Sepsis Cases

Aleksander Chotecki

Business Information Systems
Prof. Paolo Ceravolo
a.a. 2023 - 2024

Description of the case study

Description

The case study describes the process of treatment sepsis in a hospital environment. It is a condition which is triggered by the body’s reaction to an infection. By analyzing this log, effective diagnosis of the condition and treatment may be improved. Moreover, stating optimal process management is crucial for enhancing speed of the process and therefore patient survival rate and minimizing potential negative outcomes for the patient's health. The knowledge one get’s by analyzing the log could be applied in hospitals all over the world.

Actors

There are many actors involved in the process that might influence it and might benefit from event log analysis such as:

- **patients** - their treatment is documented in the event log and they are the subject of the analysis
- **medical staff** - their actions are documented in the event log and their decisions influence patient treatment
- **hospital management staff -** they are the ones responsible for how the treatment process works and how it should be changed to improve patient care

Structure of the dataset

The dataset is an event of sepsis cases from a hospital. Single case represents the pathway through the hospital. Moreover, 36 data attributes are recorded, e.g., the group responsible for the activity, the results of tests and information from checklists. There are 1050 cases with in total 15214 events that were recorded for 16 different activities. The attributes include:

- **InfectionSuspected** - Indicates whether an infection is suspected.
- **org:group** - Organizational group responsible for the case.
- **DiagnosticBlood** - Blood diagnostic test results.
- **DisfuncOrg** - Indicates organ dysfunction.
- **SIRSCritTachypnea** - Tachypnea (rapid breathing) criterion as SIRS (Systemic Inflammatory Response Syndrome)
- **Hypotensie** - Low blood pressure (hypotension).
- **SIRSCritHeartRate** - High heart rate as a SIRS (Systemic Inflammatory Response Syndrome) criterion.
- **Infusion** - Details on fluid infusion treatments.
- **DiagnosticArtAstrup** - Arterial blood gas (Astrup) test results.
- **concept:name** - Name or type of event.
- **Age** - Age of the patient.
- **DiagnosticIC** - Intensive care diagnostic tests.

- **DiagnosticSputum** - Sputum culture (phlegm that is coughed up) test results.
- **DiagnosticLiquor** - Cerebrospinal fluid (fluid found in the brain and spinal cord) test results.
- **DiagnosticOther** - Other diagnostic tests.
- **SIRSCriteria2OrMore** - Tells if two or more SIRS (Systemic Inflammatory Response Syndrome) criteria are met.
- **DiagnosticXthorax** - Chest X-ray diagnostic results.
- **SIRSCritTemperature** - Too high/low temperature as a SIRS (Systemic Inflammatory Response Syndrome) criterion.
- **time:timestamp** - timestamp of the event.
- **DiagnosticUrinaryCulture** - Urinary culture test results.
- **SIRSCritLeucos** - Too high/low leukocyte count as a SIRS (Systemic Inflammatory Response Syndrome) criterion.
- **Oligurie** - Condition of oliguria (low urine output).
- **DiagnosticLacticAcid** - Lactic acid test results.
- **Diagnose** - Diagnosis given to the patient (1 or 2 letters, given only during ER Registration).
- **Hypoxie** - Presence of hypoxia (low oxygen levels).
- **DiagnosticUrinarySediment** - Urinary sediment test results.
- **DiagnosticECG** - Electrocardiogram (ECG) test results.
- **@@classifier** - Classifier for event type (activity).
- **case:concept** - Unique identifier for the case.
- **Leucocytes** - Leukocyte (white blood cell) count.
- **CRP** - C-reactive protein (CRP) level, an inflammation marker.
- **LacticAcid** - Level of lactic acid in the blood.
- **start_timestamp** - Start timestamp of the event or case.
- **@@event_index** - Index of the event in the log.

The activities include:

- **Leucocytes** - Leukocyte (WBC - white blood cell) count.
- **CRP** - C-reactive protein (CRP) level, an inflammation marker.
- **LacticAcid** - Level of lactic acid in the blood.
- **Admission NC -** Admission to Non-Critical care
- **ER Triage -** Emergency Room Triage (prioritizing patients based on the severity of their conditions)
- **ER Registration -** Emergency Room Registration
- **ER Sepsis Triage -** specialized triage process (identifying and prioritizing patients with suspected sepsis)
- **IV Antibiotics -** Intravenous antibiotics for the patient
- **IV Liquid -** Intravenous fluids (dehydration treatment)
- **Return ER -** Patient returning to the emergency room

- **Admission IC -** Admission to Intensive Care
- **Release* -** Discharge of a patient from the hospital

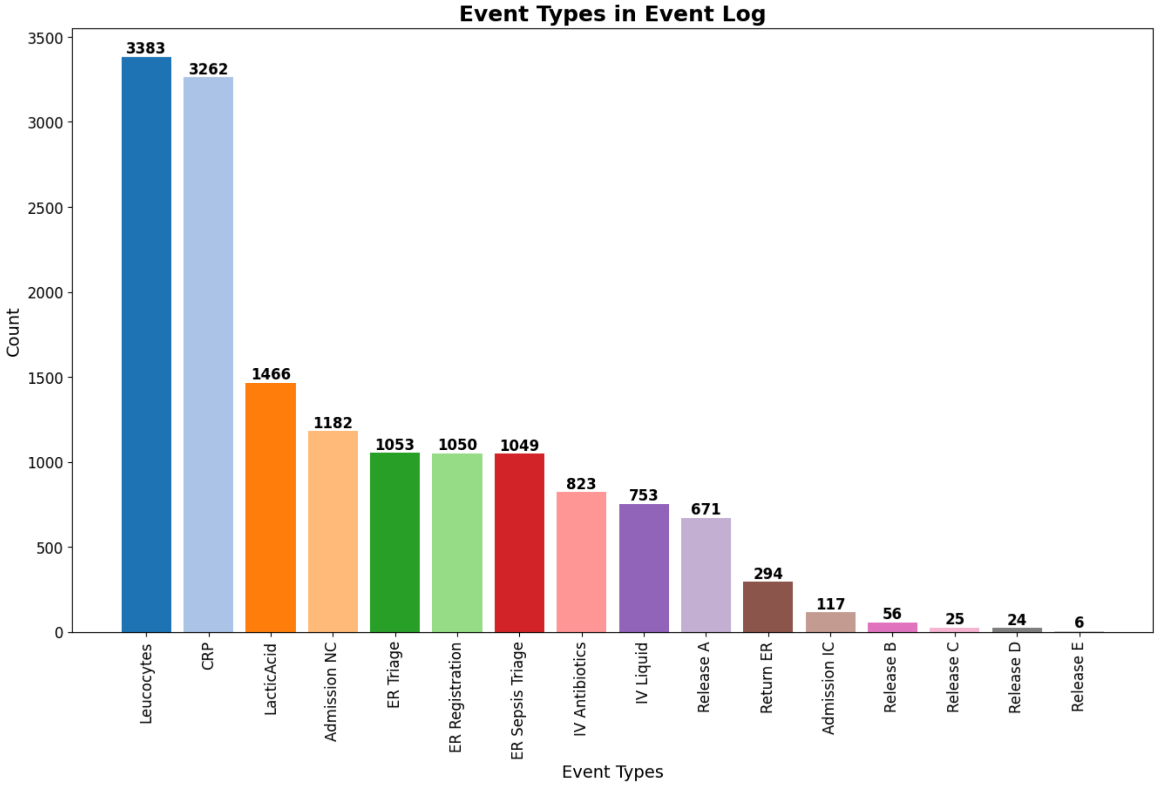


Figure 1 - Number of different event types (Activities) before filtering

We can see that some activities are more popular than others and there are different types of Release (A-E) and we don't have any further context for them.

Peculiarities

This dataset has some unique features. It's highly complex and needs advanced analysis due to its high number of variables. The quality and completeness of the data are crucial because sepsis treatment involves time-sensitive actions. There is also significant variability in treatment protocols and patient characteristics, as sepsis patients have different underlying conditions, ages, and responses to treatment.

Organizational goals

Strategic objectives

Goal	Description
Improve patients treatment	Improve patient survival rates and recovery time
Enhance organizational efficiency	Optimize process management and resource allocation
Optimize hospital resource utilization	Choose methods that minimize the cost of treatment

Operational objectives

Goal	Description
Increase condition identification speed	Identify bottlenecks and inefficiencies in the process of sepsis treatment. Identify key metrics that influence successful sepsis treatment
Standardize reaction protocols	Use analysis to standardize treatment protocol
Decrease the number of returning patients	Establish protocols that will decrease the number of cases with Return ER events.

Tactical objectives

Goal	Description
Continuous analysis of the process	Analyzing the event log and the compliance with the new protocol. Using process mining techniques.
Introduce decision support system	Make some kind of support software to support decisions real-time
Train the staff	Train the staff to the improved protocol
Improve data collection	Improve the quality and consistency of future event logs.

Knowledge Uplift Trail

The Knowledge Uplift Trail (KUT) is the method of transforming raw data, in this case event log, into useful insights. It uses a sequence of methods that are used on the data to derive necessary information and therefore acquire new knowledge. The steps are:

1) **Data cleaning and filtering -** The event log needs to be examined, filtered from wrong/inconsistent values and filled with consistent values. We will filter out the start/end activities that do not follow the standard treatment process.
2) **Descriptive analysis -** The relevant data will be plotted and visualized. Relevant statistical tools will be used to understand the characteristics of the data.
3) **Process mining -** Discovering models of the process from the filtered log and identifying key activities and sequences.
4) **Conformance checking -** identifies whether the model is in alignment with event logs
5) **Intervention strategies -** suggesting modification of existing processes and its optimization. Continuous monitoring of performance.

Project Results

1) Data cleaning and filtering

Looking at the not filtered log we have 1050 cases. We will try to examine what are the most common start/end activities and which ones should be filtered out to clean the data.

