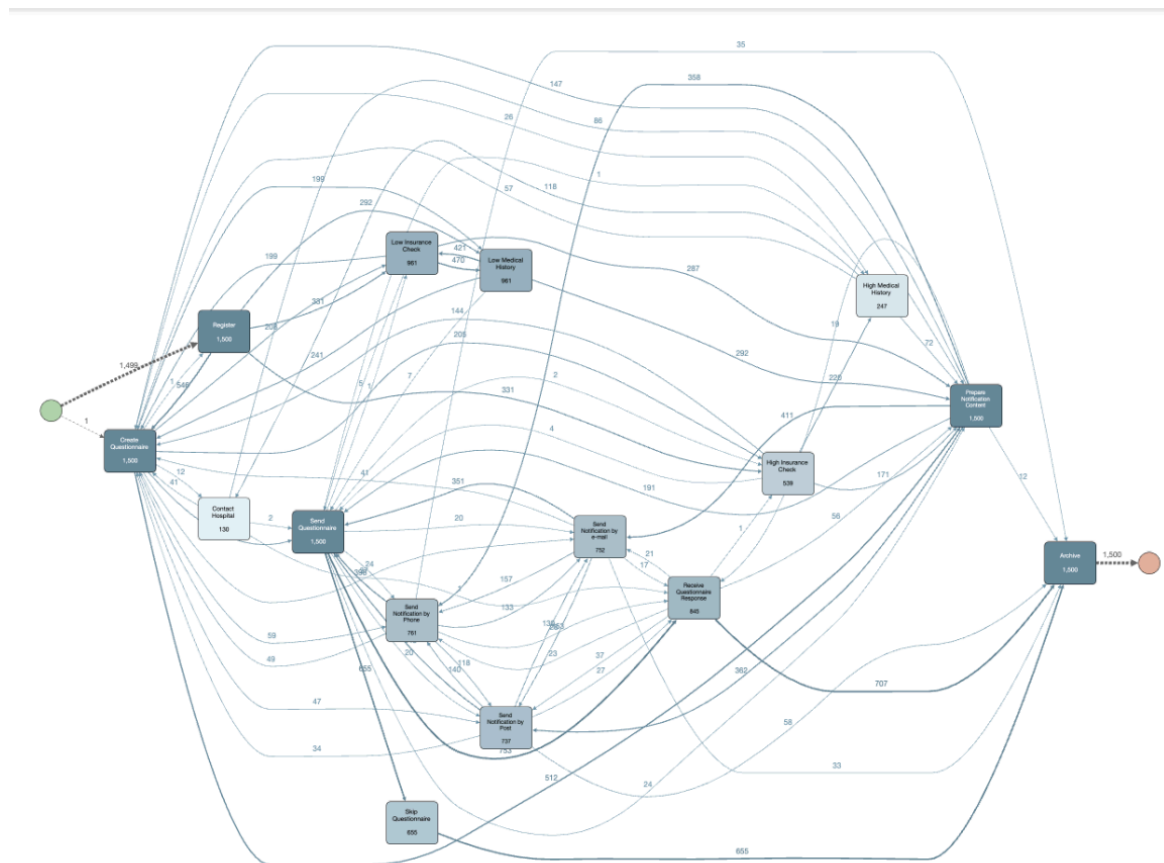


1. Automated process discovery.

**A. Use Apromore to discover three different Process Maps (a.k.a. directly follows graph). For getting different Process Maps, vary the abstraction sliders (for activity nodes and arcs) and find three interesting representations of the same event log.**

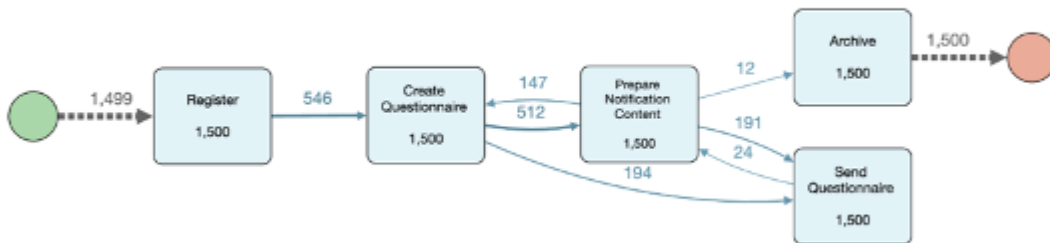
1. First, to view the overall process map, we fully understood the detailed process flow while visualizing all activities and relationships.



