includes:

- **Incomplete cases**: i.e., partial data or logging errors that can distort timestamp analysis at the end-to-end level, for example cases without "Enter the ED" as an event and/or cases without "Discharge from the ED" as the last event.
- **Cases with too few events**: in fact, a complete path must require at least Enter → Triage → Treatment → Discharge, therefore, assuming that cases with **fewer than 4 events** are very likely errors, fewer than 4 events is the threshold I set.
- **Temporal outliers**: to have a well-filtered event log and carefully analyzing the context we are studying, **less than 30 minutes** for a complete process is truly unrealistic in an ED, while **more than 48 hours** I see as exceptional cases (in medical jargon called prolonged boarding) that do not represent a standard flow (nor even a worst case of a standard flow).

With the aggregation of simultaneous events, we reduced the event log by **33%** (from 25,115 to 16,826 events) while preserving the process structure, while with forward/backward fill we reduced missing values from 89% to 0%, completely eliminating them; finally with noise filtering we removed 66 cases (3.6%) classified as:

- 0 cases with incomplete start/end (so we can say that all cases were properly logged)

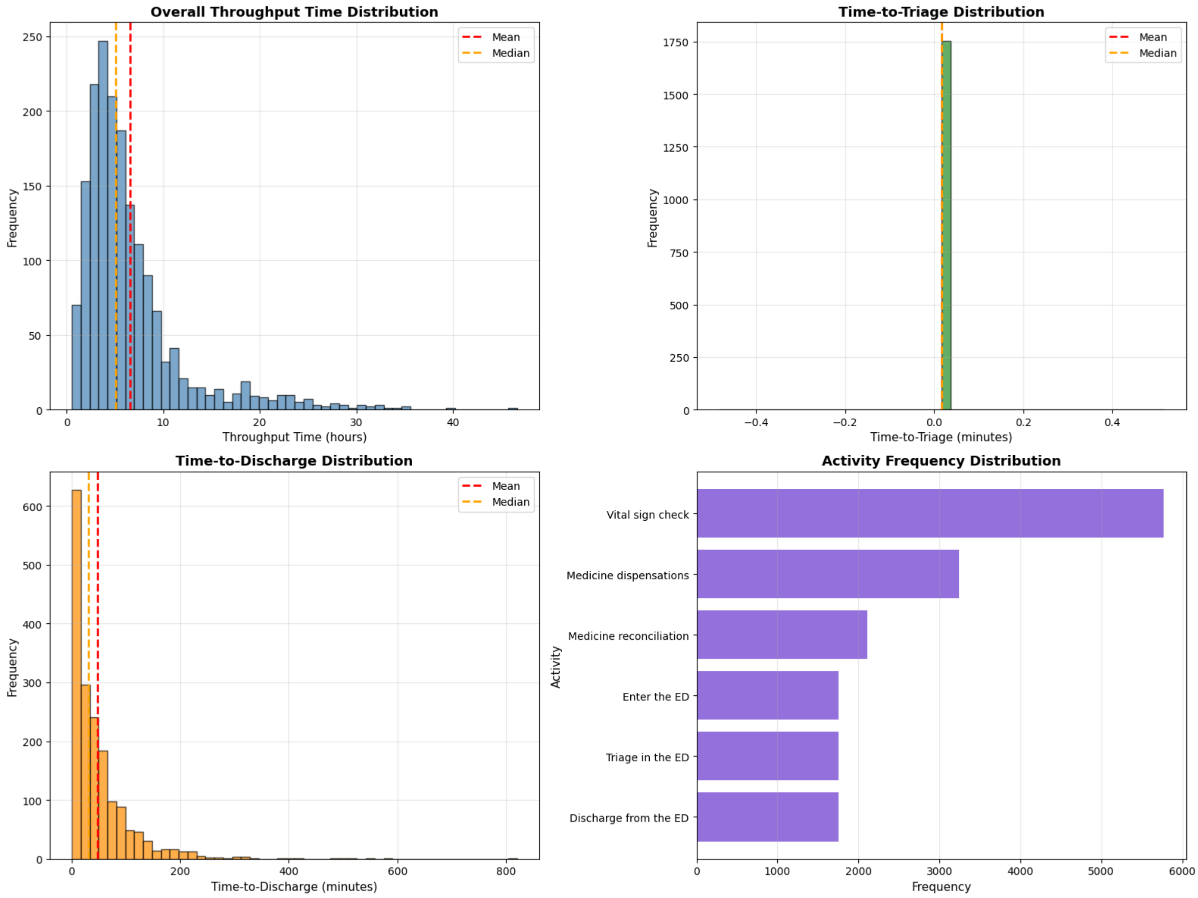
[1]Forward Fill: Propagates the last known value forward until the next non-missing value. Backward Fill: Propagates the next known value backward until the previous missing value.

- 49 cases with < 4 events (likely logging errors)
- 17 cases with extreme durations (< 30 mins or > 48 hours)

These thresholds ensure focus on the ED paths while removing statistical outliers that would **skew** performance analysis.

3 Performance Analysis

The main objective of this phase was to identify various structural inefficiencies in the process of our case study through the analysis of temporal performance and flows.



Temporal analyses were performed by calculating and implementing the following metrics.

3.1 Throughput Time (Total Duration)

We notice an asymmetric distribution and high variability; in fact, 10% of patients take less than 2.2 hours and another 10% take 12.4 hours (prolonged boarding as mentioned before, or truly complex cases).

3.2 Time-to-Triage (Enter → Triage)

```
TIME-TO-TRIAGE (Enter → Triage)

count    1.754000e+03
mean     1.666667e-02
std      3.470436e-18
min      1.666667e-02
25%      1.666667e-02
50%      1.666667e-02
75%      1.666667e-02
max      1.666667e-02
Name: time_to_triage_minutes, dtype: float64

Mean:    0.02 minutes
Median:  0.02 minutes
Std Dev: 0.00 minutes

Percentiles:
  10%: 0.02 minutes
  25%: 0.02 minutes
  50%: 0.02 minutes
  75%: 0.02 minutes
  90%: 0.02 minutes
  95%: 0.02 minutes

Cases with time-to-triage > 60 minutes: 0 (0.0%)
```

Here we have a critical finding (a "false positive"). The Time-to-Triage metric showed no variability (all cases = 1 second), indicating this is a **logging artifact** rather than true process performance. This suggests that "Enter the ED" and "Triage in the ED" are logged as combined event in the system (I assume automatically). For this reason, time-to-triage **is not a reliable KPI for this dataset** and will **not be used** for the next steps such as bottleneck identification.

