Process Mining Practices: Evidence from Interviews

Francesca Zerbato[1], Pnina Soffer[2], and Barbara Weber[1]

[1] Institute of Computer Science, University of St. Gallen, Switzerland
{francesca.zerbato|barbara.weber}@unisg.ch
[2] University of Haifa, Mount Carmel, 3498838 Haifa, Israel
spnina@is.haifa.ac.il

Abstract. Process mining provides organizations with methods and techniques to extract knowledge from event logs. In their work, process analysts can draw from the wealth of techniques developed over the years by researchers and professionals. Still, there is limited understanding of how process mining is used in practice and, in particular, how individual analysts approach the analysis stage. Towards filling this gap, we conducted semi-structured interviews with 37 practitioners and academics working with process mining. Based on the results of the interviews, we characterize common analysis strategies and examine related challenges and factors affecting their use in practice. Our findings contribute to an improved understanding of process mining practices and provide a solid empirical basis for future research developing guidance and support addressing the practical needs of process analysts.

Keywords: Process Mining · Interview Study · Mining and Analysis Stage · Analysis Strategies

1 Introduction

Process mining provides organizations with data-driven methods and techniques to extract knowledge from process execution data in the form of event logs [1].

Over the last two decades, process mining as a field of research has grown in maturity, leading to the development and consolidation of techniques and tools for analyzing event logs [10]. However, research has mainly emphasized technical work giving less attention to empirical studies focusing on understanding how process mining is used in practice.

With the increasing adoption of process mining techniques among practitioners, a growing number of empirical studies have explored how process mining is used in organizational settings [3]. For example, Grisold et al. [10] investigated process mining adoption through the eyes of managers, while Martin et al. [14] elicited a broad set of opportunities and challenges of using process mining in organizations. Still, the work practices of analysts, including the strategies they follow to conduct their work tasks, remain largely unexplored [11].

Process mining projects include different stages, starting from project planning and data extraction and progressing until the results are utilized for process improvement and support [7]. Typically, these stages unfold in many iterations, leading to different activities, goals, and, potentially, challenges, which, in turn, may require different kinds of support. In this work, we focus on one specific stage - the *mining & analysis* stage (henceforth *analysis*), in which process analysts apply process mining techniques to event logs to address analysis questions and gain insights from the data. Thus, we leave the other stages out of our scope, focusing on gaining deep insights into the practices of process mining analysis.

We aim to complement existing research by taking an *individual perspective* [3] and understanding from analysts how they work in the analysis stage. Indeed, we believe that learning how individual analysts conduct process mining analyses can help develop support to address the practical needs of process analysts and stakeholders involved in the analysis stage. In particular, we ask ourselves the following research question (RQ): **What are common strategies used in the analysis stage?**, where by strategy we refer to "an approach, a manner or a means to achieve a certain intention" [15]. To address this question, we conducted an interview study with 37 academics and practitioners working with process mining in different organizations. The interviews were designed in the context of a broader observational study, where we used a process mining task as an anchor to let participants reflect upon a concrete analysis and share their work experiences.

In this paper, we present the results of these interviews, which allowed us to look at the analysis stage from the retrospective thoughts and reflections of our participants and learn what strategies they apply in their daily work. Our findings include (i) a characterization of analysis strategies, organized into four main phases representing intermediate analysis goals; (ii) examples of recurring challenges associated with each phase; (iii) a set of factors influencing the use of the strategies in practice. By raising awareness about the work practices of process analysts, our findings can help both process analysts and business stakeholders to reflect upon their (joint) work and learn about possible difficulties. In addition, they provide a solid empirical basis for designing methods and tools to support process analysts, guiding several directions for future research.

The paper is structured as follows. Sect. 2 provides an overview of related work. Sect. 3 presents the research method followed to design, conduct and analyze the interviews. Sect. 4 reports our findings. Sect. 5 reviews our results and discusses the limitations of our work, outlining directions for future research.

2 Related Work

Our work falls into the stream of research investigating process mining practice from a general perspective, i.e., not tied to a specific organizational setting.

Following the increasing uptake of process mining in the industry, more and more researchers have investigated the use of process mining in practice using a variety of research methods and looking at the effects of process mining adoption from different angles [3].