Perspective	Activity
Abstract by	Case frequency
Abstraction	low
Node	100

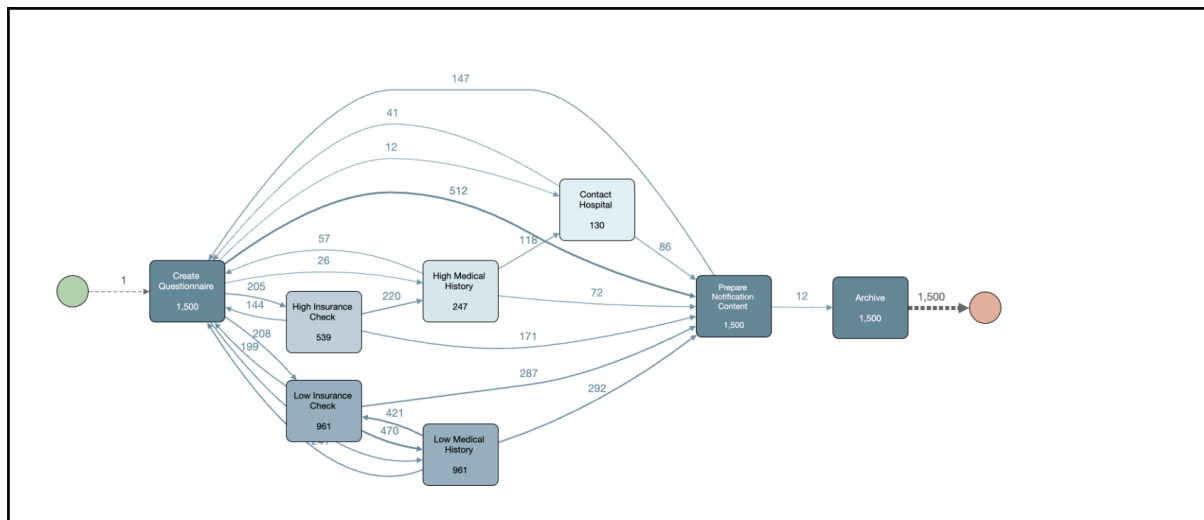
Arcs	100
------	-----

2. The second map, which highlights high-frequency activities, is an interesting point to focus on core processes, to identify critical flows and find efficient distribution of tasks.



Perspective	Activity
Abstract by	Case frequency
Abstraction	High
Node	5
Arcs	80

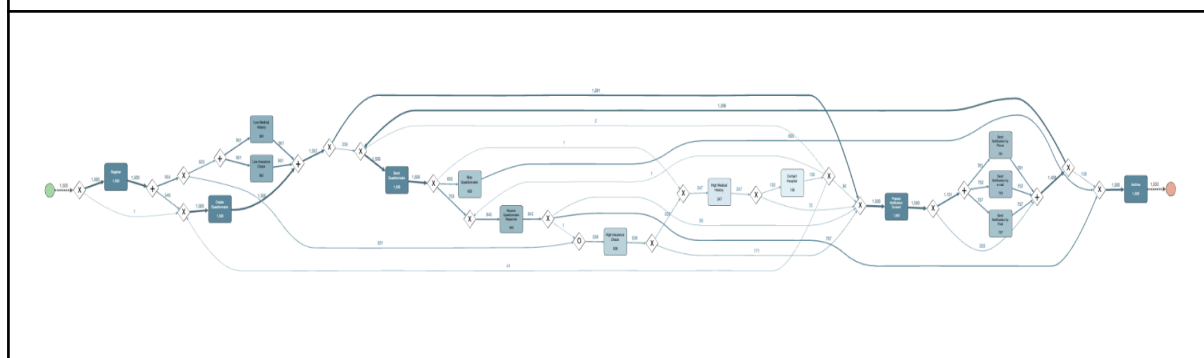
3. The third map, which highlights activities with long duration, visualizes inefficient segments, providing useful data for process optimization. This representation enables you to identify time-consuming segments and present tasks to improve speed or efficiency.



