

A Process-Oriented Approach to Analyze Analysts' Use of Visualizations: Revealing Insights into the What, When, and How

– Supplementary Material –

Authors: Lisa Zimmermann, Francesca Zerbato, Katerina Vrotsou, Barbara Weber

Last updated: 31.03.2025

This document provides supplementary material to the paper mentioned above.

Section 4: Data Collection (Process Mining Study)

1)	<i><u>Participant Details</u></i>	<i><u>p.2</u></i>
2)	<i><u>Process Mining Analysis Sessions: Encoded Perspectives</u></i>	<i><u>p.3</u></i>
a.	<i><u>Visualization Perspective – Coding Scheme</u></i>	<i><u>p.4</u></i>
b.	<i><u>Focus Perspective – Coding Scheme</u></i>	<i><u>p.9</u></i>
c.	<i><u>Intent Perspective – Coding Scheme</u></i>	<i><u>p.10</u></i>
3)	<i><u>Description of the Event Sequence Data</u></i>	<i><u>p.11</u></i>
	<i><u>References</u></i>	<i><u>p.12</u></i>

Participant Details

The table contains the overview of the 33 analysts (**A**) who participated in our study. It highlights the **sector** to which the analysts' affiliation/company belongs, their job **role** at the time of the data collection, their self-rated **expertise** in process mining (ranked on a Likert-scale with values "novice", "basic", "average", "good", "advanced"), and the number of conducted projects before the study (**#Projects**). Additionally, we highlight the **tools used** during the study and the duration of their analyses (**Duration**).

ID	Sector	Role	Expertise	#Projects	Tool(s) Used	Duration
A1	Enterprise Software	Product Manager	Good	1	Celonis	0:32:58
A2	Academia	PhD Student	Average	1	Disco, Celonis	0:38:50
A3	National Laboratory	Senior R&D Staff	Good	2-5	Disco	0:33:59
A4	Consulting	Senior R&D Staff	Average	1	Disco, ProM	0:44:41
A5	Consumer Goods	Process Analyst	Average	2-5	Celonis	0:45:19
A6	Freelance	Process Mining Consultant	Good	2-5	Disco, ProM	1:27:14
A7	Academia	Professor	Advanced	6-10	Disco	0:42:16
A8	Academia	Assistant Professor	Advanced	2-5	Celonis, ProM	0:37:57
A9	Enterprise Software	Process Mining Consultant	Good	>20	Celonis	0:33:16
A10	Academia	Senior Researcher	Novice	None	Disco	0:34:59
A11	Freelance	Business Analyst	Advanced	6-10	Disco	0:43:38
A12	Academia	Assistant Professor	Advanced	2-5	ProM	0:40:06
A13	Enterprise Software	Partner Relationships Manager	Basic	1	Celonis	0:35:12
A14	Enterprise Software	Product Manager / Engineer	Good	10-20	Celonis	0:37:59
A15	Academia	Assistant Professor	Good	2-5	ProM, Disco	0:45:39
A16	Academia	PhD Student	Average	2-5	Disco	0:34:39
A17	Enterprise Software	Process Mining Consultant	Good	10-20	Celonis	0:36:50
A18	Academia	Assistant Professor	Advanced	6-10	Disco, ProM	0:36:46
A19	Enterprise Software	Partner Relationships Manager	Basic	1	Celonis	0:29:21
A20	Academia	Graduate Student	Average	None	Disco, ProM, Pm4Py	0:49:28
A21	Academia	Graduate Student	Basic	1	Disco	0:20:41
A22	Enterprise Software	Process Mining Consultant	Average	2-5	Celonis	0:41:01
A23	Enterprise Software	Analytics Product Manager	Good	2-5	Celonis	0:34:39
A24	Packaging	Digital Project Manager	Good	2-5	Disco	0:33:20
A25	Academia	Professor	Good	2-5	Disco	0:37:29
A26	Academia	PhD Student	Good	2-5	Disco, ProM	0:37:17
A27	Professor	Professor	Average	2-5	Disco	0:31:14
A28	Academia	PhD Student	Average	2-5	Disco, ProM	0:35:21
A29	Constructions	Operational Excellence Lead	Advanced	6-10	Celonis	0:36:26
A30	Insurance	Process Analyst	Good	2-5	Disco	0:33:40
A31	Insurance	Operational Excellence Lead	Advanced	>20	Disco	0:36:11
A32	Academia	Professor	Good	6-10	Celonis, Disco	0:39:39
A33	Academia	Senior Researcher	Advanced	10-20	ProM	0:43:28