One group of related papers has looked into the practical use of process mining by using published empirical studies, case studies or process mining reports as their main source of knowledge. For example, Thiede et al. [21] explored the use and maturity of process mining technology in organizations from a service perspective. Emamjome et al. [9] investigated the diffusion and maturity of tools and methodologies from case studies discovering, for example, that the thoroughness of the analysis stage, among others, has not improved over the years. Also using published case studies as a source, Koorn et al. [12] focused on understanding the goals and methods of evaluations in which domain experts are involved. Then, based on their findings, they proposed six strategies for the qualitative evaluation of findings in process mining projects. Klinkmüller et al. [11] focused on examining visual representations from BPI Challenge reports to understand the relationships between domain problems captured by analysis questions and the information needs of process analysts. While these works provide valuable insights on a number of aspects relevant to process mining practices, they rely on case study and scientific reports, which often contain little or no information about the dynamics of the *process of process mining*.

Another group of works has focused on investigating the use of process mining in practice through explicit reporting from experts working in the field, who were directly involved in empirical studies. Syed et al. [20] conducted an interview study with 9 stakeholders in the context of a Dutch pension fund, identifying challenges and enablers of process mining adoption. Their findings touch upon different levels, going from tensions between teams to user challenges with learning tools. Grisold et al. [10] conducted a focus group study with 22 process managers to investigate the benefits and challenges of process mining adoption in organizations. Eggers et al. [8] conducted a multiple case study to investigate how organizations engage with the process transparency created by process mining to achieve increased process awareness. The study, which included 24 semi-structured expert interviews among the different data sources, revealed seven mechanisms that employ process mining to achieve increased process awareness on different levels. Recently, Martin et al. [14] conducted a Delphi study with 40 experts to explore the opportunities and challenges of process mining adoption in organizations. Some of the key challenges they discovered, which also emerged in [10,20], include elusive business value and unclear organizational anchoring. Our work also falls within this second group of studies. However, while the papers presented above reveal how process mining is used within organizations, we take an *individual* perspective [3].

The only paper so far that looks at process mining from an individual perspective is our previous work [22], which presents the results of a pilot study combining behavioral data and interviews to understand patterns and strategies of the initial exploration phase. In contrast, this paper, which is based on a newly collected data set, considers the whole analysis stage, focusing on strategies that analysts apply in their general work practices and considering factors and challenges affecting their use.

3 Research Method

This section describes the design of our study, the data collection and analysis.

Study Design. To improve our understanding of strategies used in the analysis stage (cf. RQ in Sect. 1), we designed an interview study following the empirical standards for qualitative surveys [17]. The interview was part of a broad observational study where participants engaged in a realistic process mining task. In the context of this paper, the process mining task served as an anchor to let participants work and reflect upon a concrete analysis they could use as a reference to share their work experiences. To ensure familiarity with event log analysis, we defined two requirements for participating in the observational study: (i) having analyzed at least two real-life event logs in the two years prior to the study and (ii) being knowledgeable of at least one of the process mining tools available for the task. In addition, for the interview study, we considered participants having experience in process mining projects aimed at analyzing process data for a customer to ensure that they had gained sufficient practical experience with event log analysis. Such requirements allowed us to exclude beginners from our population of interest but still include participants with different backgrounds (e.g., academics vs. practitioners) and varying levels of experience and expertise, which are needed to gain a broad understanding of analysis strategies.

Materials. The process mining task was designed to observe participants as they analyze an event log guided by a high-level question. Specifically, we used the road traffic fine management event log [6] and asked participants to investigate circumstances (scenarios) for not paying a fine and, if possible, identify potentials reasons for doing so. The event log was ready to be analyzed so that participants could fully focus on the analysis stage. For the analysis, participants had at their disposal bupaR[1], Celonis[2], Disco[3], Pm4Py[4], ProM[5], and SQL, which we selected considering the top-six tools used in BPI Challenge reports published until 2020[6]. We also prepared an online form to collect the participants' answers to the task question. Finally, we developed the interview guide following a semi-structured approach [16], organizing questions into four main themes related to the analysis stage: (i) activities and artifacts; (ii) goals; (iii) strategies; and (iv) challenges. Each question was formulated twice, first concerning the process mining task and then generalizing to the work practices of the participants. All the materials were pilot tested with two researchers external to the author team and adjusted based on their feedback.

Data Collection Procedure. We invited participants in our professional networks, ensuring diversity in their affiliation, job role and position, and tool knowledge.