Perspective	Activity
Abstract by	Average duration
Abstraction	High
Node	20
Arcs	100

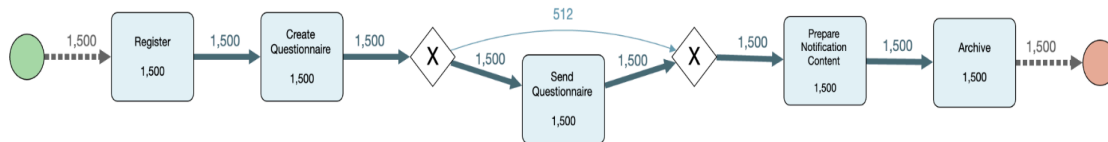
B. Using Apromore, discover a BPMN model.

1. Same as process map number 1



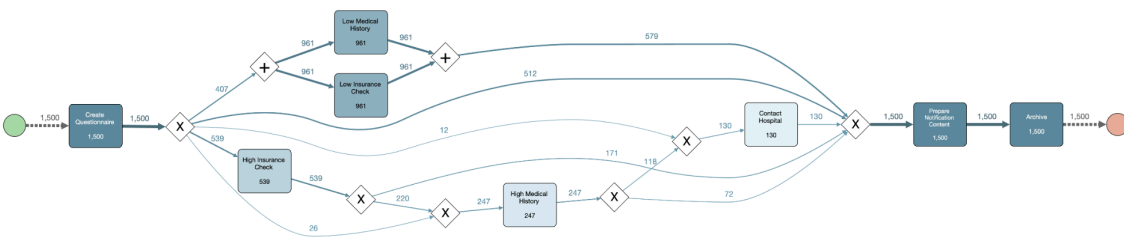
Perspective	Activity
Abstract by	Case frequency
Abstraction	low
Node	100
Arcs	100

2.Same as process map number 2



Perspective	Activity
Abstract by	Case frequency
Abstraction	High
Node	5
Arcs	80

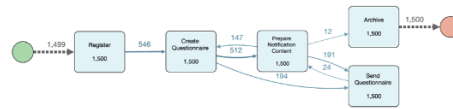
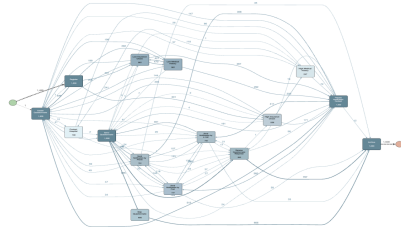
3.Same as process map number 3



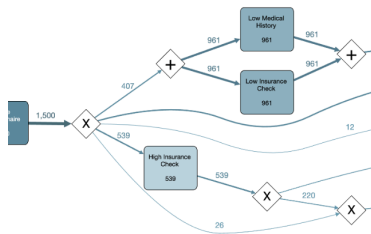
Perspective	Activity
Abstract by	Average duration
Abstraction	High
Node	20
Arcs	100

C. Compare the results of A and B.

## 1) Over-generalization



Because process maps simply display direct relationships between activities, comparing map 1 and map 2 can lead to over-generalization. This is problematic in that no path can necessarily exist.



Since BPMN models can clearly represent conditional or parallel flows, they are advantageous in preventing over-generalization problems. For example, conditional branches (XOR gateways) or parallel flows (AND gateways) can be used to clearly differentiate which paths are activated only under certain conditions.