Process Mining Analysis Sessions: Encoded Perspectives

1. Visualization Perspective

To better understand the use of visualizations during analysis, we annotated segments in the videos where a specific type of visualization is used. As such, visualization events describe *HOW* a task is performed [1].

2. Focus Perspective

The focus perspective is inspired by the *WHAT* dimension of Brehmer's typology [1]. In our case, it describes the segments in video recordings during which analysts focused on specific data aspects of the process under consideration. To simplify the analysis, we synchronized the Visualization Perspective and Focus Perspective during coding by splitting visualization events according to a change of focus by the user.

3. Intent Perspective

To capture the purpose behind analytical activities, we aimed to annotate video segments with the intent that was guiding the analyst's actions. Intent events reflect *WHY* a specific approach is taken during the analysis [1]. For the identification of events, we draw inspiration from the set of analysis processes identified by Crisan et al. [2] and defined generic analysis intents, such as exploration and experimentation for process mining analysis.

4. Interpretation Errors Perspective

To understand potential misinterpretations during analysis, especially during the use of visualizations, we identified segments where analysts made interpretation errors. Such interpretation errors could be identified from the think-aloud protocols. As an example, A6 concluded that *"They have paid much more than the amount of the payments they had to"*, when checking a single traffic fine. To arrive at this conclusion, A6 had to summarize the displayed amount with additional fine expense, which summed up to 87 while the payment amount was also 87, indicating indeed a correct payment, and not one where the payment amount was higher than the fine amount.

5. Observation Perspective

We documented observations to capture moments when analysts noted specific insights (which might be combined with interpretations) that they perceived as relevant. Observations, as Sacha et al. describe, are integral steps toward findings, bridging the gap between raw data representations (i.e., models and visualizations) and insights [3].

In the literature, observations are often used to define insights or findings [3], [4]. For us, observations represent a broader category, summarizing these terms.

In our data, observations can be identified from the think-aloud protocols, indicating moments when analysts expressed what they identified by examining a visualization. In several cases, observations are specifically indicated as unexpected, surprising, or outstanding (e.g., A7: *"OK, so they create, they send it, they receive it. It takes too long... Add the Penalty and sent to Credit Collection. I was not expecting that."*) but also capture statements about the data, its structure, and related interpretations, that are expressed without further assessment of the relevancy of the observation (e.g., A26: *"Everything starts with Create Fine so at least the starting part of the process is structured"*).

Visualization Perspective – Coding Scheme

This section provides details about the visualization types that were used during our observational study. We first provide an overview of the visualizations by describing them on three different granularity levels and indicating a reference to relevant literature, when applicable. Additionally, we indicate the type of task, inspired by the taxonomy in [5] and the data type, inspired by Shneiderman [6]. These latter two categories are provided to indicate how existing taxonomies could inform a different grouping and abstraction of visualization types, potentially useful in other analysis settings. Below, screenshots of each visualization type can be found.