[1] bupaR: https://bupar.net
[2] Celonis: https://www.celonis.com
[3] Fluxicon Disco: https://fluxicon.com/disco/
[4] PM4Py: https://pm4py.fit.fraunhofer.de
[5] ProM: https://www.promtools.org/doku.php
[6] BPI Challenges: https://www.tf-pm.org/competitions-awards/bpi-challenge

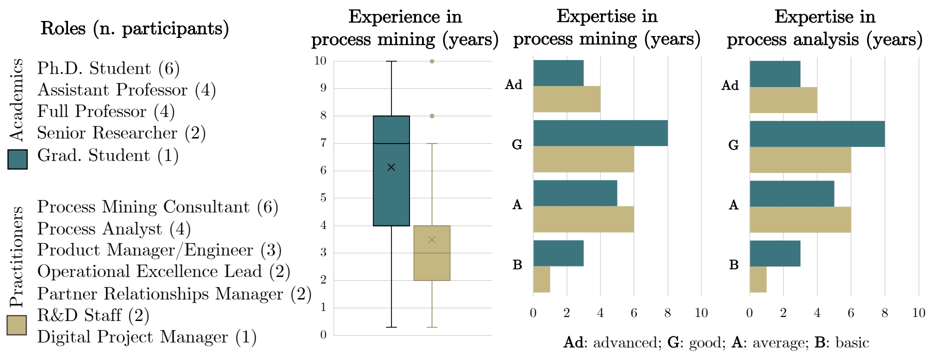

Fig. 1. Job role, process mining experience and expertise of the 37 participants.

Eleven participants were additionally recruited with the help of five participants to include in our sample users of all the five process mining tools available for the task. The participation requirements were explicitly stated in the invitation email, along with a description of the study purpose and procedure. The first author collected the data between May and July 2021 via virtual sessions with the participants. We discontinued the data collection when we achieved data saturation [19], i.e., we noticed that responses repeated across interviews.

Before each session, we collected demographic characteristics with a questionnaire and screened participants for participation requirements, process mining experience and expertise, and project experience. On the appointed day, we instructed participants about the task and gave them access to a remote desktop environment with the study materials and tools. The task was silently supervised by one author, who was available to support with questions and technical issues. Then, we asked participants to report their answers in an online form. Finally, we conducted the semi-structured interviews following our guide and prompting interviewees to share anecdotes and work experiences within their current organizations [16]. All the sessions were conducted in English and recorded.

Data Validation and Analysis. Overall, 41 people participated in our study. We averaged 83 minutes of recordings per participant, including the process mining task and the interviews, which we fully transcribed verbatim. For this paper, we excluded four participants from the analysis. Two participants did not conduct the task prior to the interview for personal and technical reasons. Two others did not have experience with process mining projects aimed at analyzing process data for a customer, which we considered relevant to our research question. Thus, in this paper, we focus on the interviews of the 37 remaining participants, i.e., 20 practitioners and 17 academics working in 29 different organizations. Their interviews lasted 1136 minutes in total and 31 minutes on average.

Fig. 1 gives an overview of the 37 participants. All the participants meet the participation requirements and have experience with at least one process mining project requiring analyzing process data for a customer. The participants have a variety of job roles, such as process mining consultant, process analyst, Ph.D.

student, and full professor, among others. Overall, the participants reported an average of 4.5 years of process mining experience. Most of them additionally indicated experience in related fields, with some indicating more than 6 years of experience in process analytics (14/37), business intelligence (17/37), and data science (15/37). More than half of the academics (11/17) reported also having worked in the process mining industry for two years on average.

Given the exploratory nature of our research, we followed an inductive approach for the analysis. More specifically, we relied on the qualitative coding guidelines in [18] and coded the interview transcripts in several rounds. First, we engaged in an initial coding round, analyzing each participant individually and fragmenting the text using in-vivo and process coding [18] to identify core concepts and steps related to strategies. Then, we used focused coding to refine and aggregate them into categories, considering codes supported by at least 10% of the participants. For example, we included strategy "choose tool" as part of "determine analysis approach", as we recognized that selecting the "right" tool is part of planning the analysis approach. Finally, we relied on axial coding to focus on the most frequent categories and find relationships among them until we achieved saturation. One author coded all the data in several iterations. The other authors checked the codes independently to ensure consistency. Throughout the analysis, all the authors revised and refined the codes collaboratively in several meetings. Our analysis led to 16 strategies related to the analysis stage. Factors affecting the use of the strategies in practice also emerged from our coding. Then, we organized the strategies into four main phases and iterated once more through the data with this structure in mind to code examples of common challenges related to each phase. The interested reader may find the documents with details about data collection and analysis based on the empirical standard on qualitative surveys [17] on: https://doi.org/10.5281/zenodo.6644982.