3.3 Time-to-Discharge (Last Activity → Discharge)

```
TIME-TO-DISCHARGE (last activity → Discharge)

count    1754.000000
mean       48.254571
std        61.767018
min         0.766667
25%         9.000000
50%        31.000000
75%        61.320833
max       820.700000
Name: time_to_discharge_minutes, dtype: float64

Last activity before Discharge:
  - Vital sign check: 1387 (79.1%)
  - Medicine dispensations: 290 (16.5%)
  - Medicine reconciliation: 57 (3.2%)
  - Triage in the ED: 20 (1.1%)
```

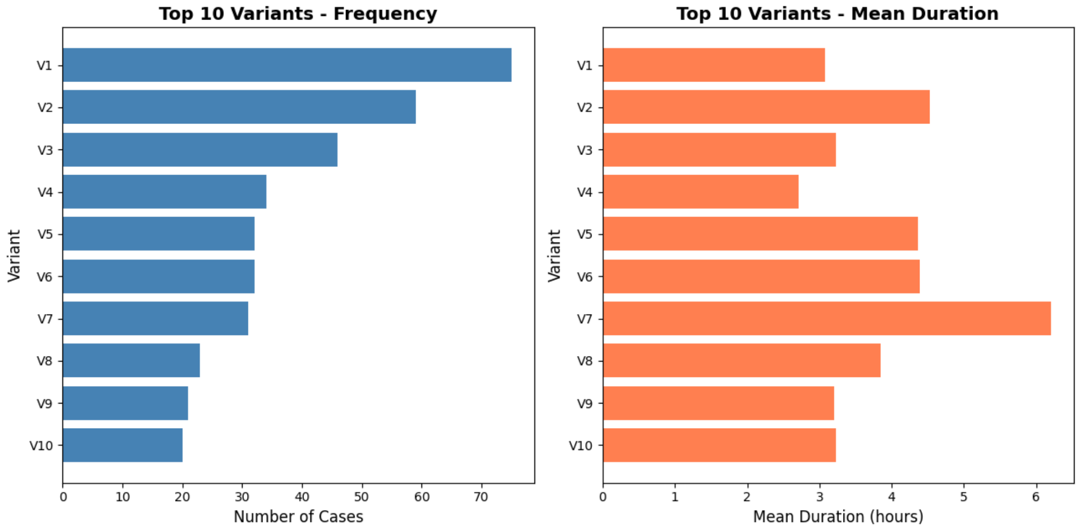
Based primarily on the mean, even though we notice truly very high variability (some cases even reach 13.7 hours), we see that on average 48 minutes of administrative waiting are employed after the patient is "clinically ready". The term administrative is not casual, as from these data one can interpret a **clear administrative/organizational bottleneck**, probably given by the time for paperwork for discharge, prescriptions to prepare, instructions/authorizations to give to the patient and/or waiting for transportation in case the disposition occurs via ambulance.

```python
#TIME-TO-DISCHARGE (from the last activity → Discharge)

#For each case, I have found the timestamp of the last activity BEFORE Discharge and the Discharge timestamp

discharge_analysis = []

for stay_id in df['stay_id'].unique():
    case_df = df[df['stay_id'] == stay_id].sort_values('time')

    # Find Discharge event
    discharge_row = case_df[case_df['activity'] == 'Discharge from the ED']
    if len(discharge_row) == 0:
        continue

    discharge_time = discharge_row.iloc[0]['time']

    # Find last activity before Discharge
    before_discharge = case_df[case_df['time'] < discharge_time]
    if len(before_discharge) == 0:
        continue

    last_activity_time = before_discharge.iloc[-1]['time']
    last_activity_name = before_discharge.iloc[-1]['activity']

    time_to_discharge_minutes = (discharge_time - last_activity_time).total_seconds() / 60

    discharge_analysis.append({
        'stay_id': stay_id,
        'last_activity': last_activity_name,
        'last_activity_time': last_activity_time,
        'discharge_time': discharge_time,
        'time_to_discharge_minutes': time_to_discharge_minutes
    })

time_to_discharge = pd.DataFrame(discharge_analysis)
```