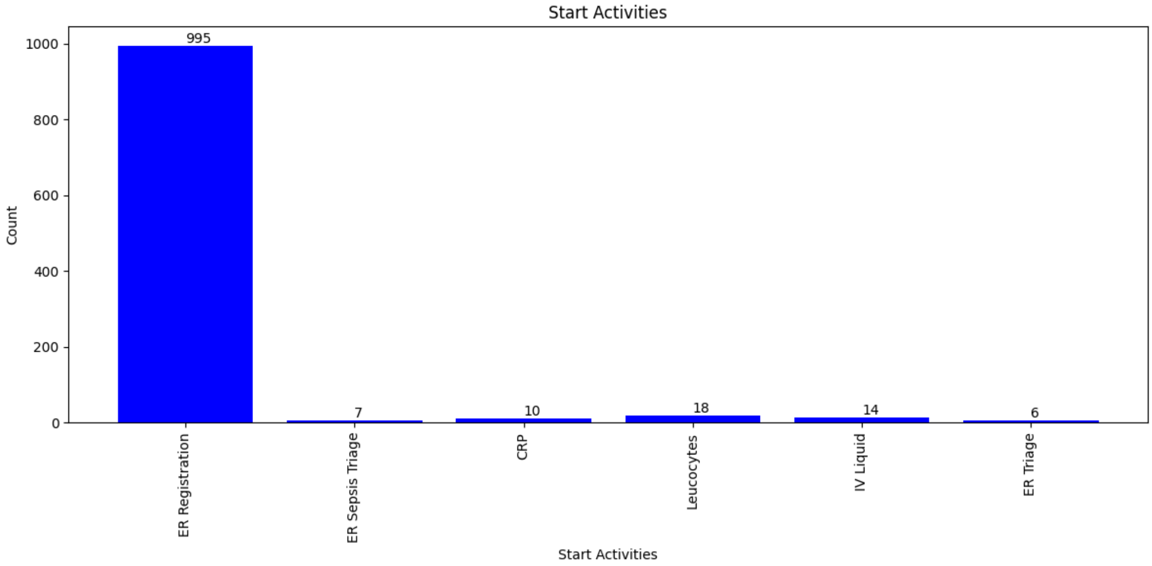


Figure 1 - Distribution of start activities

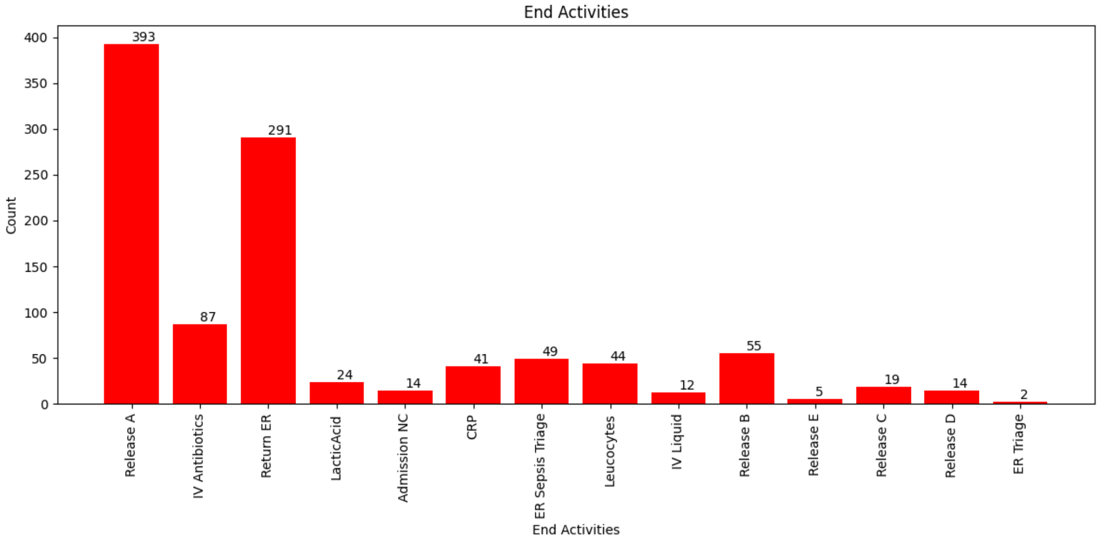


Figure 2 - Distribution of end activities

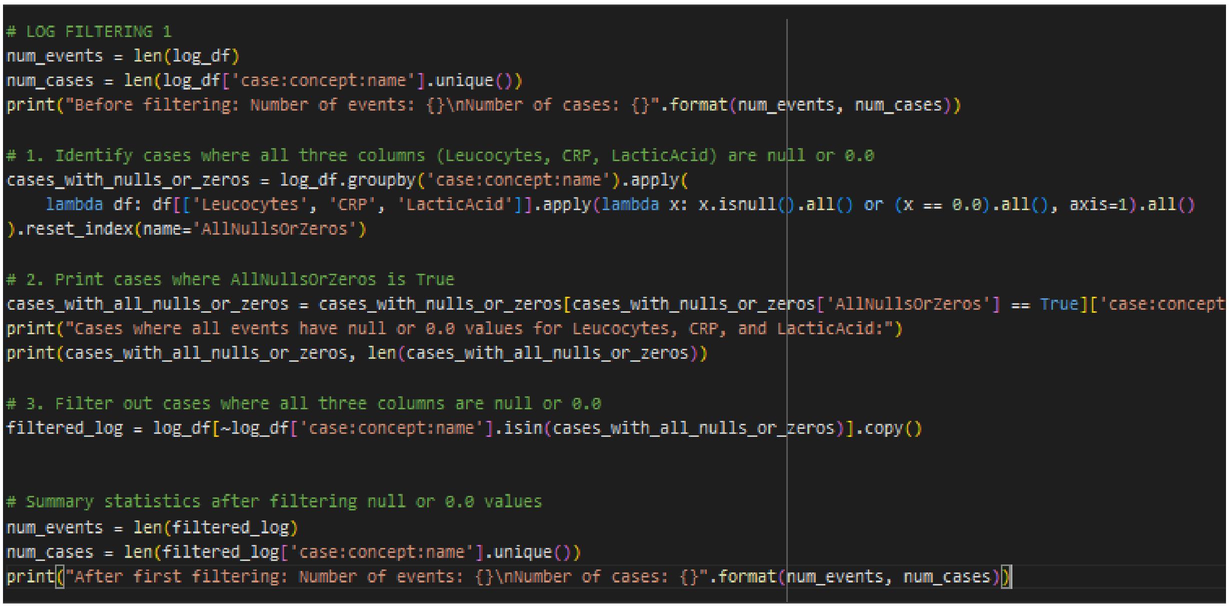
To simplify I will consider the proper cases that:

- have at least one examination done defined as at least one value for one of the columns 'Leukocytes', 'CRP', 'LacticAcid'
- start with “ER Registration”, “ER Sepsis Triage”, “ER Triage” and those that end with “Release*”, “Return ER”
- have at least 4 events per case

Also, cases that will explicitly be removed are those that:

- have duration of case more than 80 days - this eliminates the most extreme cases in terms of case duration (outliers)

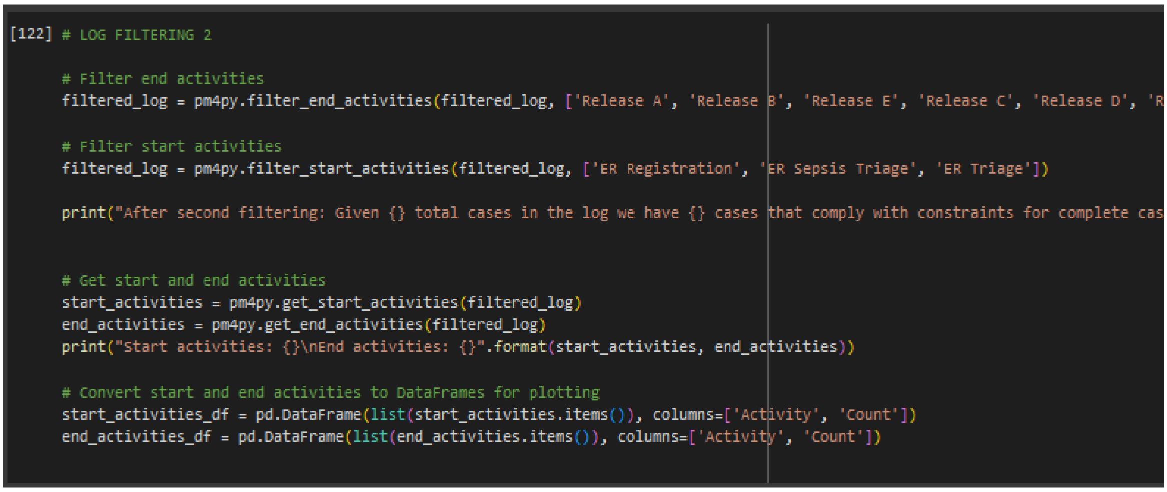
Before filtering the number of events is 15214 and the number of cases is 1050.



```
# LOG FILTERING 1
num_events = len(log_df)
num_cases = len(log_df['case:concept:name'].unique())
print("Before filtering: Number of events: {}\nNumber of cases: {}".format(num_events, num_cases))

# 1. Identify cases where all three columns (Leucocytes, CRP, LacticAcid) are null or 0.0
cases_with_nulls_or_zeros = log_df.groupby('case:concept:name').apply(
    lambda df: df[['Leucocytes', 'CRP', 'LacticAcid']].apply(lambda x: x.isnull().all() or (x == 0.0).all(), axis=1).all()
).reset_index(name='AllNullsOrZeros')

# 2. Print cases where AllNullsOrZeros is True
cases_with_all_nulls_or_zeros = cases_with_nulls_or_zeros[cases_with_nulls_or_zeros['AllNullsOrZeros'] == True]['case:concept
print("Cases where all events have null or 0.0 values for Leucocytes, CRP, and LacticAcid:")
print(cases_with_all_nulls_or_zeros, len(cases_with_all_nulls_or_zeros))

# 3. Filter out cases where all three columns are null or 0.0
filtered_log = log_df[~log_df['case:concept:name'].isin(cases_with_all_nulls_or_zeros)].copy()


# Summary statistics after filtering null or 0.0 values
num_events = len(filtered_log)
num_cases = len(filtered_log['case:concept:name'].unique())
print("After first filtering: Number of events: {}\nNumber of cases: {}".format(num_events, num_cases))
```

Figure 3 - filtering the log file no values for columns 'Leukocytes', 'CRP', 'LacticAcid'

After applying the constraints there are 15090 events in the log and 1012 cases. Now let’s apply the second filter that will take into consideration starting and ending events. We filter for those cases that start with “ER Registration”, “ER Sepsis Triage”, “ER Triage” and those that end with “Release*”, “Return ER”.



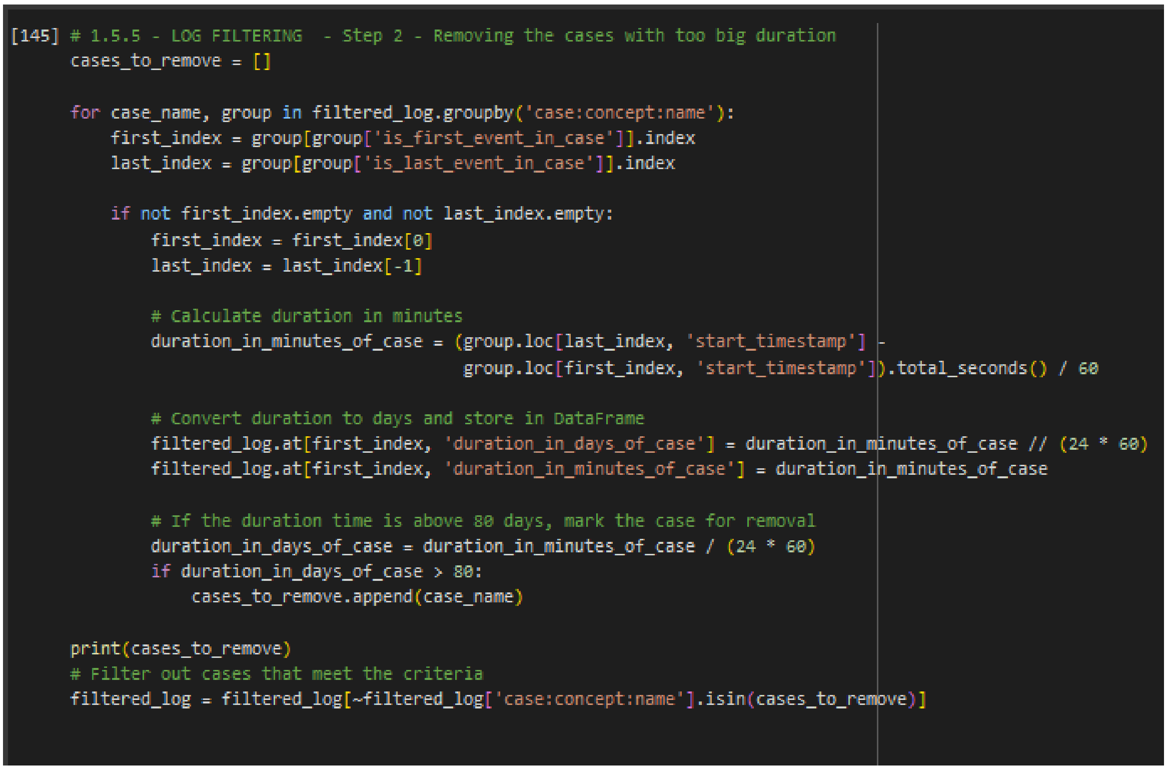
```
[122] # LOG FILTERING 2

      # Filter end activities
      filtered_log = pm4py.filter_end_activities(filtered_log, ['Release A', 'Release B', 'Release E', 'Release C', 'Release D', 'R

      # Filter start activities
      filtered_log = pm4py.filter_start_activities(filtered_log, ['ER Registration', 'ER Sepsis Triage', 'ER Triage'])

      print("After second filtering: Given {} total cases in the log we have {} cases that comply with constraints for complete cas


      # Get start and end activities
      start_activities = pm4py.get_start_activities(filtered_log)
      end_activities = pm4py.get_end_activities(filtered_log)
      print("Start activities: {}\nEnd activities: {}".format(start_activities, end_activities))

      # Convert start and end activities to DataFrames for plotting
      start_activities_df = pd.DataFrame(list(start_activities.items()), columns=['Activity', 'Count'])
      end_activities_df = pd.DataFrame(list(end_activities.items()), columns=['Activity', 'Count'])
```

Figure 4 - filtering the log file considering start and end activities

Applying the filter requiring at least 4 events for each of the cases was also performed but it didn’t remove any cases, because after already applied filters there were no such cases.