4 Findings

In this section, we report on our findings of strategies related to the analysis stage in Sect. 4.1 and factors affecting their use in practice in Sect. 4.2.

4.1 Phases and Strategies of the Analysis Stage

From our analysis, we derived 16 strategies that our participants apply in practice, some of which they also used in the process mining task (cf. Sect. 3). Based on our coding, we organized the strategies into four main phases representing intermediate analysis goals: **understand**, **plan**, **analyze**, and **evaluate**.

Understanding covers strategies to understand the problem and make sense of the needs of business stakeholders, the business domain, and the data. Planning includes strategies to devise an analysis plan. Analyzing refers to strategies to execute the analysis by applying process mining techniques within tools. Evaluating covers strategies to verify and validate analysis artifacts and findings. The identified phases resemble known phases from the problem-solving literature [4].

Since the analysis unfolds as a highly iterative and flexible process [7], phases are not strictly performed in order, but analysts rather cycle through them in several iterations, potentially skipping some phases.

In the following, we break down the *analysis* stage within the context of each phase and describe the strategies and their rationale with example statements from our participants, whose numeric IDs (p#) are reported in parentheses. For each strategy, we also indicate the number of participants explicitly mentioning it in parentheses. Then, for each phase, we summarize example challenges emerging from our interviews. A summary of the strategies is provided in Table 1.

Understand. Our participants referred to understanding as a crucial initial step of any analysis, covering (S1) problem understanding, (S2) domain understanding, and (S3) data understanding.

(S1) Understand the problem (6/37) refers to understanding the problem under analysis, including business stakeholders' needs and how process mining can be enacted within a specific organizational context. Problem understanding helps analysts orient themselves in the problem space and assess if stakeholder needs can be addressed with process mining because *"sometimes the problem is not directly a process problem but is a more statistical nature problem"* (p34). Also, it allows them to *"define the objective of the analysis"* (p1), *"translate that into an analysis that enables them* [the stakeholders] *to achieve that objective"* (p14), and figure out which type of process mining to apply.

(S2) Understand the domain (16/37) refers to gaining knowledge about the domain in which the processes under analysis are enacted, including *"which processes are in place"* (p39), what is the business meaning of the activities, who are the actors involved in the process, and what business rules are in place. Domain understanding *"helps with putting everything into context"* (p20) and *"knowing where to search"* (p14), and it seems necessary to extract and validate the data and define analysis questions that are in line with stakeholder needs.

Usually, problem and domain understanding occur during workshops with stakeholders, such as process owners and domain or IT experts, where analysts can *"ask about the problem"* and get *"an introduction to the business"* (p9). However, some participants (6/37) said that they *"don't always have the domain knowledge"* (p18) and, at times, *"lack interaction with business people"* (p34). Thus, they are used to deriving domain understanding from available documentation, such as *"secondary data and reports issued by the organization"* (p30).

(S3) Understand the data (19/37) concerns learning the data structure, the attribute values, and the related data models. Analysts aim to know what the data contains, how it is *"prepared and formatted"* (p3), and *"how it interacts [...] how it changes when I just filter"* (p9). They also inspect the data to know *"what is possible to analyze"* (p18). Some participants (6/37) mentioned focusing on the data structure to learn the main components of the log and how they can create new features to use later in the analysis. Data understanding often builds upon raw event data, accessed with process mining tools, spreadsheets, databases, and visual analytic software. Two expert participants (2/37) reported always starting from disaggregated data *"because sometimes you miss some other*