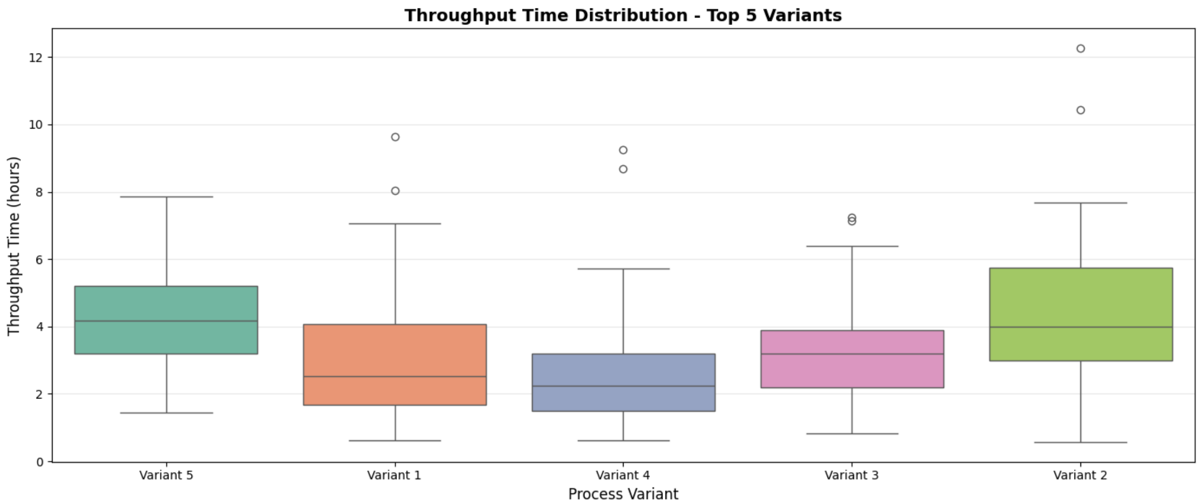
3.4 Process Variants (Frequency + Duration)

```
PROCESS VARIANTS

Total unique variants: 884
Total cases: 1754

Top 10 most frequent variants:

1. [75 cases, 4.3%]
   Enter the ED → Triage in the ED → Vital sign check → Discharge from the ED

2. [59 cases, 3.4%]
   Enter the ED → Triage in the ED → Vital sign check → Vital sign check → Discharge from the ED

3. [46 cases, 2.6%]
   Enter the ED → Triage in the ED → Medicine dispensations → Vital sign check → Discharge from the ED

4. [34 cases, 1.9%]
   Enter the ED → Triage in the ED → Medicine reconciliation → Vital sign check → Discharge from the ED

5. [32 cases, 1.8%]
   Enter the ED → Triage in the ED → Vital sign check → Medicine dispensations → Vital sign check → Discharge from the ED

6. [32 cases, 1.8%]
   Enter the ED → Triage in the ED → Vital sign check → Medicine reconciliation → Vital sign check → Discharge from the ED
...
   Enter the ED → Vital sign check → Triage in the ED → Discharge from the ED

Variants covering 80% of cases: 519
Variants covering 90% of cases: 686
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

In this phase I tried to identify which activity sequences are most common and how long they last, also trying to understand if the **Pareto Principle** could be applied in this process mining of this event log. In fact, in a well-structured process, few variants (10-20%) should cover 80% of cases, and if this were not the case, meaning that too many variants are needed to cover this 80%, then we can say that the process is fragmented/**chaotic**.



The process shows extreme fragmentation, with 884 unique variants and the most common variant covering only 4.3% of cases (hundreds of variants are needed to cover 80%, approximately 230 variants). This **heavily violates** the Pareto principle. This fragmentation indicates lack of process standardization and suggests significant opportunity for improvement through care pathway definition.



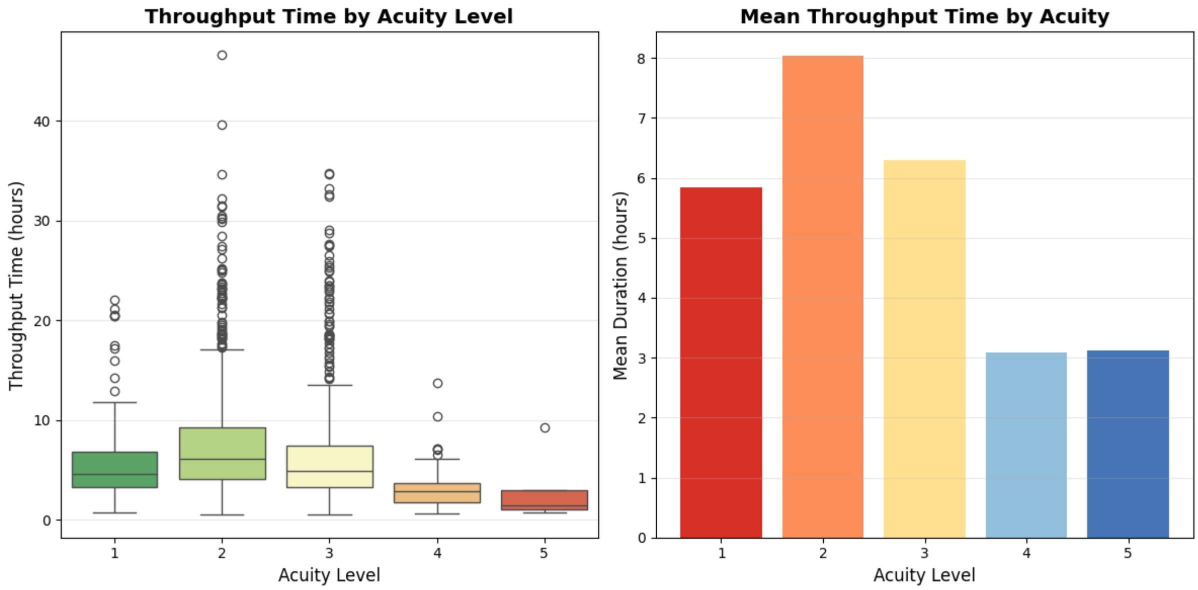
3.5 Segmentation by Acuity Level

PERFORMANCE BY ACUITY LEVEL

Note: Acuity 1-3 = High urgency, 4-5 = Low urgency

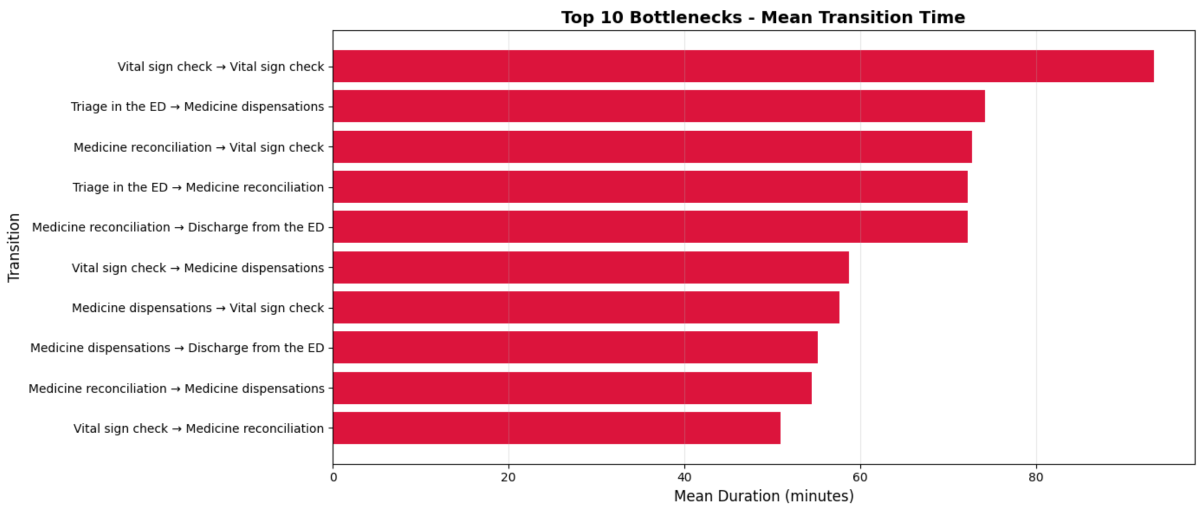
acuity	n_cases	mean_duration	median_duration	std_duration	min_duration	max_duration
1	126	5.831272	4.575000	4.319652	0.707500	22.066667
2	590	8.033486	6.100000	6.190183	0.566667	46.633333
3	851	6.295248	4.933333	5.070371	0.550000	34.766667
4	135	3.084691	2.900000	1.851045	0.616667	13.733333
5	5	3.113333	1.466667	3.579793	0.733333	9.316667

A **counter-intuitive** finding emerged: Acuity 2 (Urgent) patients have longer mean throughput time (8.03h) than Acuity 1 (Critical) patients (5.83h). This sort of **paradox** suggests an organizational bottleneck where urgent-but-not-critical patients fall into a prioritization gap. Potential causes include diagnostic complexity, boarding while awaiting admission, or resource allocation favoring extreme acuity levels. This represents a **significant structural inefficiency** requiring process redesign.



3.6 Bottleneck Identification

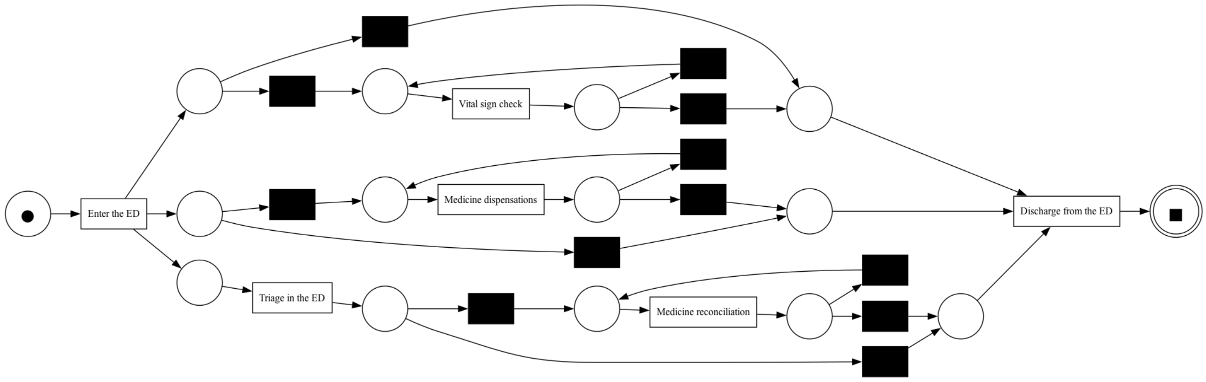
The transitions between activities that cause the greatest delays are Vital sign check → Vital sign check (mean 93 minutes, with 2159 occurrences). This is the **primary bottleneck** and represents patients in observation/monitoring status, waiting between repeated assessments. The long waiting time might suggest either insufficient nursing capacity for more frequent checks (understaffing) or, more likely, patients boarding in the ED while awaiting admission, specialist consultation, or simply test results.



4 Process Discovery and Conformance Checking

To discover the process model from the data, I used **Inductive Miner**, a model very robust to noise that is defined as "best practice" for fragmented datasets like this one.

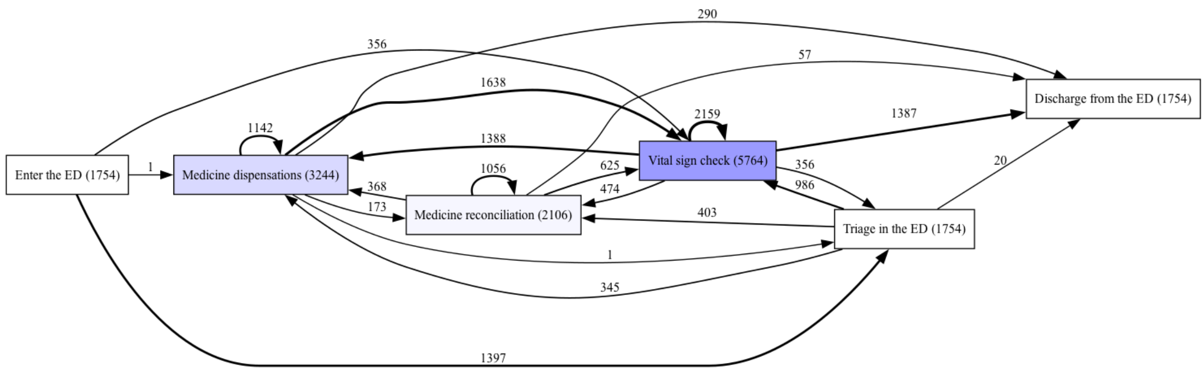
The algorithm was applied to the preprocessed event log (1754 cases, 16376 events) and converted from Process Tree to Petri Net representation for analysis.



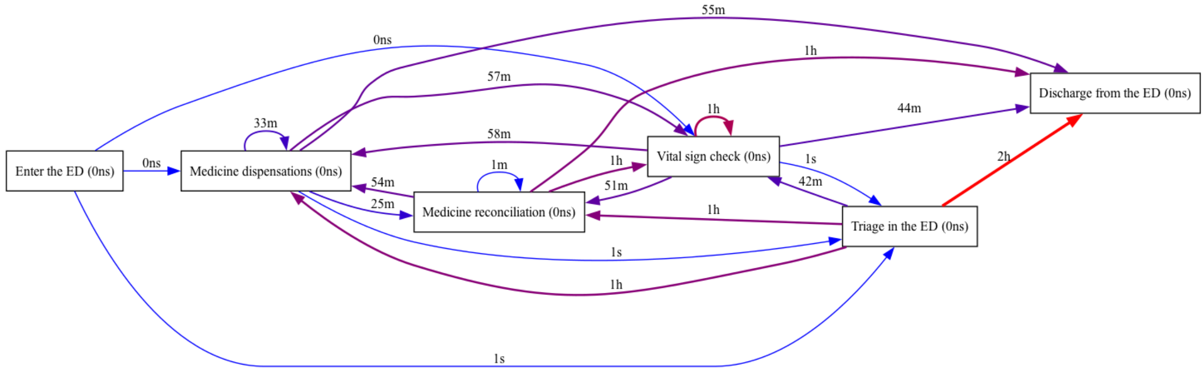
Petri Net Legend:

- Circles () = Places (states/conditions)
- Rectangles () = Transitions (activities)
- Arrows show the flow between activities
- Initial marking () shows where cases start