After filtering the log with the previous steps we have 741 cases that comply with constraints for complete cases and 12594 events.

The last stages of the filtering process will be filtering the cases with too big duration of stay in the hospital. For this we will create additional columns in our log file like: “is_first_event_in_case”, “is_last_event_in_case” or “duration_in_days_of_case”.



```
[145] # 1.5.5 - LOG FILTERING  - Step 2 - Removing the cases with too big duration
      cases_to_remove = []

      for case_name, group in filtered_log.groupby('case:concept:name'):
          first_index = group[group['is_first_event_in_case']].index
          last_index = group[group['is_last_event_in_case']].index

          if not first_index.empty and not last_index.empty:
              first_index = first_index[0]
              last_index = last_index[-1]

              # Calculate duration in minutes
              duration_in_minutes_of_case = (group.loc[last_index, 'start_timestamp'] -
                                             group.loc[first_index, 'start_timestamp']).total_seconds() / 60

              # Convert duration to days and store in DataFrame
              filtered_log.at[first_index, 'duration_in_days_of_case'] = duration_in_minutes_of_case // (24 * 60)
              filtered_log.at[first_index, 'duration_in_minutes_of_case'] = duration_in_minutes_of_case

              # If the duration time is above 80 days, mark the case for removal
              duration_in_days_of_case = duration_in_minutes_of_case / (24 * 60)
              if duration_in_days_of_case > 80:
                  cases_to_remove.append(case_name)

      print(cases_to_remove)
      # Filter out cases that meet the criteria
      filtered_log = filtered_log[~filtered_log['case:concept:name'].isin(cases_to_remove)]
```

Figure 5 - Code responsible for removing cases where stayed over 80 days in the hospital

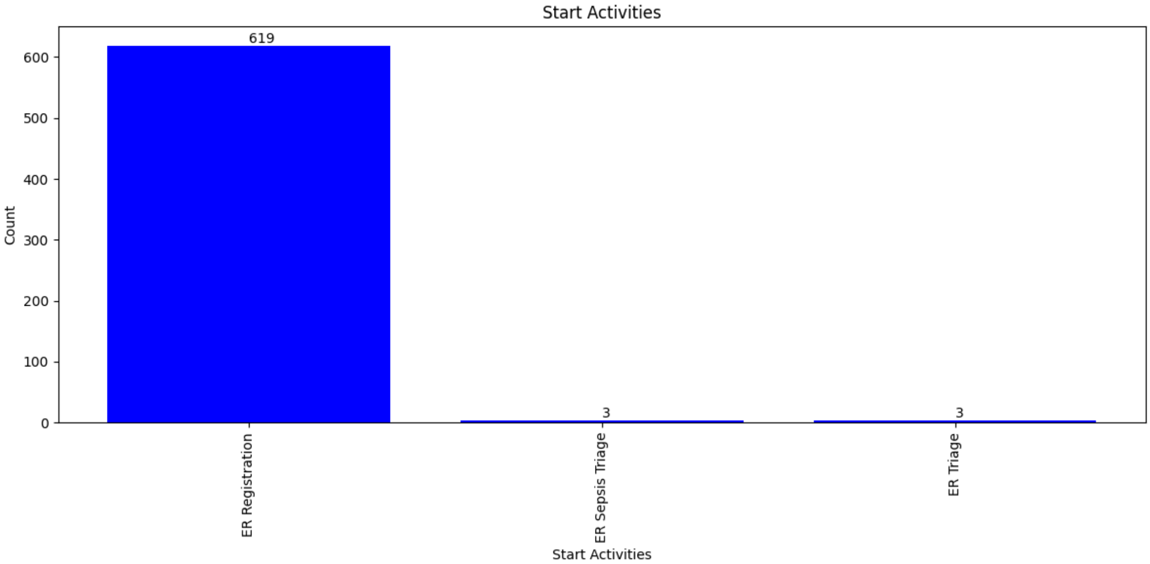


Figure 6 - Distribution of filtered start activities

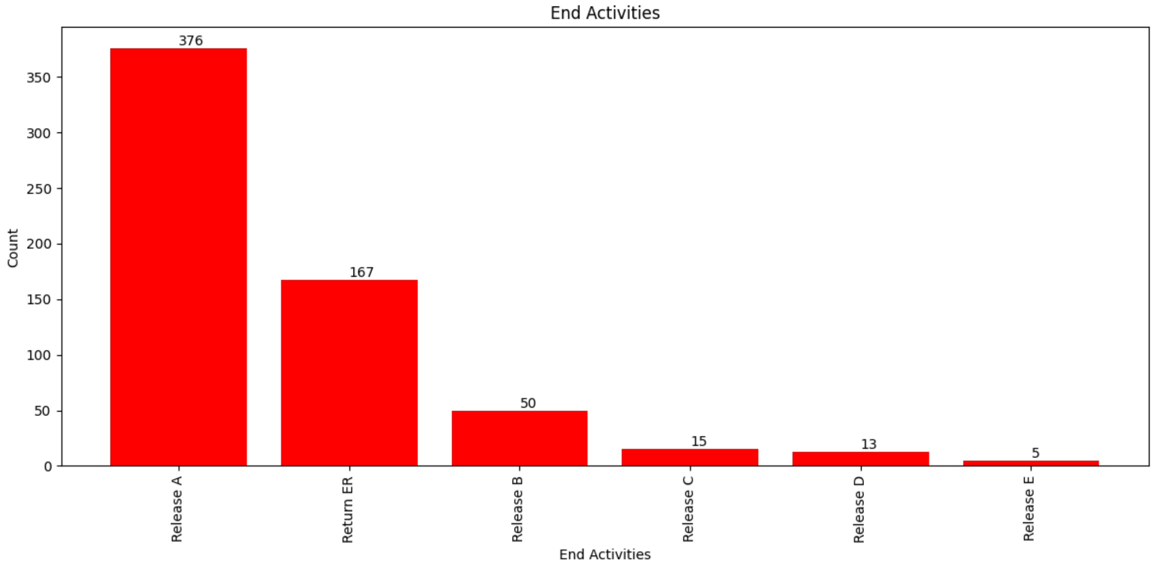


Figure 7 - Distribution of filtered end activities

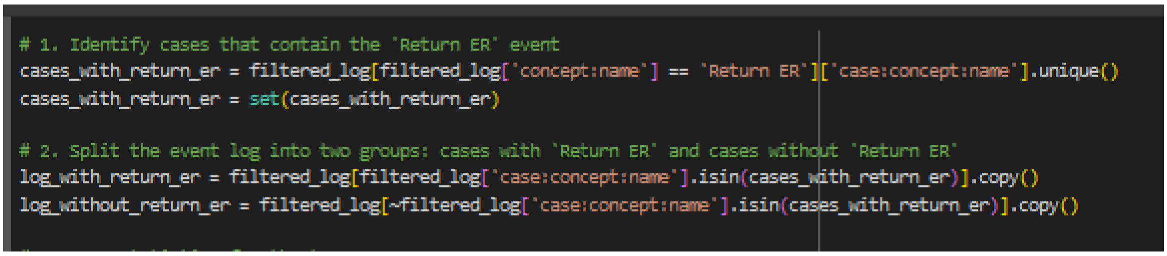
Given 1050 total cases in the log, we have 625 cases that comply with constraints for complete cases and for 15214 total events in the log, we have 10437 events that comply.

2) Descriptive analysis

Some patients come back after the initial admission, while some do not. It would be beneficial from both a financial and administrative point of view to limit the number of required returns to hospital. What I will try to examine is if there is any difference between patients that come back to the hospital (have Return ER events) and those that do not. To determine whether there is any pattern I will try to examine such aspects of each dataset as:

- event transition
- duration of treatment and probability of returning to the hospital
- frequency of important tests - CRP, Leukocytes, LacticAcid
- age distribution
- diagnose test results difference

All of those parameters will be examined on the already mentioned two sets acquired from the filtered log and divided on the basis of whether the “Return ER” event exists in the case. Therefore, I will start by dividing the set into two categories



```
# 1. Identify cases that contain the 'Return ER' event
cases_with_return_er = filtered_log[filtered_log['concept:name'] == 'Return ER']['case:concept:name'].unique()
cases_with_return_er = set(cases_with_return_er)

# 2. Split the event log into two groups: cases with 'Return ER' and cases without 'Return ER'
log_with_return_er = filtered_log[filtered_log['case:concept:name'].isin(cases_with_return_er)].copy()
log_without_return_er = filtered_log[~filtered_log['case:concept:name'].isin(cases_with_return_er)].copy()
```

Figure 8- dividing the log into two categories

We have two new sets of the log file - log_with_return_er and log_without_return_er. The parameters of the sets:

- with 'Return ER': - number of events: 3189 - number of cases: 167
- without 'Return ER': - number of events: 7248 - number of cases: 458

Heatmap of event transition

Let’s try to paint two heatmaps of events transition one for each of the logs.

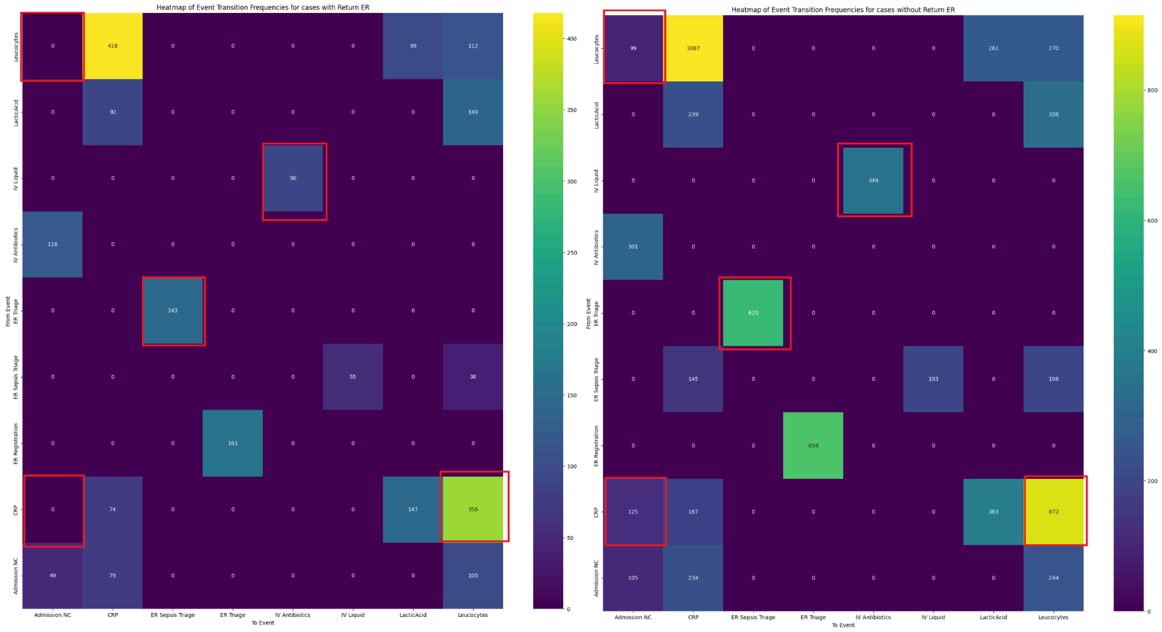


Figure 9 - Heatmap of event transition (left - with Return ER, right - without)

Event transition heatmap has been adjusted, by:

- scaling of threshold based on the number of cases
- removing columns related to specific group (Release* or Return ER)
- threshold removing other insignificant columns (for ex. types of releases of admission IC, due to small sample)
- color scale has been adjusted

There are no significant differences between log heatmaps which suggest that the process is rather similar (same colors and transition types). However, the areas with red squares have some differences:

- The heatmap with Return ER on the left has no transitions from ER Sepsis Triage to CRP. On the other hand the right heatmap (without Return ER) has almost 150 such transitions.