	Id	Strategy Description	Grp	#P	Example Statements
UNDERSTAND	S1	**Understand the problem:** understand stakeholders' needs and how process mining can support	all	6	"So, this is my way. So, opening my ears for problems. When I... I'm aware of problems, I try to determine if it is a process problem." (p34)
	S2	**Understand the domain:** learn about the domain in which the processes to analyze are enacted.	all	16	"...I try to get a grip on what the process is all about, what the pitfalls are, understand exactly the business rules and the domain..." (p11)
	S3	**Understand the data:** learn about the data structure, the attributes, and related data models.	all	19	"I would directly learn from the event log, the SQL [...] I would try to learn much more from the raw data before going to the map." (p37)
PLAN	S4	**Formulate analysis questions:** define questions based on stakeholders' needs such that they can be answered with process mining.	all	11	"We discuss what we get with the partner. [...] And with that and an overview of the field, for example, I would try to make some sense or some questions on my own." (p32)
	S5	**Prioritize analysis directions:** prioritize analyses based on the foreseen value of the findings.	P	11	"...like is this really an issue worth pursuing? [...] like if my bolts are rather inexpensive [...] then that might not be the biggest problem..." (p14)
	S6	**Determine analysis approach:** evaluate which approach suits the analysis best.	E	8	"And that's the first step. I can choose the analysis because that's part of the strategy to adapt the analysis based on the question." (p9)
	S7	**Map the question to the data:** identify the main "players" for the analysis and link them to the data	all	11	"I also will spend a lot of time on understanding the questions and map the questions to the data. How do I define a situation..." (p2)
	S8	**Make hypotheses:** hypothesize about possible answers for the question or explanations for the stakeholders' problems.	all	15	"...this is basically this CRISP thing, right? It's the data understanding and then... iterating upon and creating, like, looking at picking out hypothesis right from the question..." (p12)
ANALYZE	S9	**Understand the process:** learn about the process control flow using statistics or visualizations.	all	32	"...understand the different activities and the control flow[...] So, some understanding of the process will always be the first step" (p18)
	S10	**Discover patterns:** identify patterns in the data that can suggest new hypotheses or analysis directions.	all	20	"What I'm looking for is patterns. [...] Something that tells me 'ok, here is, uh, this is something that's worth investigating further.'" (p31)
	S11	**Classify and compare cases:** create groups of cases based on attributes or KPIs and compare them.	P	9	"...by defining these kind of KPIs, so I need a kind of central number, which helps me to understand if the situation is good or bad. " (p33)
	S12	**Look for correlations:** search for correlations among different process characteristics, e.g., attributes	all	7	"So, I would drill down in those attributes and see if I can find a correlation between a reason, between an attribute and the scenario..." (p17)
	S13	**Focus on narrow scopes of interest:** dig deep into selected parts of the event log deemed interesting.	all	24	"If I see some interesting point in the process model or in the event log, I go deeper and deeper and try to understand it." (p29)
	S14	**Test hypotheses:** check hypotheses using analysis algorithms and tools.	all	15	"Then I'll go to the conformance page and usually that validates that can help me validate or disprove my hypothesis." (p13)
EVALUATE	S15	**Verify artifacts/findings:** check findings' or artifacts' correctness	all	5	"My main steps were, first, to verify that process map generation was seemed to be correct." (p11)
	S16	**Validate artifacts/findings with stakeholders:** evaluate artifacts or findings with stakeholders.	all	18	"We always present our results to our customers, to our partners, to discuss the results with them to check if it makes sense..." (p22)

Table 1. Strategies S1-S16. **Grp**: highlights if the strategy was reported *mostly* by a specific group of participants, e.g., P practitioners or E experts; otherwise we write "all". **#P**: number of participants explicitly mentioning the use of a strategy.

parts in looking at the aggregated levels" (p24), whereas a less experienced one said that going into *"the raw data is like going down a rabbit hole"* (p13). Some participants (4/37) mentioned looking also at data quality. Although this is typically part of pre-processing, analysts check data quality when starting the analysis *"to know more about the* [provided] *log and the fields"* (p38).

Recurring challenges of the understanding phase are (i) the lack of domain knowledge and (ii) the limited availability of business stakeholders, as summarized by one participant: *"The second part of the challenge is the same, this is... the lack of understanding of the business, of the business process and, maybe, a lack of interaction with business people."*(p34). Indeed, *"the event log per se could be meaningless unless there is some semantics attached*" (p8) and, thus, analysts *"need somebody who has this business understanding"* (p25). Still, even if stakeholders are available, *"the biggest challenge is also getting those people to help"* and understand from a methodological viewpoint how *"to involve and interact with the domain experts, to feed my process with their information*" (p9).