- Final marking shows where cases end

The discovered Petri Net (15 places, 18 transitions) reveals a process structure with three main pathways after triage: medicine dispensations, vital sign monitoring (with self-loop), and medicine reconciliation. These pathways can occur in different orders or be skipped entirely, explaining the high variant count. The model's moderate complexity (simplicity = 0.70) balances comprehensibility with the need to represent 884 unique variants.



Frequency DFG generated (shows how often activities follow each other)



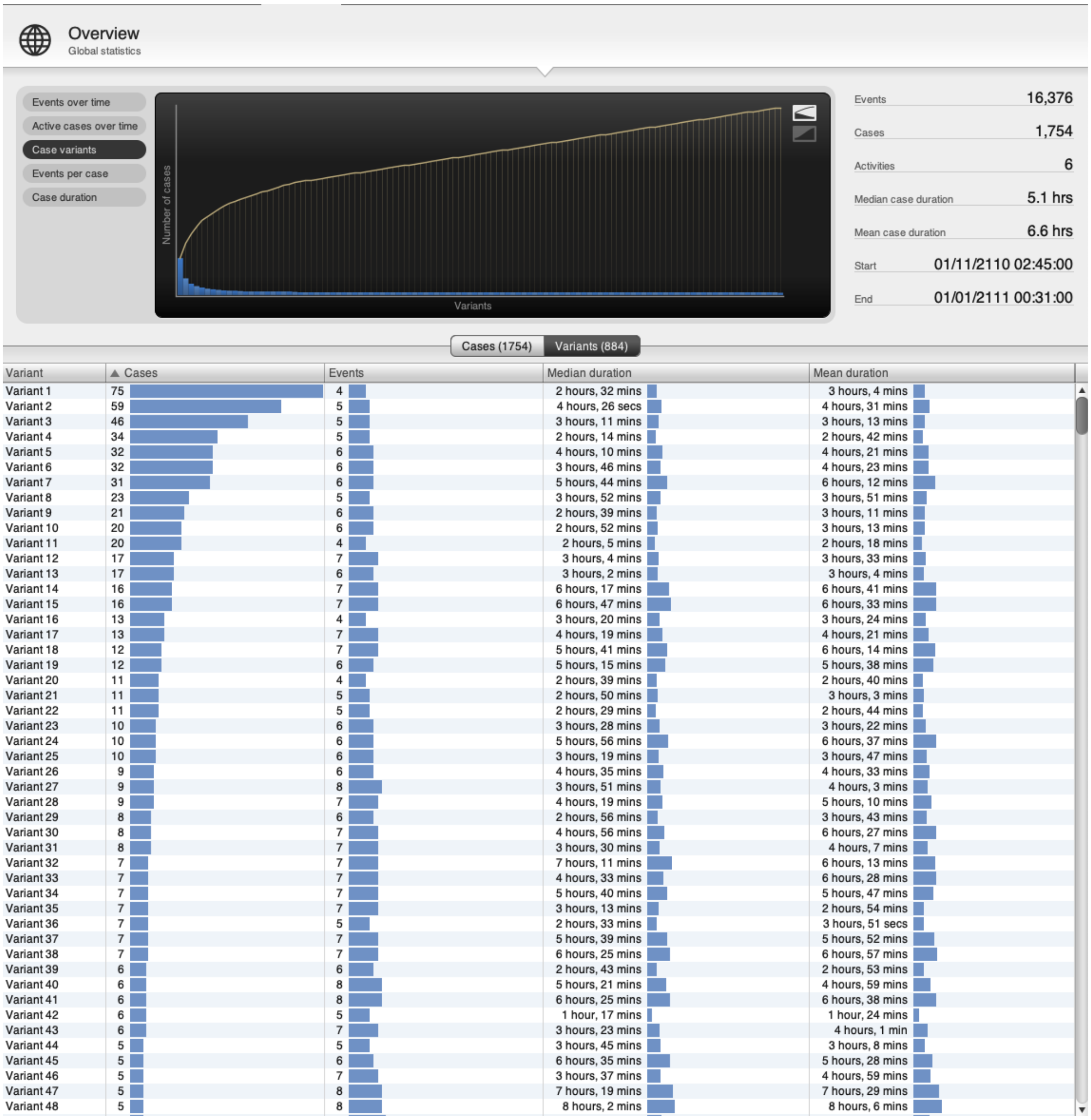
Performance DFG generated (shows average time between activities)

Conformance checking revealed another paradoxical result: perfect fitness (1.0) with moderate precision (0.71). While all 1754 cases conform to the discovered model (**zero deviant cases**), the model is **overly permissive**, allowing process flows never observed in the log. This reflects the fundamental issue: **the ED lacks a normative process definition**. The model describes "what happens" across 884 variants but cannot prescribe "what should happen" because no standard pathway exists.

In a well-structured process, one would expect 5-10 standard variants covering 80% of cases (Pareto principle as we said before), fitness around 0.90-0.95, precision values around 0.85-0.95 (model represents standards tightly), and 5-10% deviant cases requiring investigation.

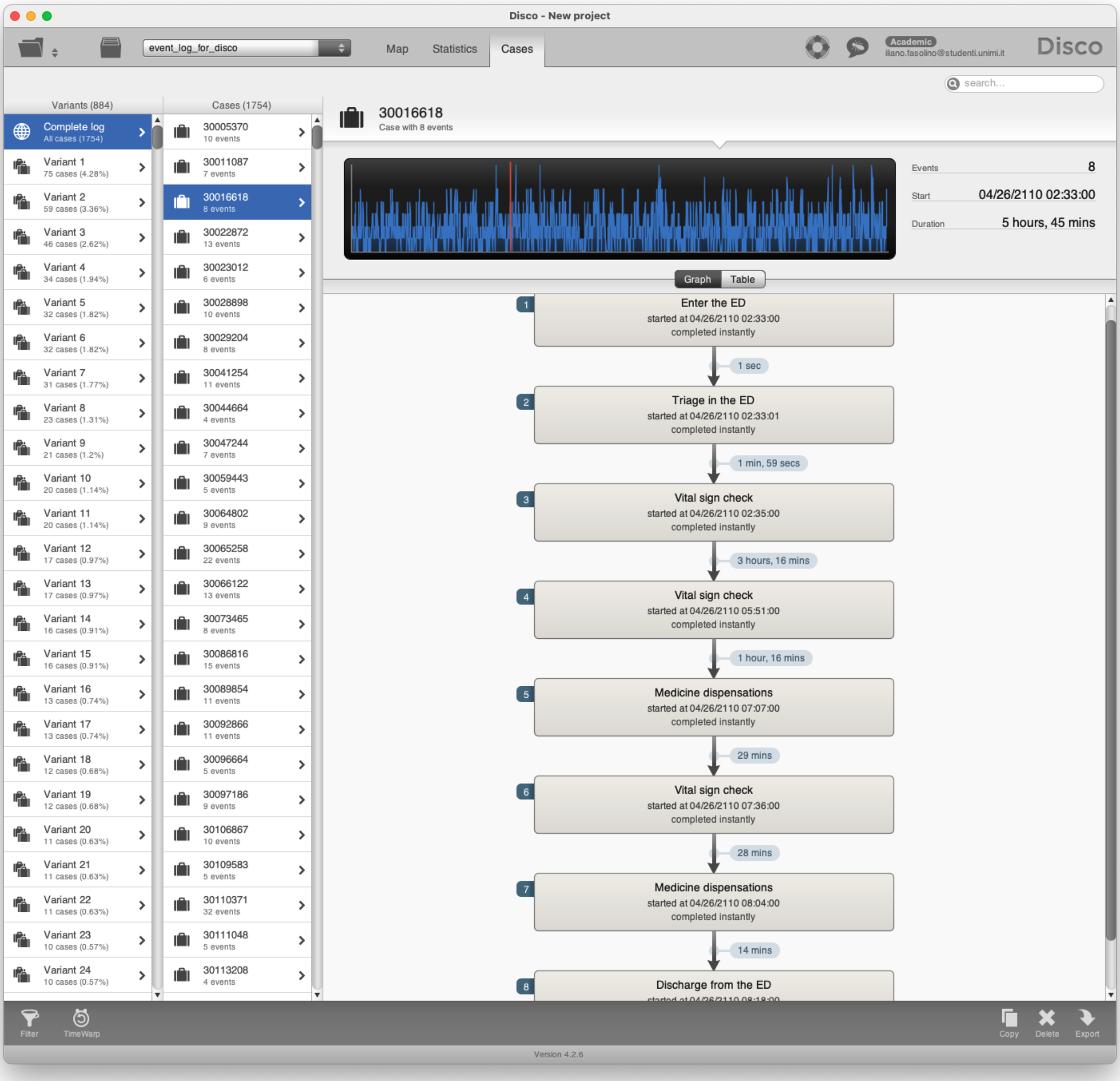
Here we can observe the **inverse trend**: 884 variants, perfect fitness, low precision, zero deviants, indicating a sort of reactive, ad-hoc care rather than protocol-driven standardization.

5 Additional Visuals from Disco

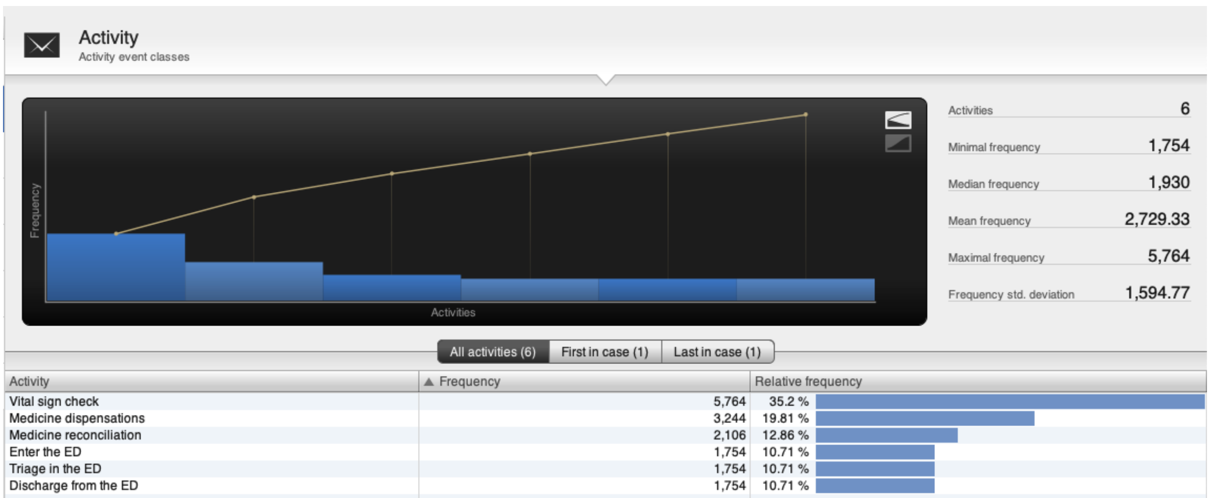


Variant	▲ Cases	Events	Median duration	Mean duration
Variant 1	75	4	2 hours, 32 mins	3 hours, 4 mins
Variant 2	59	5	4 hours, 26 secs	4 hours, 31 mins
Variant 3	46	5	3 hours, 11 mins	3 hours, 13 mins
Variant 4	34	5	2 hours, 14 mins	2 hours, 42 mins
Variant 5	32	6	4 hours, 10 mins	4 hours, 21 mins
Variant 6	32	6	3 hours, 46 mins	4 hours, 23 mins
Variant 7	31	6	5 hours, 44 mins	6 hours, 12 mins
Variant 8	23	5	3 hours, 52 mins	3 hours, 51 mins
Variant 9	21	6	2 hours, 39 mins	3 hours, 11 mins
Variant 10	20	6	2 hours, 52 mins	3 hours, 13 mins
Variant 11	20	4	2 hours, 5 mins	2 hours, 18 mins
Variant 12	17	7	3 hours, 4 mins	3 hours, 33 mins
Variant 13	17	6	3 hours, 2 mins	3 hours, 4 mins
Variant 14	16	7	6 hours, 17 mins	6 hours, 41 mins
Variant 15	16	7	6 hours, 47 mins	6 hours, 33 mins
Variant 16	13	4	3 hours, 20 mins	3 hours, 24 mins
Variant 17	13	7	4 hours, 19 mins	4 hours, 21 mins
Variant 18	12	7	5 hours, 41 mins	6 hours, 14 mins
Variant 19	12	6	5 hours, 15 mins	5 hours, 38 mins
Variant 20	11	4	2 hours, 39 mins	2 hours, 40 mins
Variant 21	11	5	2 hours, 50 mins	3 hours, 3 mins
Variant 22	11	5	2 hours, 29 mins	2 hours, 44 mins
Variant 23	10	6	3 hours, 28 mins	3 hours, 22 mins
Variant 24	10	6	5 hours, 56 mins	6 hours, 37 mins
Variant 25	10	6	3 hours, 19 mins	3 hours, 47 mins
Variant 26	9	6	4 hours, 35 mins	4 hours, 33 mins
Variant 27	9	8	3 hours, 51 mins	4 hours, 3 mins
Variant 28	9	7	4 hours, 19 mins	5 hours, 10 mins
Variant 29	8	6	2 hours, 56 mins	3 hours, 43 mins
Variant 30	8	7	4 hours, 56 mins	6 hours, 27 mins
Variant 31	8	7	3 hours, 30 mins	4 hours, 7 mins
Variant 32	7	7	7 hours, 11 mins	6 hours, 13 mins
Variant 33	7	7	4 hours, 33 mins	6 hours, 28 mins
Variant 34	7	7	5 hours, 40 mins	5 hours, 47 mins
Variant 35	7	7	3 hours, 13 mins	2 hours, 54 mins
Variant 36	7	5	2 hours, 33 mins	3 hours, 51 secs
Variant 37	7	7	5 hours, 39 mins	5 hours, 52 mins
Variant 38	7	7	6 hours, 25 mins	6 hours, 57 mins
Variant 39	6	6	2 hours, 43 mins	2 hours, 53 mins
Variant 40	6	8	5 hours, 21 mins	4 hours, 59 mins
Variant 41	6	8	6 hours, 25 mins	6 hours, 38 mins
Variant 42	6	5	1 hour, 17 mins	1 hour, 24 mins
Variant 43	6	7	3 hours, 23 mins	4 hours, 1 min
Variant 44	5	5	3 hours, 45 mins	3 hours, 8 mins
Variant 45	5	6	6 hours, 35 mins	5 hours, 28 mins
Variant 46	5	7	3 hours, 37 mins	4 hours, 59 mins
Variant 47	5	8	7 hours, 19 mins	7 hours, 29 mins
Variant 48	5	8	8 hours, 2 mins	8 hours, 6 mins

Overview of Case variants

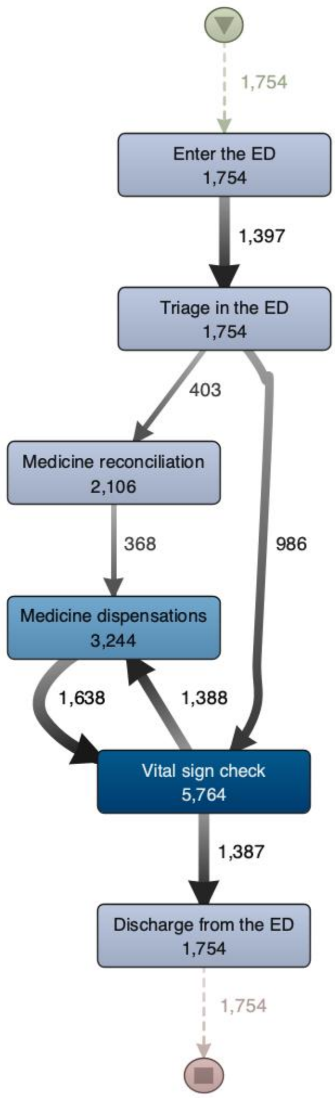


Sample Case



Activity	Frequency	Relative frequency
Vital sign check	5,764	35.2 %