- On a heatmap of non-return cases (right) we may see that there is notably more transition from Admission to Triage and from Triage to Sepsis Triage, which may indicate some more procedural organized manner of the treatment process. This slight difference might also be spotted on the Petri Net (in the process mining part) and it should be asked why there is such change between the two processes. It might be the case that due to omission of some procedure steps the patients are not fully diagnosed.

The rest of the differences are not that statistically significant, however the business process questions might be asked, why there are slight differences between process of these two groups of patients

Duration of treatment and probability of returning to the hospital

Now let’s try to examine the distribution of duration of treatment with division for those cases (already mentioned groups) where the patient returns (Return ER) and those where he doesn’t.

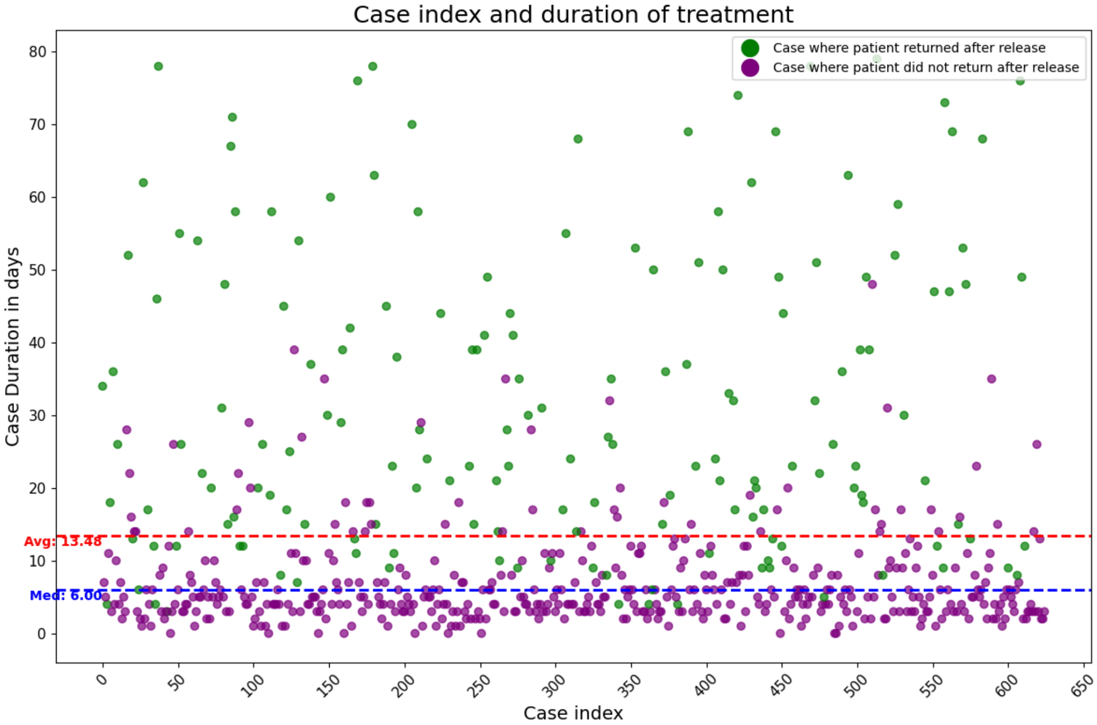


Figure 10 - Duration of days spent in hospital by case index for both patients that return and those that do not

As we can see there is a significant difference between the distribution of time spent in the hospital for both of these groups. Patients who come back have had their stay at the hospital longer than those who don't. The average duration of stay is 13.46 days and median is 6.00. Now let's break this plot into two separate plots and recalculate average and median.

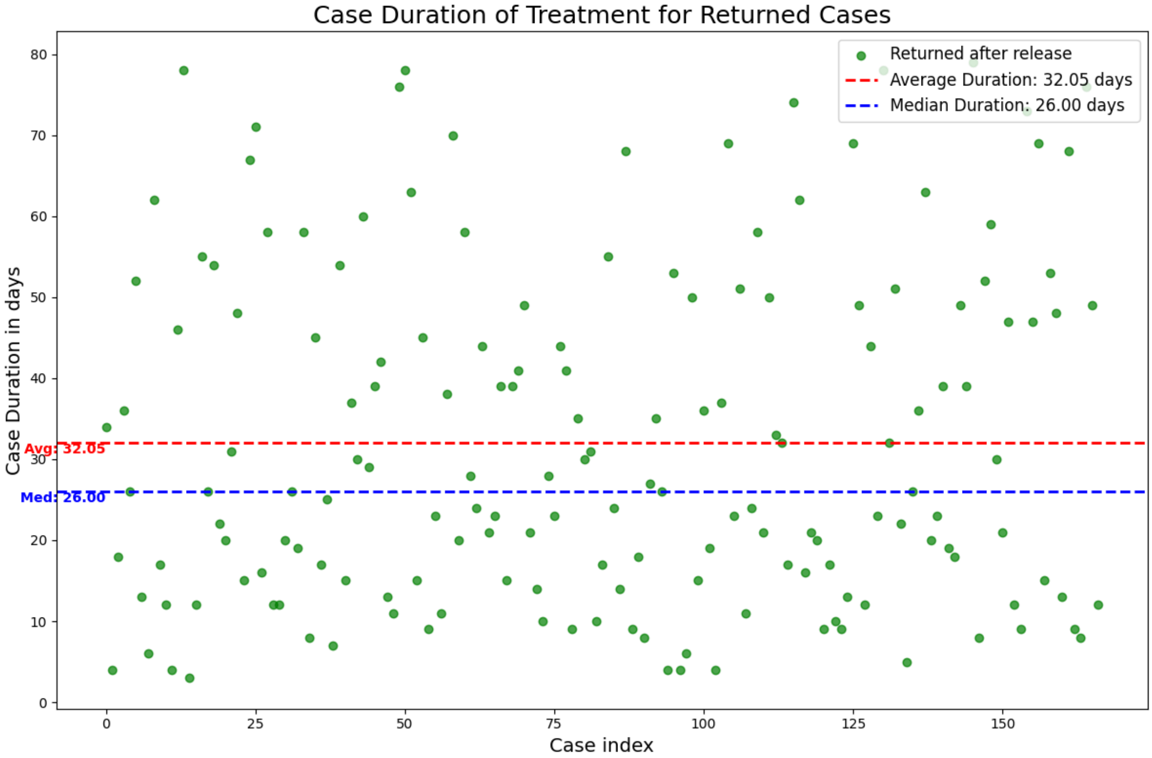


Figure 11 - Duration of stay for patients who return, distributed per case index

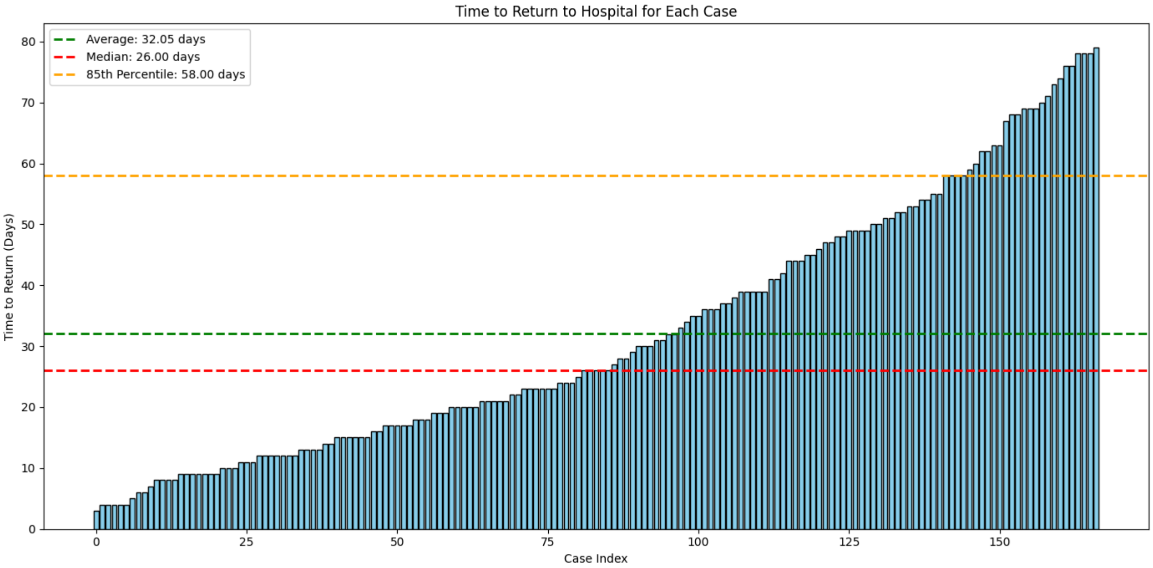


Figure 12 - Duration of stay for patients who return distributed per case index, different representation

We can see that patients that come back have the average duration of 32.05 days and a median of 26.00.

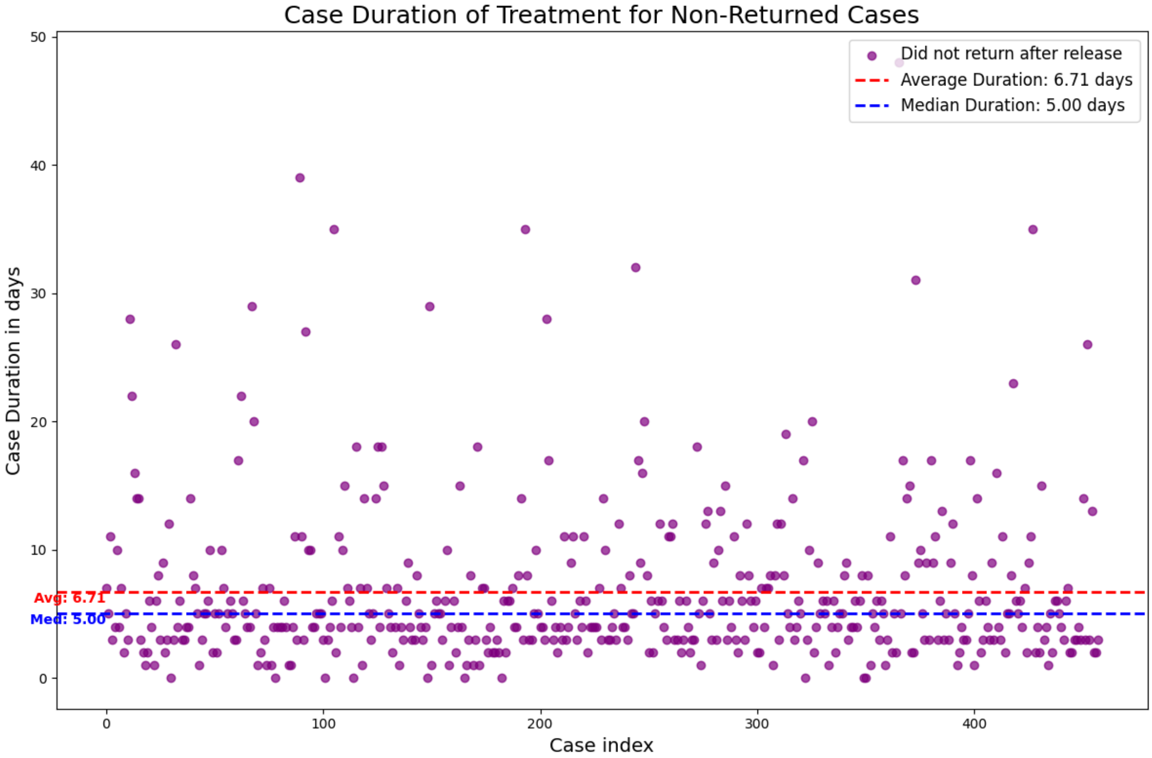


Figure 13 - Duration of stay for patients who don't return distributed per case index

In the case of patients who don't come back the average duration of stay at the hospital is 6.71 and median is 5.00.