Attributes	Process Map (Directly Follows Graph)	BPMN Model
<b>Complexity</b>	Simple and direct	Describes complex business processes
<b>Readability</b>	Simple and direct	Complex rules and flows, more difficult to understand
<b>Generalization</b>	Can include impossible paths. Only shows direct relationships between activities, so real-world processes may differ.	Defines complex flows and branches clearly to solve generalization issues.
<b>Condition Branching and Divergence</b>	Cannot express condition branching or exception handling. Only shows direct flows, so complex structures cannot be expressed.	Condition branching and exception handling can be defined using XOR gateways and AND gateways.

**Loop  
Definition**

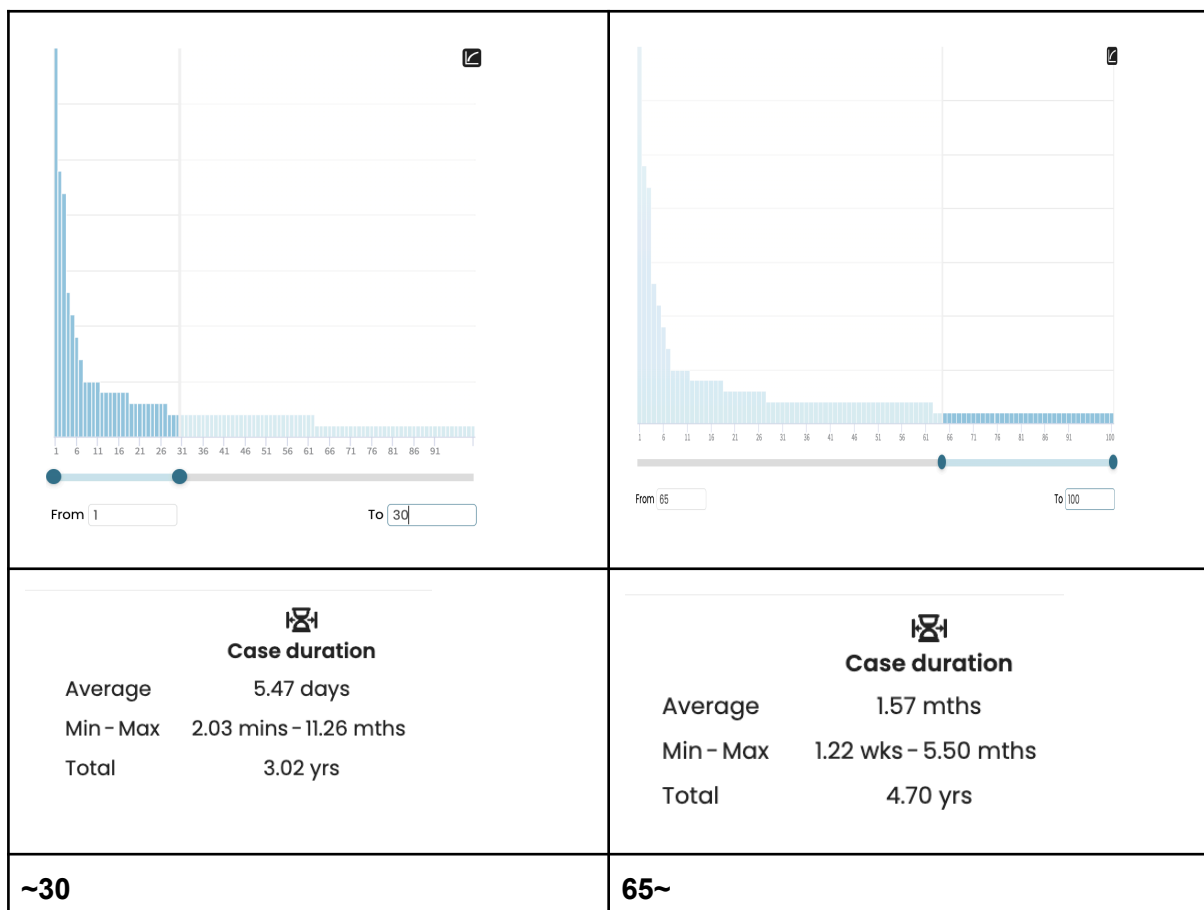
Difficult to define looping or repeated activities.

Loops and repeated activities can be explicitly defined.