Visualization Type					
L1	L2	L3	Ref.	Type of Task	Data Type
Process Model	Directly-Follows Graph (DFG)	Process Explorer	[7]	Structure, Connectivity	Network Data
	Directly-Follows Graph (DFG)	Variant Explorer		Structure, Connectivity	Network Data
	Inductive Visual Miner (IvM)	IvM	[8]	Structure, Connectivity	Network Data
	Petri net	Petri net	[9]	Structure, Connectivity	Network Data
	Business Process Model and Notation (BPMN)	BPMN	[10]	Structure, Connectivity	Network Data
	Causal Net	Causal Net	[11]	Structure, Connectivity	Network Data
	Social Network	Social Network	[12]	Structure, Connectivity	Network Data
Other Process Representation	Sequence Chart	Variant Sequence		Comparison, overview	1-dimensional data
	Sequence Chart	Trace Sequence		Overview, Trend	Temporal data
Other Chart Visualization	Scatterplot	Dotted Chart	[13]	Correlation	Multi-dimensional
	Scatterplot	Scatterplot		Correlation	Multi-dimensional
	Line Graph	Line Graph		Correlation, Trend	Temporal data
	Bar Chart	Bar Chart		Comparison	2-dimensional data
	Bar Chart	Histogram		Distribution	Temporal data

Directly-Follows Graph (DFG) – Process Explorer

The Process Explorer displays a process model in a graph-like structure. Based on the implementation, the level of detail (how many relations and activities are displayed) can be selected. Additionally, annotations can be added to the nodes and edges of the graph. Those typically depict the case (trace) frequency (in how many traces of the event log the activity or relation was observed) or performance information, such as the mean or median throughput time for activities or the transition from one activity to another.



Figure 1. DFG in the “Process Explorer” in Celonis. The sliders on the right allow users to change the level of detail for the graph. Annotations can be changed by selecting the top left icon. Here also custom metrics can be defined.

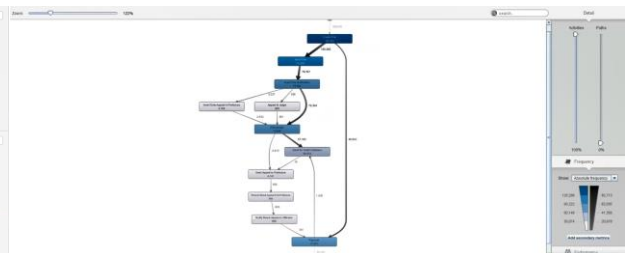


Figure 2. DFG in Disco (called “Process Map”). In the window on the right, users can change the level of detail for the graph and select annotations based on pre-defined metrics.

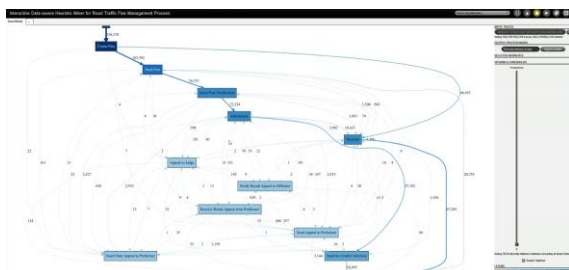


Figure 3. Process Explorer in ProM.

Directly-Follows Graph (DFG) – Variant Explorer

Like the Process Explorer the Variant Explorer displays a DFG. However, instead of setting the level of detail for activities and paths, users can select variants (sets of traces with the same sequence of activities).

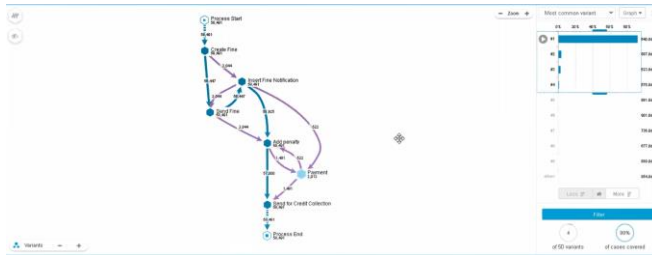


Figure 4. DFG in the "Variant Explorer" in Celonis.

Inductive Visual Miner (IvM) Model

The inductive visual miner is a process mining discovery algorithm that comes with its own visualization. The model visualization follows the principles of process trees. According to the author, "IvM shows models in an intuitive formalism that closely resembles Petri nets, process trees, and BPMN models" [2].