	Patients who return	**Patients who don't return**	**Difference**
AVG duration of stay	32.05	6.71	25.34
Median duration of stay	26.00	5.00	21.0.0

It means that patients who return, spend on average 25.34 days more in the hospital than those who don't. Let's now examine what is the probability of returning to the hospital, considering duration intervals for the cases of patients who come back to the hospital.

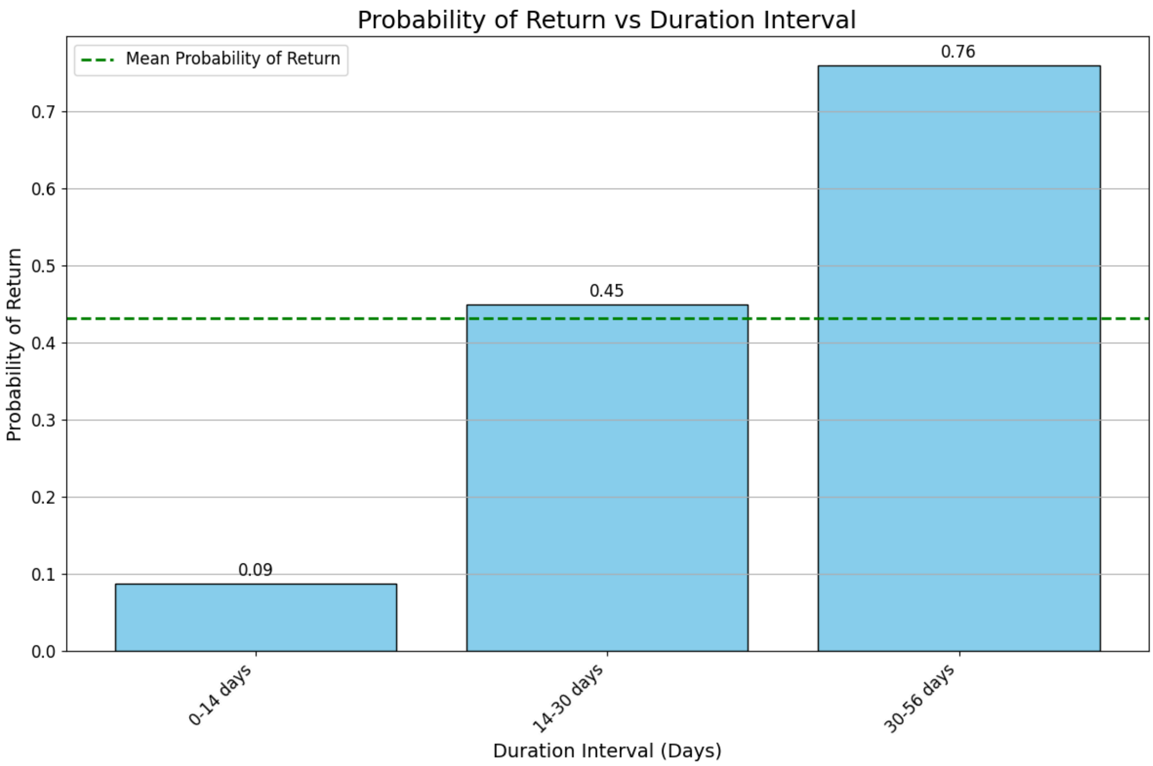


Figure 14 - probability of returning to the hospital, after spending days, divided to intervals.

We can see that the probability of return has an increasing pattern with every week of stay. In the first week the probability is close to 0.1. In the second two-week interval the probability is over 4 times bigger. And the final interval step (30-56 days) makes the probability 0.76 which is almost 1.7 times more compared to the previous interval.

Frequency of important tests - CRP, Leukocytes, LacticAcid

Let's examine how often there are activities in each case that are CRP, Leukocytes or LactidAcid examination and compare those numbers in the set of cases with “Return ER” and without. The plots have been normalized because of the different number of cases in each of the sets.

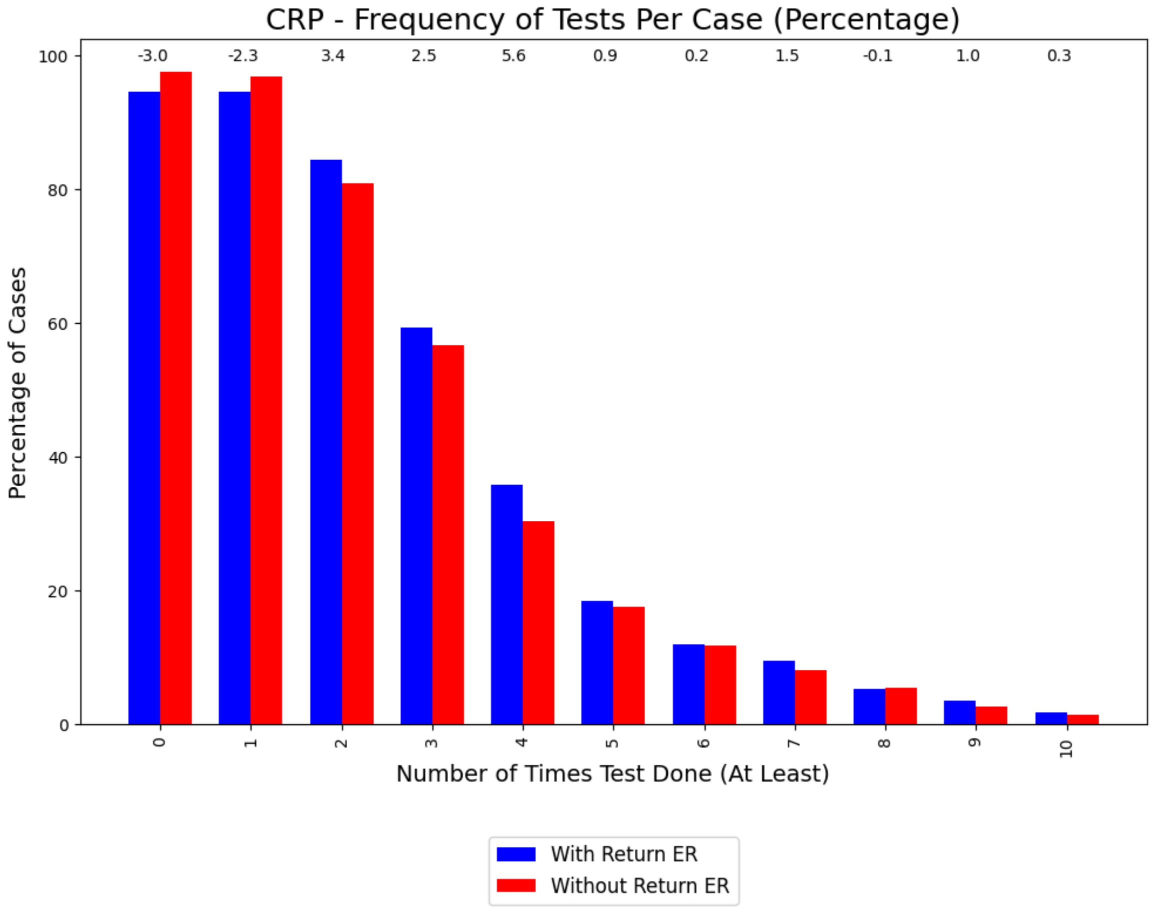


Figure 15 - Number of times (at least) CRP test was done in the percentage of cases

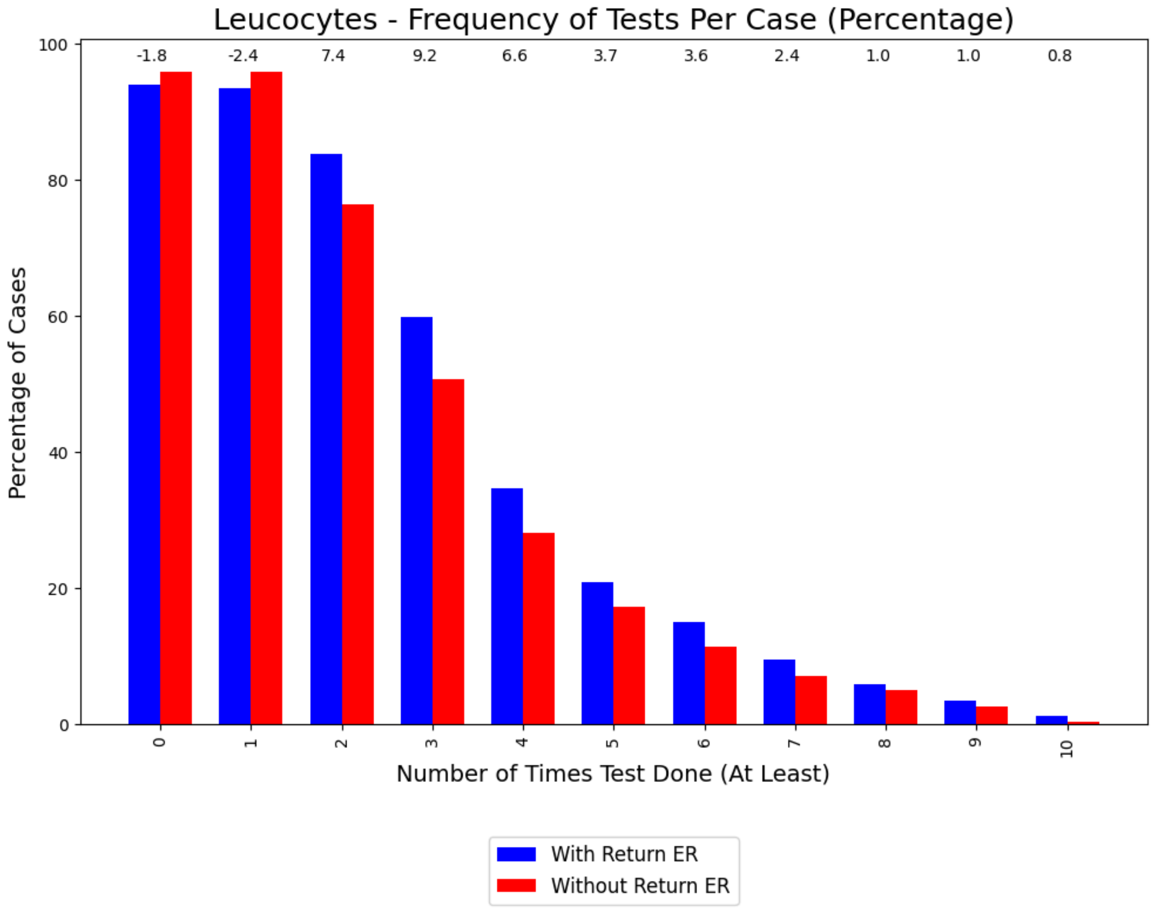


Figure 16 - Number of times (at least) Leukocytes test was done in the percentage of cases