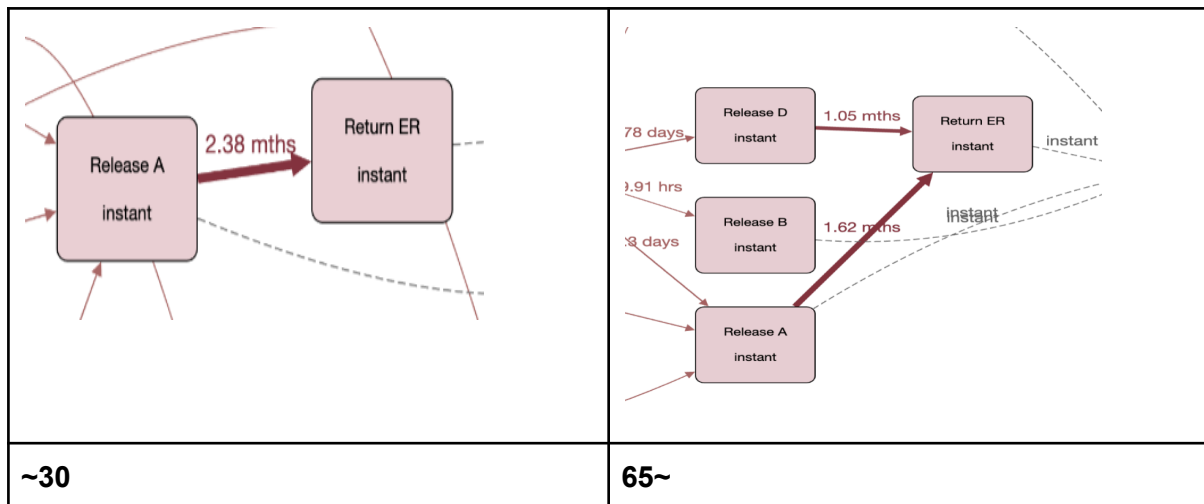
**While the process map provides a simple and intuitive flow, the BPMN model can clearly represent conditional flows, parallel processing, loops, etc. of complex processes, which is advantageous for dealing with the complexity of processes.**

**2. Variant Analysis. Consider the SEPSIS log that was used during our class. This is a log related to 1050 paKents treated for sepsis in a Dutch hospital. Generate two variants of this log, one where the paKents age is within 30 years (SEPSIS\_young), the other where the age is 65 or above (SEPSIS\_old). Based on these two variants, answer the following quesKons using a process mining tool (show proof of your answers with diagrams and/or screenshots of the tool of your preference):**

**A. What is the cycle time of each variant?**



**B. Where are the bottlenecks (highest waitKng Kmes) in each of the two variants and how do these bottlenecks differ?**

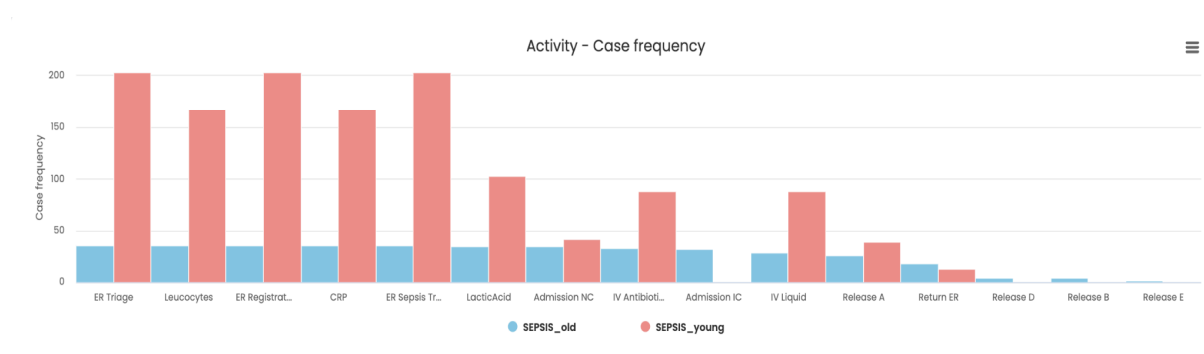


The wait time between Release A → Return ER is shown to be 2.38 months, which can be a bottleneck if this value is much longer than the wait time between other activities.