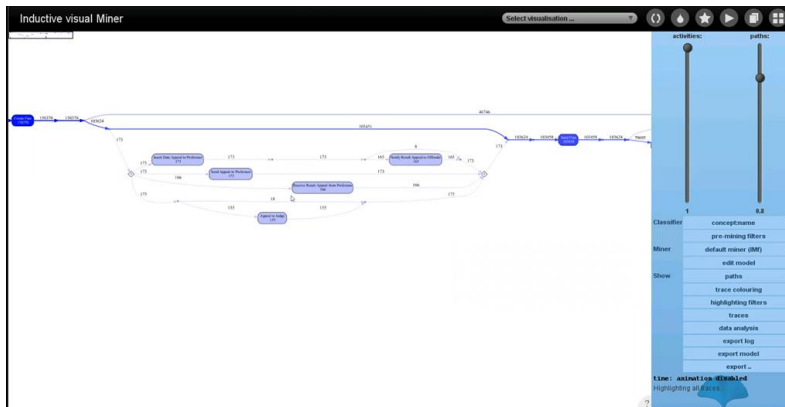


Figure 5. IvM Model in ProM.

Petri net

The Petri net is a mathematical modeling language that is mainly used to describe distributed systems. However, it can be applied to processes to model them as a set of states and transitions.

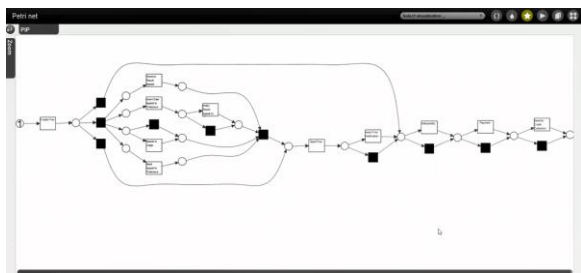


Figure 6. "Standard" Petri net in ProM.

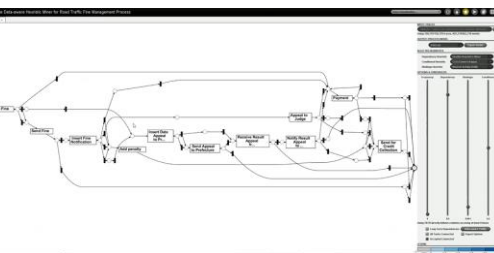


Figure 7. Petri net like representation produced by the data aware heuristic miner in ProM.

Business Process Model and Notation (BPMN)

BPMN is a graphical language for process management and provides symbols that allow to record, model, document, design, execute, measure, monitor, and control business processes.

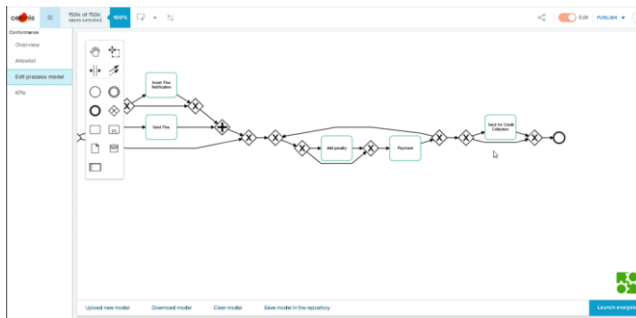


Figure 8. BPMN Interface in Celonis. Users can generate a BPMN model by selecting variants of their choice or develop their own model from scratch.

Causal Net

The Causal net, also called C-net is a “graph where nodes represent activities and arcs represent causal dependencies” [11]. Compared to Petri nets, in Causal nets, no places are displayed and the routing logic is only represented by the direct bindings of nodes.

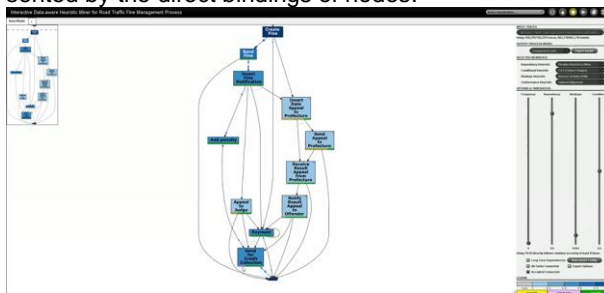


Figure 9. Causal Net in ProM. Analysts can modify multiple parameters to mine a model that represents the event log well.