Medicine dispensations	3,244	19.81 %
Medicine reconciliation	2,106	12.86 %
Enter the ED	1,754	10.71 %
Triage in the ED	1,754	10.71 %
Discharge from the ED	1,754	10.71 %

Activity Frequency



Event log Tree

6 Conclusion and Improvement Suggestions

The ED data analysis through the previous steps revealed significant structural inefficiencies requiring targeted interventions. The mean throughput time of 6.58 hours exhibits high variability (standard deviation of 5.42 hours), with 10% of patients remaining beyond 12.4 hours.

The analysis identified as the central problem not so much non-compliance with existing processes, but rather the **absence of standardized processes**: with 884 unique variants and the most common variant representing **only** 4.3% of cases, the department operates in **reactive** mode rather than following protocols.

The most critical finding emerges from the acuity level segmentation analysis: patients classified as Urgent (Acuity 2) experience a mean throughput time of 8.03 hours, 38% higher than Critical patients (Acuity 1, 5.83 hours). This sort of paradox represents a counter-intuitive inversion of clinical priorities and indicates a **severe organizational inefficiency**. Bottleneck identification highlighted that time accumulates primarily in the vital sign monitoring loop (93 minutes mean between consecutive checks, with 2159 occurrences) and in the final discharge process (48 minutes mean after the last clinical activity).

6.1 Scenario 1: Creation of Standard Care Pathways Based on Acuity Level

The extreme fragmentation of the ED process represents the most pervasive systemic problem. The existence of 884 unique variants indicates that no **standardized** care pathways exist and that each patient follows an **almost unique path**. This lack of standardization has multiple consequences: it makes protocol-based training impossible, prevents resource optimization, and creates a paradox where urgent patients end up in an "organizational limbo" receiving less attention than critical cases.

The proposed solution consists of defining four distinct care pathways, each optimized for a specific acuity level. These are not rigid prescriptive sequences, but rather frameworks that define expected activities, target timeframes, and dedicated resources for each patient category. The goal is to reduce variants from 884 to fewer than 50, maintaining the clinical flexibility necessary to manage the complexity of individual cases but eliminating non-value-added variability resulting from lack of coordination.

The **Pathway for Critical patients (Acuity 1)** is designed to ensure the fastest possible throughput for life-threatening conditions. The ideal path includes immediate triage within 5 minutes of arrival, followed by treatments without intermediate waits. These patients require continuous vital sign monitoring every 15-30 minutes and a rapid decision (within 60-90 minutes) regarding admission to intensive care, operating room, or transfer. The throughput target is maintained at 4-5 hours, roughly in line with current performance (5.83 hours) but with reduced variability. The key is to **eliminate any wait between steps** through dedicated resources and fast-track protocols.

The **Pathway for Urgent patients (Acuity 2)** represents the **absolute** priority of this intervention, given that this category manifests the most severe paradox (8.03 hours vs 5.83 hours for Acuity 1). These patients require a structured approach that balances urgency and diagnostic necessity. The standard pathway begins with triage within 10 minutes, followed by initial assessment and diagnostic test ordering within 30 minutes. The diagnostic phase (laboratory and imaging) should complete within 60-90 minutes, followed by results-based treatment (30-60 minutes), reassessment (30 minutes, if necessary), and final decision regarding discharge or admission (approximately 30 additional minutes). The overall target is to reduce throughput to 5-6 hours, a 25-31% reduction from the current state.