Wait times between Release D → Return ER are shown as 1.05 months, and wait times between Release A → Return ER are shown as 1.62 months, which may indicate a bottleneck.

C. Describe the differences between the frequency and order in which activities are executed in the two variants. Hint: If you use process maps, you should consider using the abstraction slider to hide some of the most infrequent arcs so as to make the maps more readable.

### 1) frequency

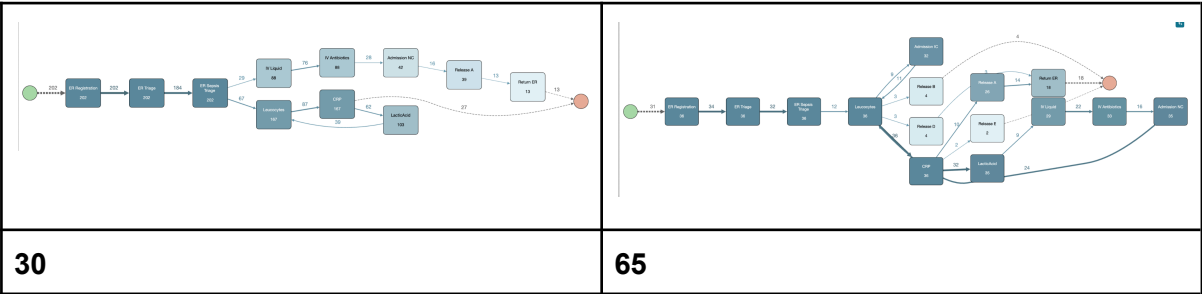


In SEPSIS\_young, activities such as Leucocytes, CRP, ER Sepsis Treatment, and Lactic Acid are more frequent, and overall, there are activities that are more frequent in SEPSIS\_young variants.

On the other hand, in SEPSIS\_old, Admission NC, Release A, Release

D, Release B, Release E, etc. tend to occur more frequently.

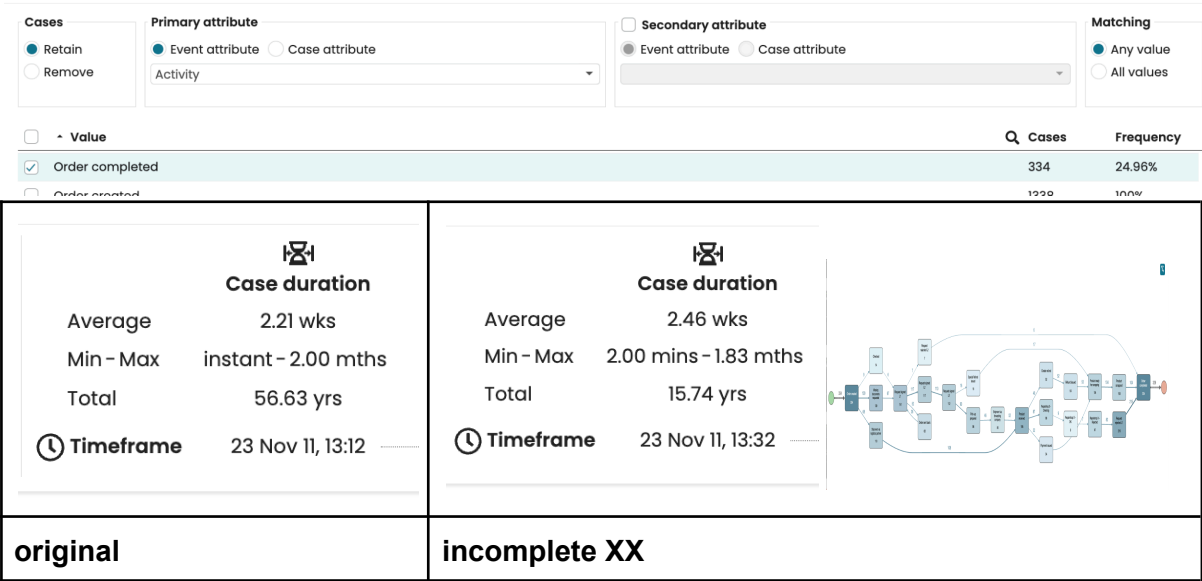
2) 순서비교



In the sequence of activities, CRP and Lactic Acid are ordered in reverse in the two variants. Also, SEPSIS\_young is followed by Return ER after Release A, and SEPSIS\_old is followed by Return ER after Release A