Social Network

The Social Network highlights the interactions and relationships between individuals or roles involved in a process. It typically uses metrics like frequency of communication or task handovers to reveal collaboration patterns, responsibility distribution, or potential bottlenecks in team dynamics.

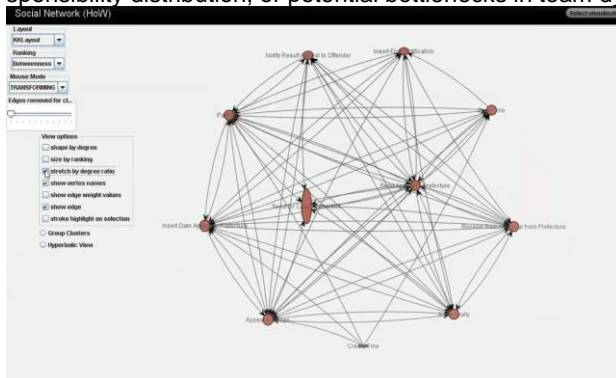


Figure 10. Social Network in ProM. In our study, analyst used it to assess the relationship of activities instead of resources.

Sequence Chart - Variant Sequence

Provides a visual overview of all variants in the log by displaying each group of traces that has the same control-flow order of activities. For easier recognition, activities are colored. When a variant is selected, all single traces are shown on the right and attribute values can be checked. Also, some metrics for each variant, as well as for each trace are shown.



Figure 12. Variant Sequence in ProM.

Sequence Chart - Trace Sequence

Trace sequence visualizations display the events and attributes of one case (trace) in the log. They allow to assess the sequence of activities for the case, as well as attributes (e.g., timestamps) related to them.

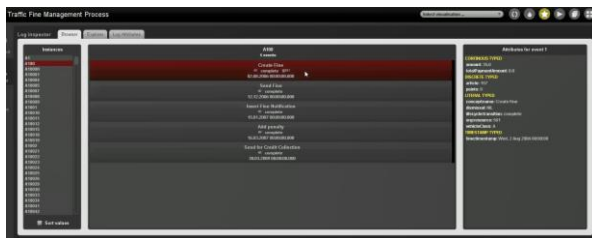


Figure 13. Trace Sequence in ProM.



Figure 14. Trace Sequence in Disco.

Scatterplot - Dotted Chart

The dotted chart visualizes the raw data from the event log, displaying each event (row) in the data as one dot in a chart. Options allow to flexibly set the x- and y-axis and to specify the meaning of colors.

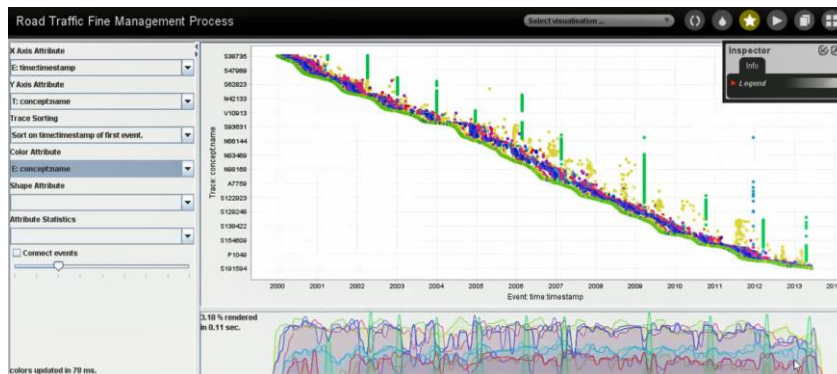


Figure 15. Dotted Chart in ProM

Scatterplot

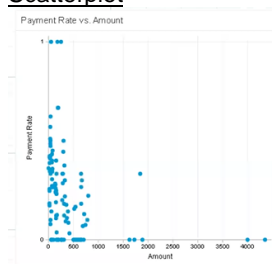


Figure 16. Classical Scatterplot in Celonis.