The key to success of this pathway is creating a dedicated "*Urgent Care Team*" that is not continuously redirected toward critical emergencies. This is because Acuity 2 patients are often started in the pathway but then "parked" when Acuity 1 cases arrive, accumulating delays. The dedicated team, supported by fast-track diagnostics (also for laboratory-related actions), ensures continuity of care. It is also fundamental to anticipate the admission decision: instead of waiting for all tests to complete before deciding, the protocol includes early evaluation (within 4 hours) to identify early on patients who will require admission, enabling advance coordination with inpatient units.

The **Pathway for Less Urgent patients (Acuity 3)** recognizes that for this category (representing 48.5% of total volume) some waiting time is acceptable, but the pathway must still be efficient once started. After triage (0-15 minutes) and possible waiting room time, medical team assessment should complete in 20-30 minutes. If diagnostic tests are needed, these are performed with standard priority (30-60 minutes), followed by treatment (30 minutes) and discharge instructions (10 minutes). The target is to reduce throughput from the current 6.30 hours to 4-5 hours, primarily through operational efficiency and nursing protocols for minor conditions that do not require complex medical assessment.

The **Fast-Track Pathway (Acuity 4-5)** leverages the fact that these patients already experience relatively brief throughput (3.08 hours). The opportunity is to create a completely separate lane, physically distinct from the main ED, managed by interns and physician assistants following standardized protocols for minor conditions (sprains, minor lacerations, flu-like symptoms). The ideal pathway includes quick triage (0-5 minutes), provider assessment (10-15 minutes), rapid treatment (15-20 minutes), and discharge (5 minutes), for a total target under 2 hours. Although this category represents only 8% of volume, physical separation frees resources in the main ED for more complex cases. Also important is the impact of **moderate** but not substantial expense to face, along with training time and organization of an existing space (which may take a few months) or in worst cases construction of a dedicated space (6-12 months).

Implementation of these pathways requires a **phased approach**. The **pilot** phase (1-2 months) focuses on Pathways 2 (Acuity 2) and 4 (Fast-track), the two with greatest potential for immediate impact. During this period, KPIs are monitored weekly (rather than monthly) and protocols are refined based on staff feedback. The **full implementation** phase (3-6 months) extends pathways to all acuity levels, including complete staff training and information system updates to support new flows. After that, one can move to the final phase of continuous monitoring that uses process mining (this time monthly) and real-time KPI dashboards to verify adherence to pathways and identify deviations requiring intervention.

Expected outcomes are significant: variant reduction from 884 to fewer than 50 (with 80% of cases following the top 10 pathways), resolution of the Acuity 2 paradox with throughput reduced to 5.5 hours (lower than Acuity 1 and Acuity 3), and improvement of process model precision from 0.71 to above 0.85, indicating that the process is now **defined** and **standardized**.

6.2 Scenario 2: Optimization of the Discharge Process

The analysis revealed that the mean time between the last clinical activity and formal discharge is 48 minutes, with high variability (standard deviation of 61 minutes) and extreme cases reaching 820 minutes (13.7 hours!!). This time represents, as mentioned before, a "non-value-added" window from the patient's perspective: clinically ready to go home but still retained in ED for purely administrative and logistical reasons. In 79% of cases, the last clinical activity is a vital sign check, suggesting that physicians await confirmation of patient stability before proceeding with discharge, but the long subsequent delay indicates inefficiencies in the post-decision process.