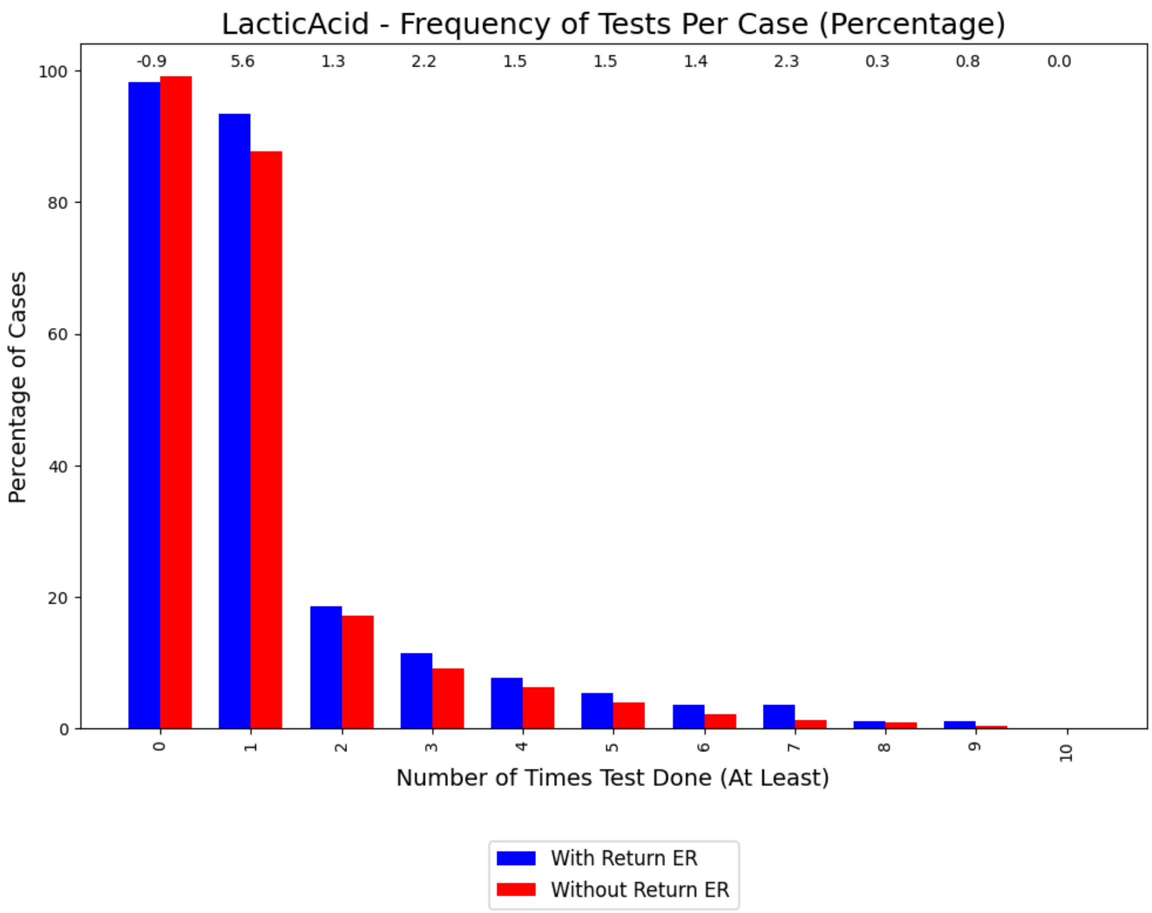


Figure 17 - Number of times (at least) LacticAcid test was done in the percentage of cases

We might see some slight tendency there - that for each of the above activities there is a slight difference in the frequency of examination. It might be spotted on the plot that cases in which patients return to the emergency room tend to undergo the specified diagnostic tests (CRP, Leucocytes, Lactic Acid) more frequently after the second test compared to cases which do not return. However, the difference is not that statistically significant and saying that there is a visible pattern would be far fetched.

Age distribution

Now, let’s examine the age distribution for each of the sets.

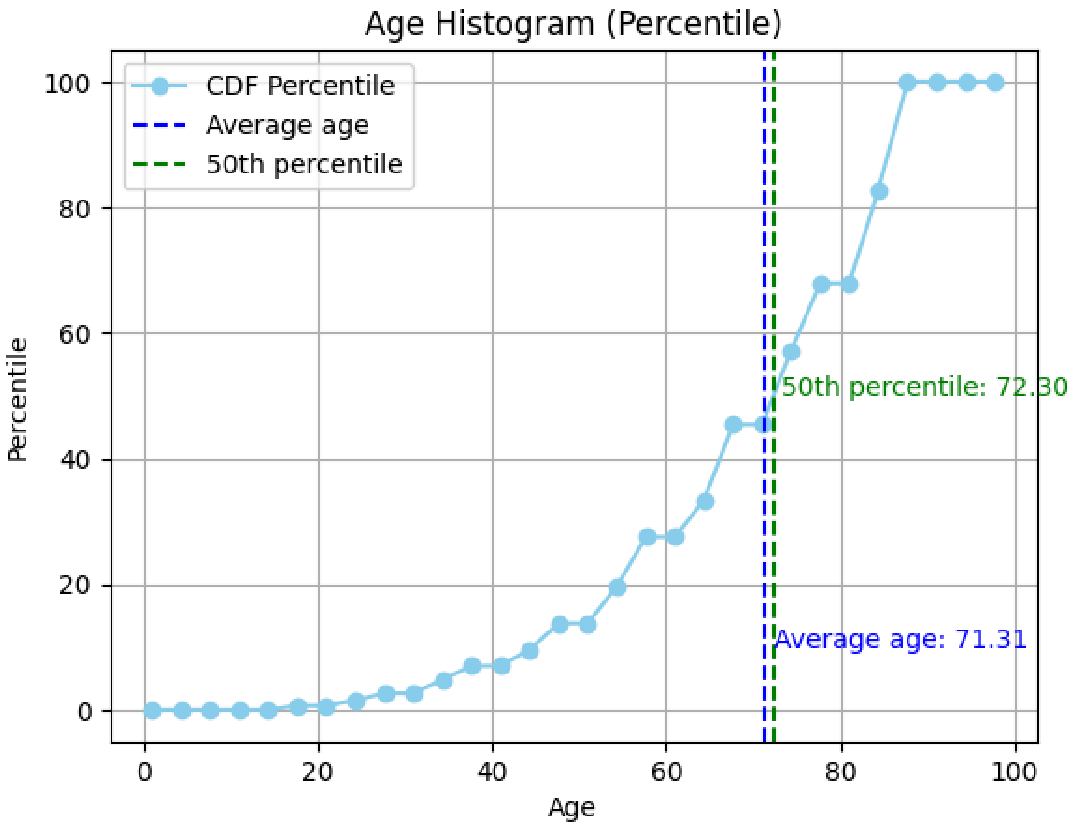


Figure 18 - Age distribution for patients that do not return (without return ER)

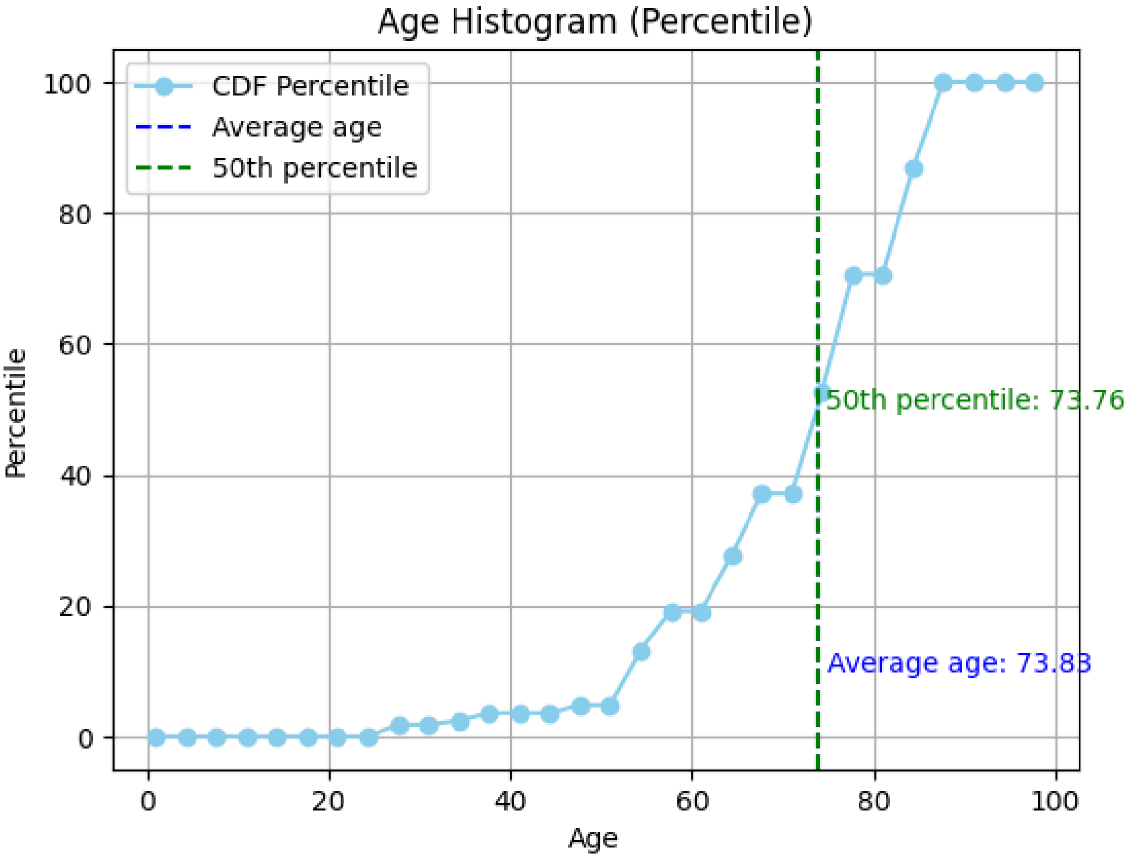


Figure 19 - Age distribution for patients that do return (with return ER)

The data printed on the plot indicate that patients who return to the ER tend to be, on average, 2.5 years older than those who do not return. This age difference suggests that older patients might be at a higher risk of returning to the hospital. Although the age difference is not that big between these two sets, it should be considered that even 2,5 years of difference when talking about the age of between 70 to 80 might mean a big difference in a patient's health conditions.

Serious sepsis case is defined as in the exercise:

- Leukocytes count less than 4,000/µl or more than 12,000/µl.
- LacticAcid levels > 2.0 µg/L
- CRP more than 50 mg/l

We can try to examine the age distribution for the two sets and try to add sepsis severity for those cases.