Plan. Planning concerns coming up with a plan for conducting the analysis, including finding a *direction* for the analysis by (S4) defining analysis questions and (S5) prioritizing analysis directions, giving a *structure* to the analysis by (S6) determining an analysis approach, and finding concrete *entry points* for the analysis by (S7) mapping the question to the data and (S8) making hypotheses.

(S4) Formulate analysis questions (11/37) entails deriving questions from higher-level stakeholder needs and formulating them in suitable way for process mining. Analysts define questions to have a *"clear direction to look into the data"* (p33). Indeed, *"if you understand the question deeply, then you can go to that special part of the analysis that you need to answer that question*" (p6). Questions also help to *"avoid getting lost in the complexity"* (p33) as *"this could be dangerous because you have to keep focus"* (p26). Analysts formulate questions independently or during workshops with stakeholders, at times using standard business hypotheses and pre-defined analyses as templates to *"break the ice [...] so that you don't start with an empty piece of paper"* (p33).

(S5) Prioritize analysis directions (11/37) refers to prioritizing analysis directions based on the foreseen value or impact of the findings. Prioritizing helps analysts find analyses that are valuable for the stakeholder, rather than focusing on things *"already known to the customer"* (p11) or *"the small ones that could be accidental"* (p7) because *"you can spend hours and hours doing something that doesn't have an impact"* (p24). To prioritize, analysts combine different criteria, such as the monetary value of the objects involved in the process or indications about process execution frequency and time performance. This strategy was reported by 11 participants, of which nine were practitioners (cf. Table 1).

(S6) Determine analysis approach (8/37) consists of evaluating which method suits the analysis best, given an analysis question or lack thereof, the tools available, and the desired outcomes. Our participants, especially experts, reported following a *"structured approach"* (p33) that they *"adapt based on the question*" (p9) or *"the* [end] *user group"*, because *"if it's done too complex, then people won't use it"* (p17). One part of determining the approach is also selecting the tools to use for the analysis. Analysts do so based on the tools' strengths, the artifacts that the tools allow them to create, and the need for customization, e.g., *"depending on the questions, I will select a different tool, and with a different tool. I will go for a different thing"*(p7). Six participants reporting this strategy indicated "advanced" expertise, while two have "good" expertise (cf. Table 1).

(S7) Map the question to the data (11/37) allows analysts to break the question down into relevant "players" for the analysis that can be linked to key "data objects" in the event log, serving as concrete entry points for the analysis, e.g., *"I'm immediately looking at those event attributes because they were or they seem to be the most central data object for the guiding questions"* (p18).

(S8) Make hypotheses (15/37) refers to conjecturing about (i) explanations for the problem at hand, e.g., *"I try to create a hypothesis which explains the situations or the main problem for the initiative"* (p27), (ii) possible answers for the analysis questions, e.g., *"trying to build hypotheses related to the general question"* or (iii) observations in the data. Analysts often make hypotheses based on their experience, belief, or (limited) evidence in the data. Participants report *"picking out hypothesis right from the question"* (p12) to find entry points for the analysis, e.g., *"I look for the research questions and I start a preliminary hypothesis of... how it could look like or what patterns I could find"* (p7).

From our interviews, it emerged that analysts often plan their analysis very close to the question to avoid *"getting lost in complexity"* (p33) and to *"have a structure at hand"* (p9). Still, formulating questions is perceived as difficult. Indeed, *"it is very often hard to identify the correct question"* (p36), and *"often there would be a lack of concrete questions"* (p12) and *"you do not really know what you actually look out for"* (p37). While using pre-defined analyses seems to help analysts formulate initial questions, the lack of a concrete "structure" for planning the analysis and of methodologies that can help elicit questions in collaboration with stakeholders remains a challenge.

Analyze. In the *analyze* phase, process mining and visual analytic techniques are applied. Common strategies are (S9) understanding the process, (S10) discovering patterns, (S11) classifying and comparing cases, (S12) looking for correlations, (S13) focusing on narrow scopes of interest, and (S14) testing hypotheses.