The causes of this delay were identified through flow analysis and informal staff interviews. Manual paperwork for discharge summary and prescriptions requires significant time: the physician must manually complete the summary, write prescriptions, and prepare patient instructions, often while simultaneously managing other patients. Waiting for prescriptions from the pharmacy creates a second delay: even when prescriptions are written promptly, the patient must wait for the pharmacy to prepare them. Patient education, when not prepared in advance, requires improvised time: explaining home care instructions, warning signs to monitor, and follow-up appointments. Finally, waiting for transportation, particularly for patients requiring ambulance or ride service, can add significant delays if not coordinated in advance.

The proposed solution consists of **three complementary interventions** at relatively **low** cost.

Anticipatory Discharge Planning (ADP) redefines the process by shifting the start of discharge preparation from when the patient is ready (current approach) to when the clinical decision to discharge is made. Concretely, when a physician decides that a patient will be discharged after the last vital sign check, the nurse immediately begins preparing the discharge summary using pre-populated templates. While the patient awaits the last check (which serves to confirm stability), the nurse completes the paperwork and the pharmacy is alerted to prepare prescriptions in advance. Patient education is prepared using standardized materials by condition, and if necessary transportation is organized proactively. When the time for formal discharge arrives, everything is ready and the patient can leave immediately. This process redesign requires only training and change management, with low cost and timeline of 1-2 months (as well as bringing advantages to the experience, a point I will address shortly).

The **Digital Discharge Process** instead introduces technological automation where possible. The electronic discharge summary (**EDS** - Electronic Discharge Summary) uses auto-populated templates from the EHR system with clinical data already recorded during the stay, requiring the provider only to verify and add specific notes instead of rewriting everything manually. "E-prescriptions" sent directly to the patient's pharmacy eliminate in-hospital waiting and reduce transcription errors. Digital patient instructions sent via SMS or email allow the patient to access instructions at home, reducing time needed for detailed ED explanations. Implementation requires integration with existing EHR and has medium cost with timeline of 3-6 months, but benefits include not only speed but also error reduction and improved patient satisfaction.

Nurse-Led Discharge (NLD) for appropriate cases represents an opportunity to free physicians from tasks that do not require complex clinical judgment. For Acuity 4-5 patients with simple conditions and clear discharge plan, specially trained nurses can manage the entire discharge process with remote provider approval (through electronic documentation review). The goal here is not only discharge for these patients but freeing physician time for more complex cases. Implementation requires development of specific protocols that define which cases are appropriate for nurse-led discharge and targeted nursing staff training, with low cost and timeline of 2-3 months.

Performance targets are ambitious but achievable. In the short term (3 months), the goal is to reduce mean time-to-discharge to 30 minutes (38% reduction), while in the medium term (6 months) the target is 20 minutes (58% reduction). These goals are based on case studies of EDs that have successfully implemented anticipatory discharge planning and digitalization[2][3], typically achieving reductions of 50-60%.

[2] Adrian Baker et al. "Anticipatory care planning and integration: a primary care pilot study aimed at reducing unplanned hospitalisation". In: *The British Journal of General Practice* 62.595 (2012), e113

[3] Attakrit Leckcivilize et al. "Impact of an anticipatory care planning intervention on unscheduled acute hospital care using difference-in-difference analysis". In: *BMJ health & care informatics* 28.1 (2021), e100305

The impact on overall ED throughput is estimated at a 5% reduction, apparently modest but significant considering it affects all patients (100% of cases go through discharge). More importantly, the impact on patient experience is substantial, as discharge is the "last emotional note" the patient has of the ED, and reducing final waiting from almost an hour to 15-20 minutes significantly improves the overall perception of service; and at the same time from an operational perspective, reducing discharge time more quickly frees spaces and resources for new patients, improving overall flow.