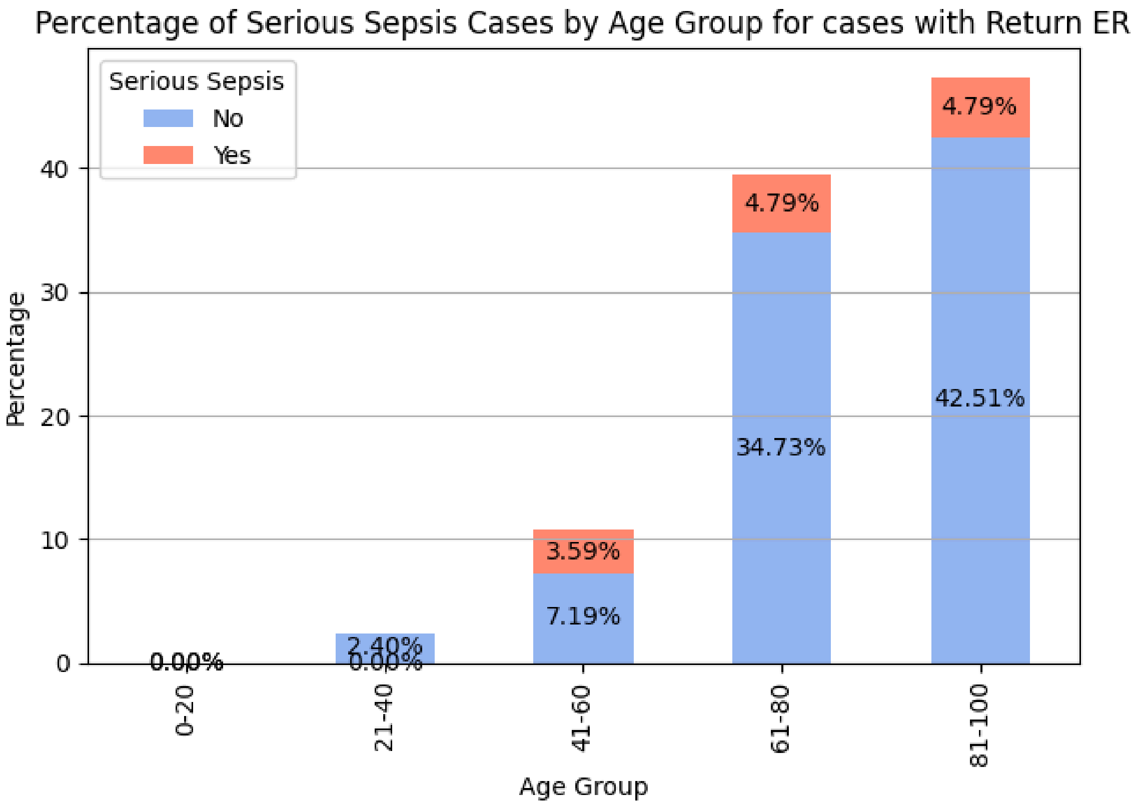


Figure 20 - Age distribution for cases with Return ER with serious/not serious sepsis division

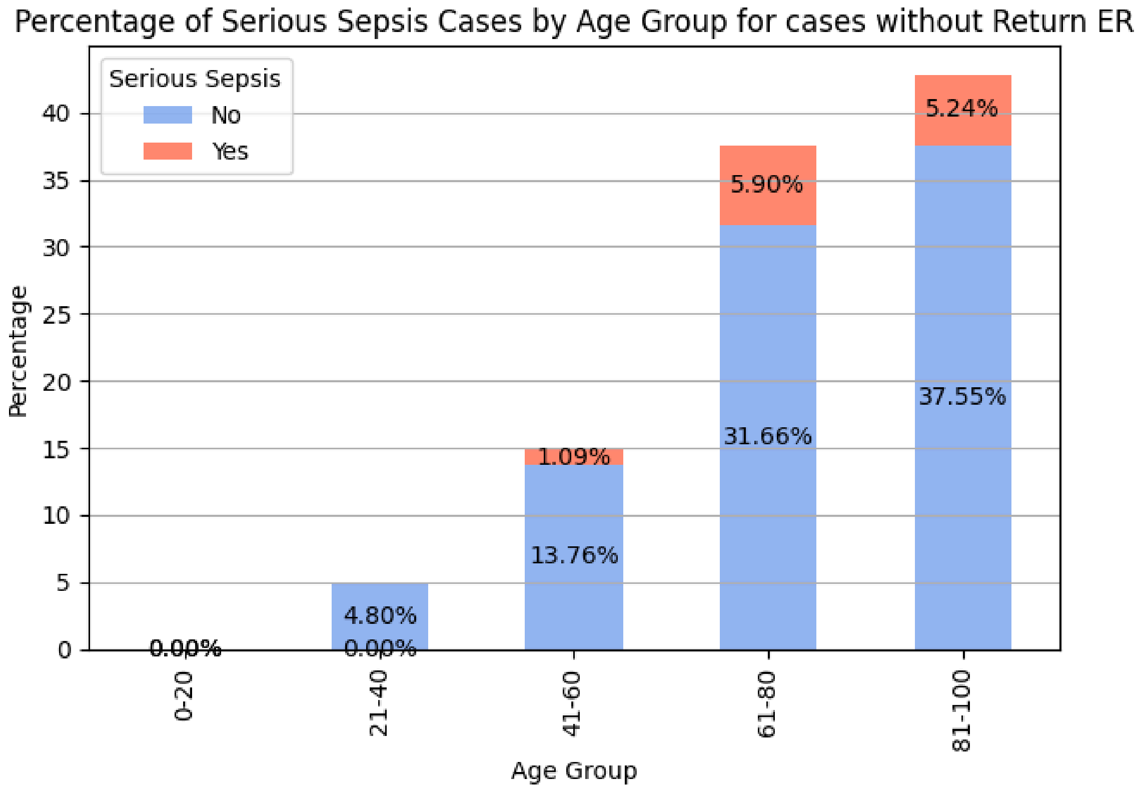


Figure 21 - Age distribution for cases without Return ER with serious/not serious sepsis division

There is no statistically significant severity distribution.

Diagnose test results difference

Let's examine the different parameters of the event log that are located in many columns as part of ER Registration activity. These columns might inform us about the first diagnosis of the patient during registration or whether some examination has been done.

The parameters have been plotted and counted for each of the sets (those patients who return and those who don't). There were no significant differences of diagnosis observed except for 1 - SIRSCritTahypnea which is a parameter for diagnosis of rapid, abnormal breathing. Patients who returned later to the hospital met 13.01% more often the SIRSCritTahypnea. Tachypnea is a significant indicator of potential systemic inflammatory responses, such as sepsis.

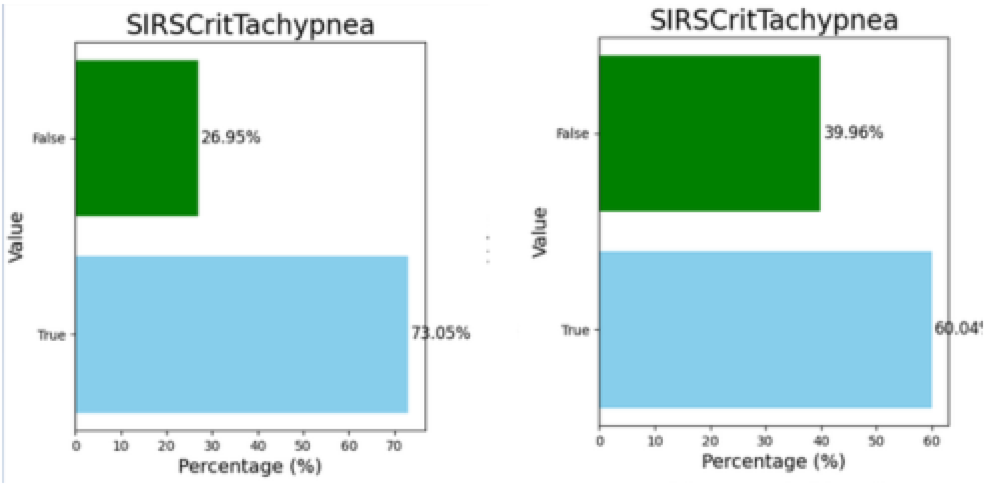


Figure 22 - Tachypnea parameter - left for the patients who return, right for those who don't

3) Process Mining

To discover the business process behind sepsis handling I will use different process discovery algorithms like Petri nets, Alpha Miner, Heuristics Miner, and Inductive Miner. I will also try to plot business processes separately for each of the sets (Return ER vs non-return).

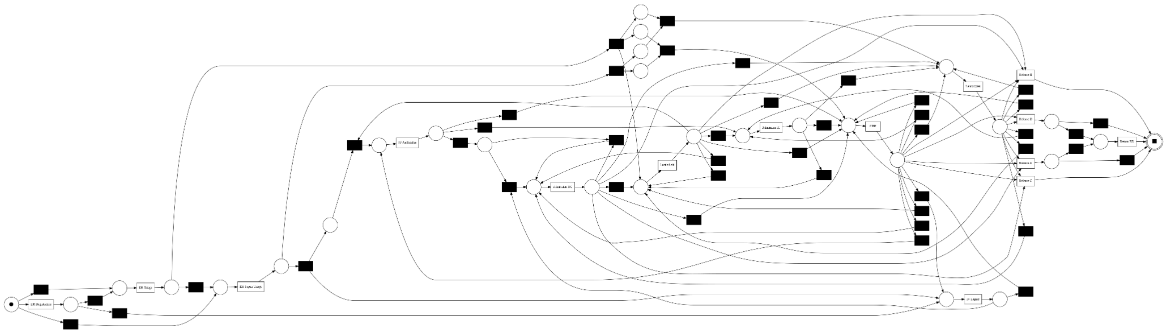


Figure 23 - Heuristic miner applied on the event log

We can try to plot the business process with petri net technique with noise_threshold=0.8.

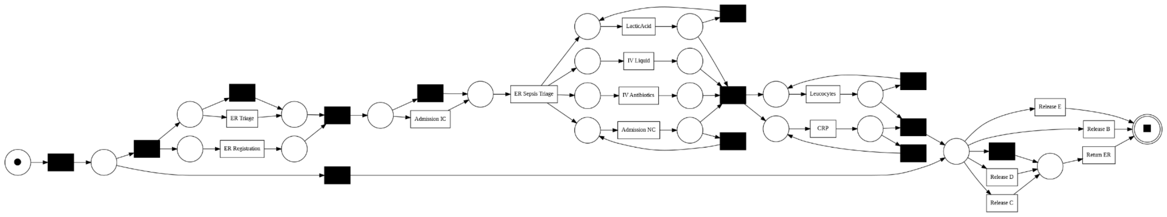


Figure 24 - Petri net applied on the event log

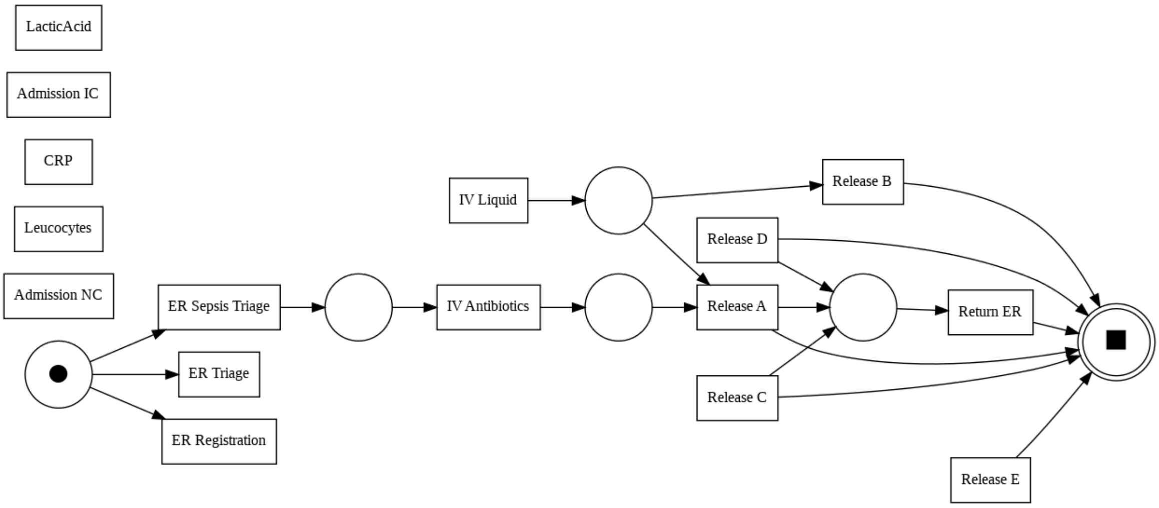


Figure 25 - Alpha miner applied on the event log

Let’s try to draw the petri net for these two sets to examine the differences.