3. Performance and Compliance Analysis. Consider the dataset of a refund process from an electronics manufacturer. Customer complaints and the inspection of individual cases indicate that this process suffers from inefficiencies and overly long cycle times. Assume that only cases that have reached the 'Order completed' event are finished. Using the RefundProcess event log, answer the following questions using a process mining tool (show proof of your answers with diagrams and/or screenshots):

A. Is it a problem if you take the average cycle time of all cases, also the ones that have not finished yet?



Because unfinished cases are not fully processed, their cycle times are not fully measured. If unfinished cases are included, the cycle



time may be longer than the actually finished process. In this case, the average cycle time tends to be distorted and stretched. --> Therefore, analyzing only completed cases is advantageous for accurately identifying the average cycle time.

**B. In general, which channel(s) have the biggest problems with missing documents that need to be requested from the customer?**

**Cases**  
☒ Retain  
☐ Remove

**Primary attribute**  
☒ Event attribute ☐ Case attribute  
 Activity

☒ **Secondary attribute**  
☒ Event attribute ☐ Case attribute  
 Channel

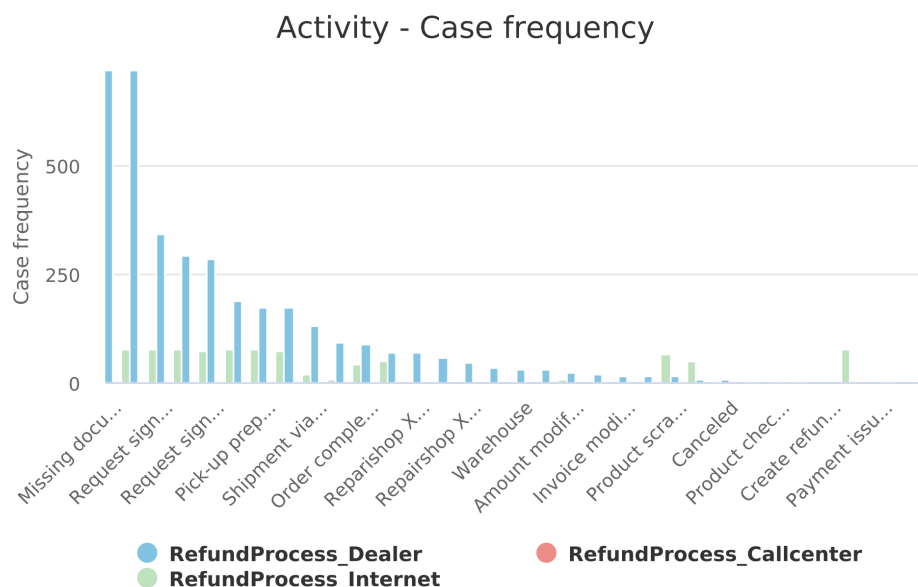
**Matching**  
☒ Any value  
☐ All values

Value	Cases	Frequency
<input type="radio"/> Invoice requested	40	2.99%
<input checked="" type="radio"/> Missing documents requested	799	59.72%
<input type="radio"/> Order completed	334	24.96%
<input type="radio"/> Order created	1338	100%

☐ Value

<input type="checkbox"/> Callcenter
<input type="checkbox"/> Dealer
<input type="checkbox"/> Internet

Missing documents requested' event was set as the main attribute and Channel was set as the secondary attribute, and the frequency of document requests for each channel was analyzed.



- **RefundProcess\_Dealer (blue):** This is the channel with the highest number of document requests. Document requests are occurring on this channel at a very high frequency. As a result, the issue of missing documents can be determined to occur

frequently on the Dealer channel.