6.3 Considerations about the project

This project demonstrated how through process mining raw data are transformed into real insights that guide and prompt reflection on strategies that can be implemented always based on evident facts, the data. Quantitative analysis left no room for subjective perceptions, but rather revealed precise patterns and structural inefficiencies that without it would have remained hidden or unclear in the great complexity of processes and values.

The main value of this work lies in the precise quantification of problems required by the assignment and proposed solutions. Obviously, it is not sufficient to observe that "the process is slow" or "there are bottlenecks," but thanks to process mining I demonstrated that Acuity 2 waits precisely 2.2 hours more than Acuity 1, that 79% of final time is spent on post-clinical administrative activities, and that 884 different ways exist to traverse the process. At the same time, my proposed solutions are not generic recommendations ("improve efficiency") but specific strategies (also based on various consulted scientific papers, present in various footnotes) with measurable targets: reduce variants to fewer than 50, bring Acuity 2 below 5.5 hours, reduce time-to-discharge to 20 minutes.

This project sheds light on the concept that process mining is not simply an analytical tool but a true "strategic trigger" for organizational transformation in healthcare based on objective data.

6.4 Knowledge Uplift Trail

Step	Phase	Main Activities	Key Outputs	Critical Decisions
Step 1	**Preprocessing**	- Dataset loading (25,115 events, 1,820 cases) - Dimension and missing value analysis - Simultaneous event aggregation - Forward/backward fill for case-level attributes - Noise filtering (incomplete cases, <4 events, temporal outliers)	- Clean dataset: 1,754 cases, 16,376 events - 9 selected columns (main + case-level) - Missing values reduced: 89% $\rightarrow$ 0% - 66 cases removed (3.6%) as noise	- Simultaneous events: aggregate for macro-flow focus, not granular detail - Missing values: distinguish case-level vs event-level - Noise definition: incomplete cases, <4 events, duration <30mins or >48h
Step 2	**Performance Analysis**	- Throughput time analysis - Time-to-triage - Time-to-discharge - Process variant identification - Acuity-based segmentation - Bottleneck identification (through transition times)	- Throughput: 6.58h mean, high variability (std 5.42h) - Time-to-triage: 1 sec (artifact, unusable KPI) - Time-to-discharge: 48 min mean - 884 unique variants (top: 4.3%) - **Paradox**: Acuity 2 (8.03h) > Acuity 1 (5.83h) - Primary bottleneck: Vital $\rightarrow$ Vital (93 min, 2159 occurrences)	- Priority metrics: throughput, bottleneck, acuity segmentation - Visualizations via matplotlib/seaborn - Pareto violation confirmed: extreme fragmentation

Step 3	**Process Discovery & Conformance**	- Conversion to PM4Py event log format - Process discovery with Inductive Miner - Visualizations: Petri Net, DFG (frequency + performance) - Conformance checking (fitness, precision, generalization, simplicity) - Deviant case analysis	- Petri Net: 15 places, 18 transitions, 40 arcs - Fitness: 1.0000 (perfect) - Precision: 0.7139 (moderate) - Generalization: 0.9694 - Simplicity: 0.7021 - Deviant cases: 0 (all conformant) - Self-loop confirmed on Vital sign check	- Tool selection: PM4Py for integrated workflow (Disco additional for other visuals) - Algorithm: Inductive Miner for noise robustness - Conformance Paradox: perfect fitness but low precision = absence of process standard, not violation - Zero deviants with 884 variants = fragmentation
Step 4	**Potential Improvements**	- Synthesis of critical findings from Steps 1-3 - Development of 2 scenarios/proposals: 1. Pathway standardization 2. Discharge optimization process - KPI and timeline definition - Expected impact quantification	- Proposal 1: 4 acuity pathways, target 884→<50 variants - Proposal 2: Anticipatory discharge + digital, target 48→20 min	- Primary focus: Acuity 2 pathway to resolve paradox