Line Graph



Figure 17. Line Graphs in Celonis and Disco.

Bar Chart – Bar Chart

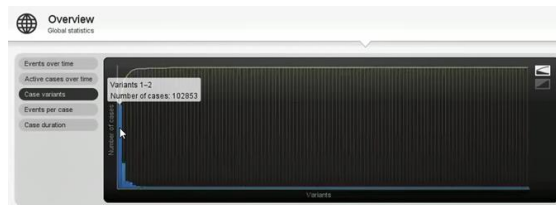


Figure 18. Bar Chart In Disco

Bar Chart – Histogram

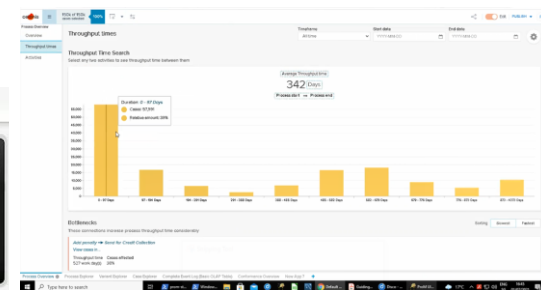


Figure 19. Histogram in Celonis

Focus Perspective – Coding Scheme

Focus			Description	min
L1	L2	L3		
Other Artifacts	Tool Documentation	Tool Documentation	Publicly available online sources, personal documentation, or "cheat sheets" that analysts might refer to.	2.6
Provided Material	Guiding Question	Guiding Question	A (pdf) document stating the guiding question for the analysis task.	45.6
	Artifacts	Artifacts	A (pdf) document providing domain knowledge for the road fine traffic management process (e.g., activity and attribute descriptions).	190
	Instructions	Instructions	A (pdf) document providing instructions for the analysis task.	4.45
Analysis Focus	Deviations	Deviations	Discrepancies between the observed process and an expected or predefined process model.	9.43
	Decision Points	Decision Points	Points in the process where branching occurs. Tools might provide decision rules for decision points.	12.9
	Flow	Flow	Structure of the process, including frequent and less frequent paths or transitions between activities.	265
	Event log	Event log	Raw data provided as .csv or .xes file that can be checked in a csv-reader (e.g., to analyze the available column names or the structure of the data).	39.1
	Performance	Performance	Time-related aspects of the process, such as the cycle time, throughput, or the distribution of events over time.	103
	Attributes	Attributes	Further data attributes, such as payment amount, additional fine expenses, or car type.	112
	Statistics	Number of events	Count of events (activities) for the complete or filtered event log.	13
		Number of cases	Count of cases (traffic fines) for the complete or filtered event log.	139
		User-derived Metrics	Metrics generated by analysts to summarize or quantify aspects (e.g., payment rate) that do not directly exist in the event log.	53.6
		Others	General statistics, e.g., produced by process mining algorithms to indicate the fitness of the model or confidence metrics.	24.9
	Cases	Cases	Individual cases or process instances are typically assessed to understand complexity or gain an overview of a single case without a more specific focus on its flow or attributes.	27.8
	Variants	Variants	Groups of control-flow paths of the process.	101
		Happy Path	The "ideal" or most frequent sequence of activities in the process, often considered a benchmark for comparison to other variants.	6.43
	Activities	Activities	Focus on specific activities (events) performed during the process to understand their role or identify them within a process model	163
	Activities	End Activities	Activities that conclude a process instance (traffic fine).	46.6