- **RefundProcess\_Callcenter (red):**The frequency of document requests is relatively low; however, Callcenter also experiences document requests at regular intervals, and you also need to consider the issue of missing documents on this channel.
- **RefundProcess\_Internet (green):**The frequency of paper requests is lower than that of Dealers and Callcenter. This is one channel that receives fewer paper requests, but it is still one of them.

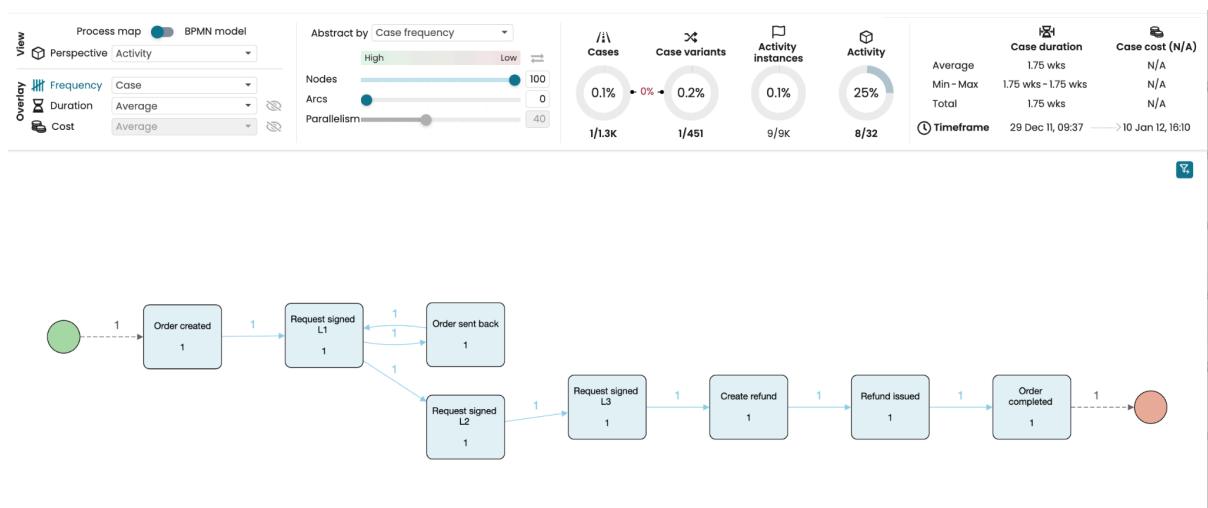
**C. How many customers have received a refund without the product being received by the electronics manufacturing company? This should not happen.**

### Criteria

Retain all cases that contain 'Activity' ['Refund issued']

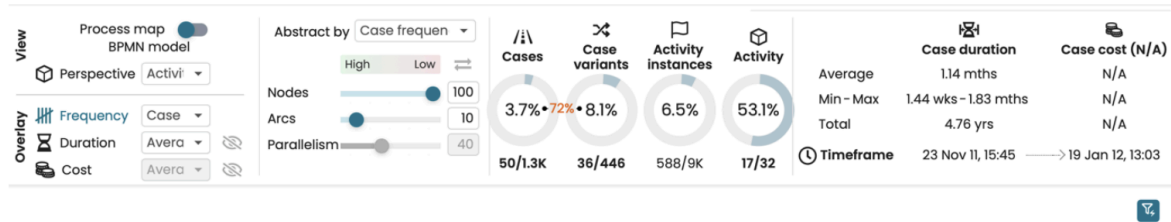
Remove all cases that contain 'Activity' ['Product received']

**Extract the case where the 'Product received' event did not occur after the 'Refund issued' event occurred.**



Only one person in that process map appears to have a problem between the 'Refund issued' event and the 'Product received' event. As a result, it is identified as a case in which one customer received a refund without receiving the product.

**D. Has a customer ever received a double payment? This should not happen in this process.**



**This can be confirmed through the "Issued Payment" event.**

**-->There have been 2 cases where the customer has received double payments.**