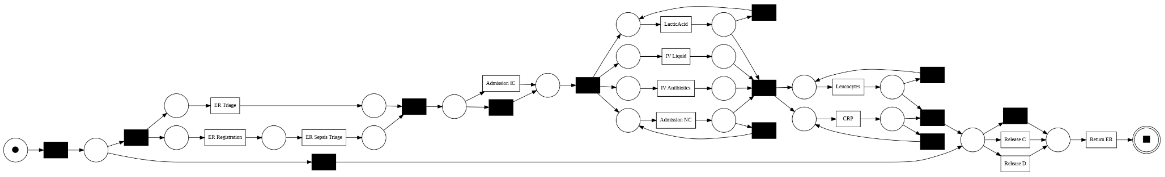


Figure 26 - Petri net applied on the set of cases with Return ER

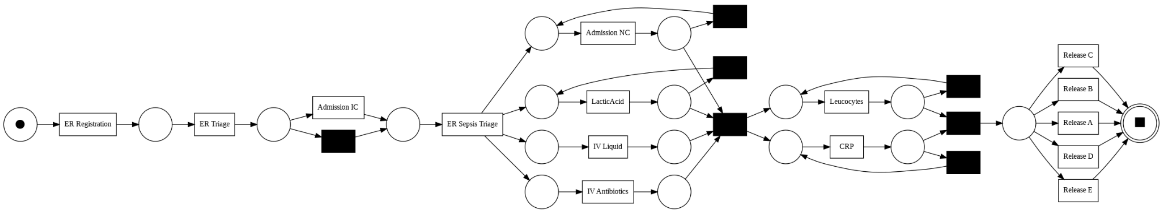
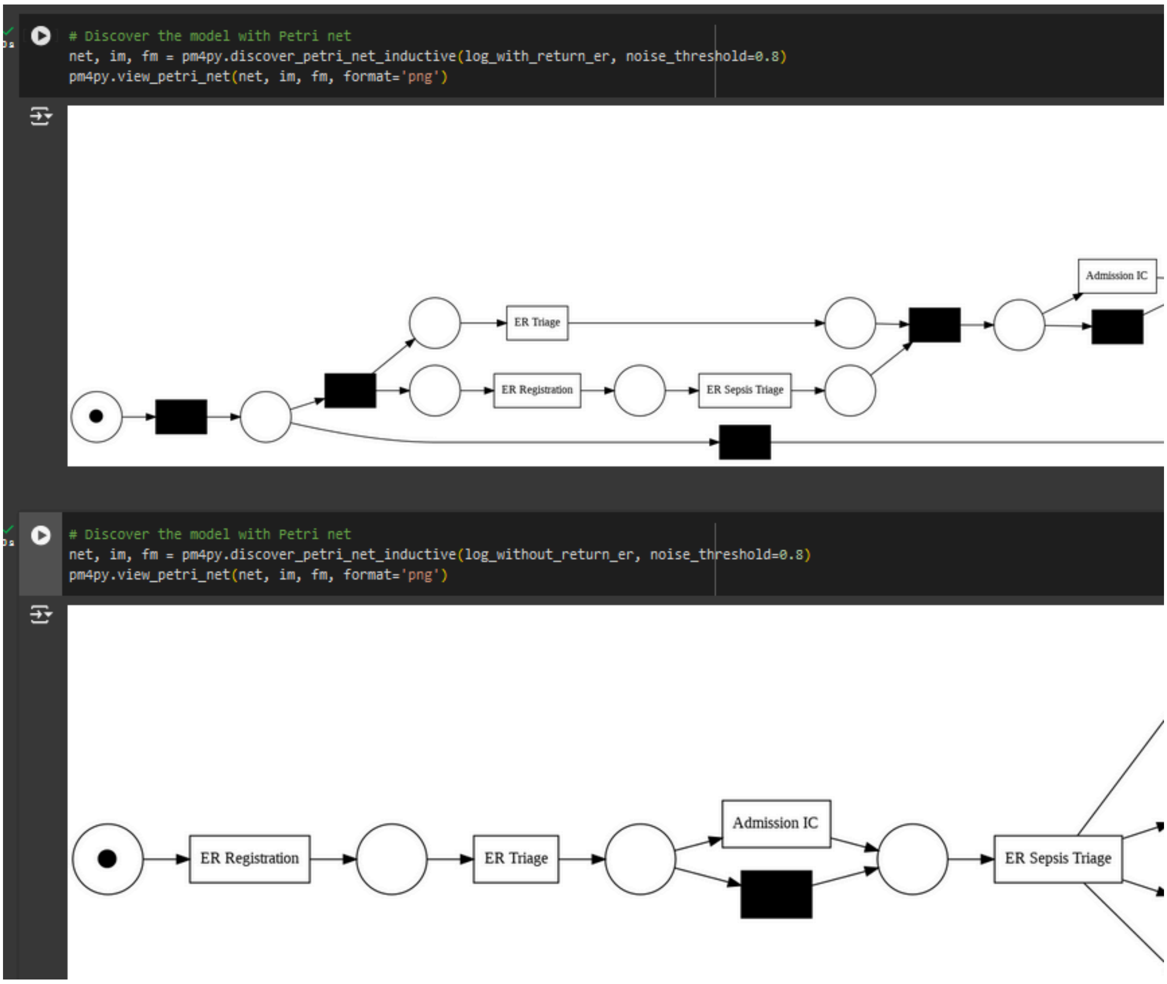


Figure 27 - Petri net applied on the set of cases without Return ER

Let’s compare the above process. We might spot that in the beginning and in the end of the processes there are significant differences.



```
# Discover the model with Petri net
net, im, fm = pm4py.discover_petri_net_inductive(log_with_return_er, noise_threshold=0.8)
pm4py.view_petri_net(net, im, fm, format='png')
```

```
# Discover the model with Petri net
net, im, fm = pm4py.discover_petri_net_inductive(log_without_return_er, noise_threshold=0.8)
pm4py.view_petri_net(net, im, fm, format='png')
```

Figure 28 - Petri net comparison of logs with and without return - beginning

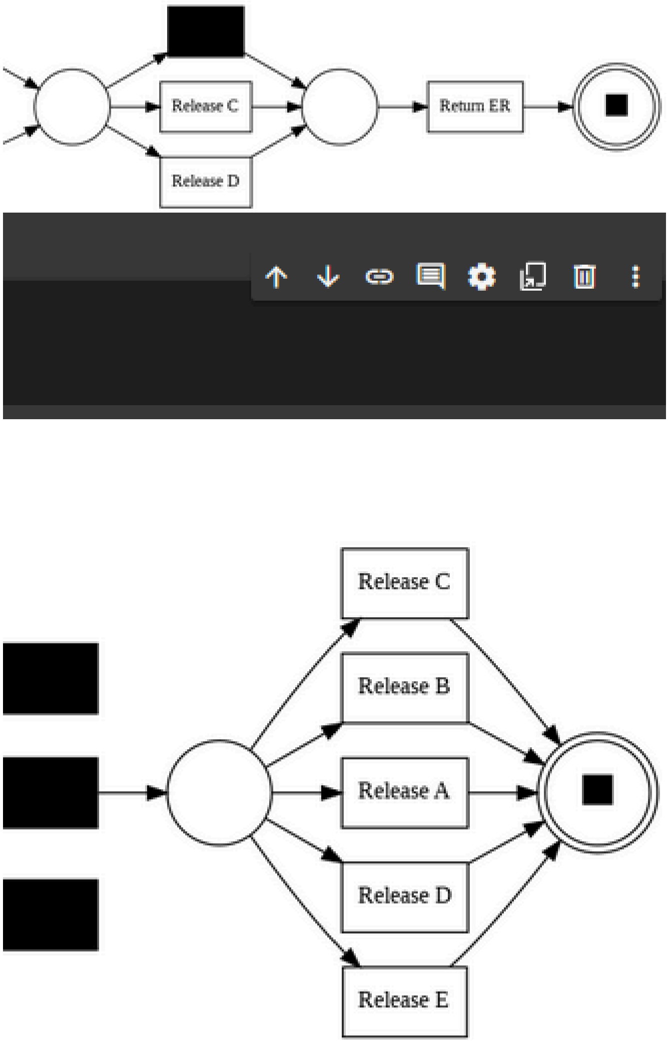


Figure 29 - Petri net comparison of logs with and without return - end

What might be spotted is that:

- There is a difference in the beginning - patients that don’t come back, go through the triage process in a very orderly, repetitive manner (ER Registration -> ER Triage -> Admission IC -> Sepsis Triage). On the other hand those patients who come back (Return ER) might go directly to one of these: ER Registration, ER Triage. Sepsis is a condition which cannot be easily diagnosed without careful patient examination and advanced diagnosis methods. We might have the hypothesis that due to the bad classification of patients and improper beginning of the process and omitting procedure steps, some people are misdiagnosed and sent for further sepsis examination. Then they are released due to lack of sepsis detection and they return because they are not healthy.

- For some reason there is no Release type A, B, E in the process discovered for log with return. On the other hand in the log without return there are all types of releases. For some reason release type A and B might make patients unlikely to come back. Unfortunately A-E classification is not a standardized medical classification universally recognized, therefore we don’t have any further information on what it means.

4) Conformance checking

The conformance checking applied for the created models is token-based replay diagnostics. he models were assessed based on four key criteria:

- **fitness** - how well the cases from the log align with the model, comparing the model's behavior to the behavior observed in the log.
- **precision** - accuracy of the model, how accurately is the behavior represented
- **generalization** - means how well the model can generalize the behavior observed in the log
- **simplicity** - assesses the complexity of the created model. A simpler model is easier to understand and interpret

	Heuristic miner	**Petri net inductive**	**Alpha miner**
Fitness	0,90	0,67	0,73
Precision	0,88	0,57	0,31
Generalization	0,84	0,87	0,90
Simplicity	0,49	0,62	1.0

Figure 30 - Comparison of 3 models characteristics for the filtered log

	Heuristic miner	**Petri net inductive**	**Alpha miner**
Fitness	0,88	0,73	0,66
Precision	0,81	0,57	0,25
Generalization	0,79	0,85	0,84
Simplicity	0.53	0,67	1.0

Figure 31 - Comparison of 3 models characteristics for the log_with_return_er

	Heuristic miner	**Petri net inductive**	**Alpha miner**
Fitness	0,91	0,83	0,80
Precision	0,70	0,72	0,337
Generalization	0,84	0,90	0,88
Simplicity	0,47	0,62	1,0

Figure 32 - Comparison of 3 models characteristics for the log_without_return_er

5) Intervention strategies

Basing on the descriptive analysis and process mining steps, some main observations may be formulated:

- Patients staying longer after ER admission are generally more likely to experience return to the ER
- Despite the above, they do not undergo more test on average
- Process discovered for cases of patients who do not return to the ER is more clear and organized at the beginning than that of returning patients
- Critical tachypnea is present more often with patients who later need to return to the ER.

Based on the above formulations, three main intervention strategies may be determined on order to reduce the percentage of patients' returns:

- During the stay on hospital during sepsis treatment, important blood tests should be conducted with more regular frequency (so that patients staying for a longer period of time should experience proportionally more tests) - this would allow more up-to-date patients' health monitoring and could prevent releasing the patient whose state is only temporarily better, but examination results are still not right (possibly indicating an ongoing infection)

- If a patient is released after longer than a fixed period of time (ex. 30 days, based on Figure 14) they should always be referred for a check-up examination after a week or two. This could show any underlying infection in an early stage before it spreads, resulting in the need to come back
- Some check-up tests (as above) should also be ordered for patients displaying symptoms of critical tachypnea
- A clear ER admittance procedure should be defined and strictly followed by medical personnel. This could result in limiting the number of misdiagnosed patients, incomplete medical records and omitting some important steps on treatment resulting later in the need for ER comeback.

Conclusions

After in-depth log analysis it can be stated that overall there is not much difference between cases of patients who need to return to the ER and those who don’t - it may be due to the medical nature of the data and the fact that sepsis has no one, easily defined course.

This finding suggests a need for standardizing reaction protocols as part of the operational objectives to ensure a consistent treatment process, potentially reducing the variability in patient outcomes.

The patient treatment could be improved by ensuring regular testing and follow-up examinations, especially with long stay durations. Moreover, reducing the number of patients returning to the hospital would minimize the cost of treatment.

Code:
https://colab.research.google.com/drive/1ayjk7Ok7g5e5ZbkeOts4jC11xHHFFLso?usp=sharing