Intent Perspective – Coding Scheme

Intent	Description	Example	min
Defining Need	Understanding the objectives and goals of the analysis.	A8: "Ok, so the question is the three most prevalent circumstances in which the fines and related expenses are not paid or paid not in full. All right. And for each circumstance, if possible from the information at your disposal, can you provide at least one reason that explains why the fines are not paid in full? Very nice."	66.8
Profiling	Assessing the event log to understand its structure, characteristics, and distribution of timestamps, events, and control-flow patterns.	A22: "OK, so what I first would do if I would start from the beginning, I would look at my event logs and all the other guidelines that I have with the data model in this part. So far, it seems to me at the beginning, quite, um, complicated. I would go back in this case to, um, to Celonis the environment, just look if the process has already been uploaded. So, I can see so I can get a high-level understanding of it."	256
Data Wrangling	Involves cleaning, transforming, or organizing data to ensure it is suitable for analysis.	A9: "The main problem for me currently is I want to see the ratio of my bad cases, let's say in any pie chart or any ratio, like it's partially are not paid in 10 percent and then it's partially not paid in 50 percent to go into a root cause analysis. I want to have one KPI where I can differentiate bad and good cases."	112
Experimentation	Involves assessing or validating a hypothesis derived either from prior analysis insights or external knowledge.	A18: "My hypothesis is that between Create Fine and Send Fine, let's move to Disco to check that between Create Fine and Send Fine there is, um there might be some delay. So how can I do that?"	63.4
Exploration	Involves investigating the event log or process model without a predefined hypothesis. May include examining different process paths, variants, or attributes to find relevant insights related to the task without a predefined focused on any particular aspects.	A3: "So, let me take a little bit uh browse a little bit more. If I can spot something else. I'm not going to be able to find the three most prevalent reasons. But at least I feel that find one, um.."	351
Descriptive Modeling	While still exploratory, descriptive modeling begins with the intent to describe specific features, metrics, attribute distributions, or control-flow aspects of the event log.	A16: "Then we're interested in the cases in which there was no payments, not in full. So we'll have a look at the paths that are also here, because that might give us an idea of the frequency of cases where there was no payments before the process closes. I can see here already that ..."	195
Diagnostic Modeling	Aims to uncover the underlying reasons for observed process behaviors, attribute values, or metrics.	A17: [Attempts to look into root causes for the situation he identified]: "There is a high number here. This one, I could do it in a different way, but there is also a high number. And this one is small, small, relatively small and high, so what I'm going to do is I'm going to filter that one and I'm going to try to understand what's going on in this case? Ok, that is 40% of the cases. Well. Let's see. I will try to be a bit more selective. Ok, though, it seems that it was credit recollection, that is a complicated one."	115
Verification	Reflecting or actually checking the accuracy, validity, and completeness of the analysis results or the process to get there	A14: "Ok, so actually then we'll go back to process overview or to the Process Explorer, Variant Explorer doesn't end with payment, remove this condition. Ok, so it still looks quite similar. Maybe, but it seems like there was definitely a flaw in my logic."	14.2
Interpretation	Making sense of the analysis results, understanding them in the context of the problem domain, or translating them into actionable insights or recommendations.	A11: "Already by looking here at the flow we can see those coming from another process step then payment and ending just there is a candidate for the root causes. But still keeping in mind that we need to do some filtering because it could be that the appeal court has not decided anything yet. So we are waiting for that. So more process steps or directions will occur or it could also be that we are going to pay, but we are not overdue yet. So, there is no reason for saying that, that the payment outstanding is a violation."	17.6

Event Sequences Data

This section provides an overview of the dataset accompanying our paper. The dataset captures event sequences from video-coded sessions of analysts interacting with visualization tools during process mining analysis. It serves as the raw data for our study and is made available in the file **Event_Sequences.csv**.

Data Description

The dataset is structured as a CSV file, where each row represents a coded event segment. The columns provide information on the document (session), the assigned codes, perspectives, timestamps, and durations. Below is a description of each column:

- **Document name:** Identifier for the session (i.e., the participant ID, cf. page 2).
- **Code:** Assigned code representing a specific activity or event including the full code name with all granularity levels as exported from MaxQDA.
- **Perspective:** High-level categorization of the event based on analytical perspectives (cf. page 3).
- **CodeName_L2:** Second-level categorization of the coded event.
- **CodeName_L3:** Third-level categorization of the coded event.
- **CodeName_L4:** Fourth-level categorization of the coded event (if applicable).
- **start_ms:** Start time of the event in milliseconds, relative to the beginning of the session, which always starts at 0 for each participant.
- **end_ms:** End time of the event in milliseconds relative to the beginning of the session.
- **duration_ms:** Duration of the event in milliseconds.
- **session_duration_ms:** Total duration of the analysis session of the specific participant (Document name) in milliseconds.
- **session_duration_min:** Total duration of the analysis session of the specific participant (Document name) in minutes.
- **starttime_normalized:** Normalized start time of the event within the session (relative to total session duration) on a scale from 0 to 1.
- **endtime_normalized:** Normalized end time of the event within the session (relative to total session duration) on a scale from 0 to 1.
- **Segment:** Text of the think-aloud protocol associated with the coded event. Only available for events of the Observation and Interpretation Errors Perspective.

This dataset provides a structured foundation for analyzing how and when visualizations are used in process mining analysis. Researchers can utilize it to explore behavioral patterns and interactions with visual tools or to evaluate new visualizations or analysis tools for multi-perspective and multi-granular datasets.

References

- [1] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Trans Vis Comput Graph*, vol. 19, no. 12, pp. 2376–2385, 2013, doi: 10.1109/TVCG.2013.124.
- [2] A. Crisan, B. Fiore-Gartland, and M. Tory, "Passing the Data Baton : A Retrospective Analysis on Data Science Work and Workers," *IEEE Trans Vis Comput Graph*, vol. 27, no. 2, pp. 1860–1870, Feb. 2021, doi: 10.1109/TVCG.2020.3030340.
- [3] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim, "Knowledge Generation Model for Visual Analytics," *IEEE Trans Vis Comput Graph*, vol. 20, no. 1, pp. 1604–1613, 2014, doi: 10.1109/TVCG.2014.2346481.
- [4] C. North, P. Saraiya, and K. Duca, "A comparison of benchmark task and insight evaluation methods for information visualization," *Inf Vis*, vol. 10, no. 3, pp. 162–181, Jul. 2011, doi: 10.1177/1473871611415989.
- [5] M. Tory and T. Moller, "Rethinking Visualization: A High-Level Taxonomy," in *IEEE Symposium on Information Visualization*, IEEE, pp. 151–158. doi: 10.1109/INFVIS.2004.59.
- [6] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," in *The Craft of Information Visualization*, Elsevier, 2003, pp. 364–371. doi: 10.1016/B978-155860915-0/50046-9.
- [7] S. J. J. Leemans, E. Poppe, and M. T. Wynn, "Directly follows-based process mining: Exploration & a case study," *Proceedings - 2019 International Conference on Process Mining, ICPM 2019*, pp. 25–32, Jun. 2019, doi: 10.1109/ICPM.2019.00015.
- [8] S. J. Leemans and Prom, "Inductive visual Miner manual," 2017.
- [9] B. F. Van Dongen, A. K. Alves De Medeiros, and L. Wen, "Process Mining: Overview and Outlook of Petri Net Discovery Algorithms," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5460 LNCS, pp. 225–242, 2009, doi: 10.1007/978-3-642-00899-3_13.
- [10] M. Chinosi and A. Trombetta, "BPMN: An introduction to the standard," *Comput Stand Interfaces*, vol. 34, no. 1, pp. 124–134, Jan. 2012, doi: 10.1016/J.CSI.2011.06.002.
- [11] W. Van Der Aalst, A. Adriansyah, and B. Van Dongen, "Causal Nets: A Modeling Language Tailored towards Process Discovery," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6901 LNCS, pp. 28–42, 2011, doi: 10.1007/978-3-642-23217-6_3.
- [12] A. Jalali, "Supporting social network analysis using chord diagram in process mining," *Lecture Notes in Business Information Processing*, vol. 261, pp. 16–32, 2016, doi: 10.1007/978-3-319-45321-7_2/FIGURES/9.
- [13] M. Song and W. M. van der Aalst, "Supporting process mining by showing events at a glance," *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)*, 2007, pp. 139–145.