(S9) Understand the process (32/37) concerns making sense of the control flow, including activities and their relations, variants, the happy path, the process structuredness, and the desired flow. It helps analysts to *"get familiar with the data and the process"* (p28) and *"characterize process behavior"* (p37), assess if additional data or artifacts are needed for the analysis and check their understanding of the process from stakeholders or documents, e.g., *"trying to confirm also my initial feeling on how the process looks like"* (p17). Process understanding is mostly supported by visualizations of directly-follows graphs (DFGs), variants, and dotted charts. DFGs seem to be by far the most used as they are *"very straightforward"* (p10) and allow one *"to imagine what is the problem you have inside the process"* (p26). Still, some expert participants (5/37) advised using DFGs carefully because *"they are easily misinterpreted"* (p39), and if *"you start with a filtered view, it can be really misleading"* (p12).

(S10) Discover patterns (20/37) aims to identify patterns and relationships among observed phenomena, scenarios, or outliers that help analysts make (new) hypotheses or find new analysis directions, e.g., *"you first try to understand what's going on, try yourself to look at meaningful patterns, high-frequent patterns"* (p39). Some participants (6/37) also reported trying to *"explore beyond a question"* (p30) and *"keep an eye open for things that stand out"* (p19) to *"attract those who are going to listen to my analysis afterward"* (p5).

(S11) Classify and compare cases (9/37) concerns partitioning the event log into *"subsets that are meaningful"* (p31) or *"dimensions that could be relevant"* (P24) to describe and classify clusters of cases and compare them to find differences between groups, e.g., *"I want to see the category versus the non-category"* (p31). Analysts classify and compare cases based on data attribute values, enumerated or numeric, or a custom KPI, i.e., a *"kind of central number, which helps me to understand if the situation is good or bad"* (p33). Comparing groups of cases is a way to search for correlations and potential root causes. This strategy was reported by nine participants, of which eight were practitioners (cf. Table 1). In particular, the term "KPI" was explicitly used only by practitioners involved in improvement initiatives where KPIs help monitor *"later on when I implement improvement measures, if the situation has really improved"* (p33).

(S12) Look for correlations (7/37) concerns looking for relations among different process characteristics, for example, by combining control-flow-related characteristics with data attributes or performance metrics to find *"influencing factors"* (p33). Correlations usually serve as hints for generating new hypotheses or for finding potential root causes of a problem, e.g., *"Usually, the root cause is somewhere hidden in the attributes. So, we have some pattern [...] and then you want to know why, but it's really seldom that it's because of the actual process flow. Usually, you have to correlate it with attributes"* (p37).

(S13) Focus on narrow scopes of interest (24/37) concerns focusing on specific parts of the log that capture the analysts' attention or are suggested by stakeholders. Analysts describe "drilling down" or "deep-diving" into the data with the help of filters to focus on interesting things, such as patterns, scenarios, issues, specific process behavior, or, simply, something *"that sticks out"* (p17). Usually, they narrow their scopes of interest to explain observations or spot inconsistencies, e.g., *"you know from a benchmark where the critical process steps are and you can deep-dive into those and see if it's really an issue"* (p23).

(S14) Test hypotheses (15/37) concerns gathering evidence in the data to confirm or disprove a hypothesis. Analysts test ideas when attempting to answer analysis questions, usually checking if hypotheses can be *"backed up by real data"* (p11). Hypotheses are tested with tools and algorithms such as conformance checking. Our participants described testing hypotheses by *"applying filters, looking at the results from different points of view"* (p34) or searching for *"violations to the hypotheses"* (p8). They also mention keeping track of *"tests to present later"* to show to stakeholders *"what we found and rejected"* (p32).

As reported by our participants, the main challenge related to the analysis phase is the lack of techniques for identifying causality, which makes it difficult to *"get real value from the analysis"* (p12). Indeed, *"understanding the real why is a big question"* (p1), and with process mining, *"you don't get the clear reason out of it [...] but it gives you indications for further deep dives"* (p33). Moreover, not knowing the root causes of a problem makes it difficult to recommend solutions, e.g., *"you found... Like the root cause or where it may be, but then what is the next step? So, process mining is like not helping you to solve the issue"* (p17).

Evaluate. Evaluation entails verifying (S15) and validating (S16) analysis artifacts and findings, helping analysts to determine the end of an analysis iteration.

(S15) Verify artifacts/findings (5/37) is about checking the correctness and accuracy of analysis artifacts, such as filters, dashboards, and visualizations, or (intermediate) results, usually by comparison with the original data. Participants reported checking the logic of filters to ensure that they "*build something tested, validated on like a raw table"* (p14), combining different tools *"to understand just for sanity check if we have the same insights in both tools"* (p27), or writing scripts to check "programmatically" if things do not "add up."

(S16) Validate artifacts/findings with stakeholders (18/37) concerns evaluating analysis artifacts or findings together with business stakeholders. Validation helps analysts to confirm if analysis artifacts and findings *"make sense in terms of domain knowledge"* (p22) for the stakeholders, who can assess if *"their questions were answered"* (p16) or if *"they see any additional things which might be interesting"* (p33). While this strategy is often used at the end of an analysis iteration, it can also occur spontaneously, for example, if analysts want to validate their hypotheses or discover something interesting during the analysis.

Besides the lack of domain knowledge, which also affects evaluation, a challenge related to this phase is the lack of tool support for validating findings, e.g., *"there is very little support for validation. So, you see something that stands out, is it actually true?"* (p19). Analysts can combine different tools for "sanity-check", but they mainly depend on stakeholders for validating what they observe *"because the data will not tell you that the answer is invalid"* (p12). Still, *"how do you evaluate your results with domain experts?"* (p16) remains an open question.

4.2 Factors Affecting the Use of Strategies in Practice

From our analysis, we derived four *factors* affecting the use of strategies in practice: Q the analysis questions, R the analyst's role, S the availability of business stakeholders, and T the tools used. Here, we reflect on such factors, reporting the strategies they affect and representative examples from our interviews.

Q **Analysis questions (S4, S6, S7, S8, S9, S10, S11).** Often, strategies *"depend on the question"* (p15), especially when it comes to planning and analyzing. In some cases, stakeholders *"have limited knowledge of what is possible"* (p39) and may not pose questions, making analysts formulate questions starting from pre-defined analyses or start with *"finding out patterns in the event log"* (p36). If present, questions are reported to be specific or broad, affecting how analysts plan their analysis, e.g., *"If that's a clear question, that is for us very nice, because we can just dive deep straight away. And then a pipeline is always to* *start exploring [...] If the question is not so clear [...] then we have a more fixed scheme"* (p39). In particular, specific questions seem easier to be mapped to the data and allow analysts to spend little or no time on hypotheses making, pattern discovery, and, even, process understanding. Indeed, in this setting, *"you know what to watch out for"* (p37) and *"you might not even go to a process model or process mining tool"* (p19). Ultimately, questions can be of many kinds, such as "statistical" or about "deviations", affecting what analyses can be done, e.g., *"If there's a simple statistic question [...] statistical analysis will be enough"* (p2).

R **Analyst's role (S2, S3, S12):** Among our interviewees, we could discern between "generalists", i.e., analysts who oversee all process mining stages, and "specialists", i.e., analysts specialized in specific steps such as creating event logs or dashboards based on stakeholders' needs, e.g., *"I'm very technical... most time I spend in event collections or setting up the data model, setting up the activities, instead of running the analytics"* (p24). Generalists often analyze the processes enacted within the company for which they work, thus gaining domain and data knowledge as part of their work experience, e.g., *"speaking about roles and capabilities, we own everything in our team [...] we know all the data, we create the event logs [...] we do all this ourselves, including the analysis"*. By contrast, specialists tend to work in teams. Thus, they may gather domain and data understanding from team members in closer contact with stakeholders or even work on the data without much domain knowledge, focusing on its structure. We also observed that analysts with a technical background in data engineering are more comfortable than others with strategies requiring deep data understanding and manipulation because with *"a relational database background, it's probably not as tricky. But if you come from, like, just a BI background, you probably aren't thinking in this way. And so, I think the aggregation can be the challenge"* (p14).

S **Availability of business stakeholders (S1, S2, S4, S8, S16).** Stakeholders' availability mainly affects problem and domain understanding and validation strategies. Indeed, analysts organize *"interactive sessions with the data owner to understand things"* (p34) and, if they are in close touch, also validate intermediate results, e.g., *"For me, it's not hard to talk during analysis [...] I am accustomed to getting some feedback. So my goal usually would be to do the analysis, get feedback, adjust it."* (p15). Stakeholders' availability can also influence planning. Indeed, if stakeholders are available, they can contribute to the generation of questions and hypotheses, e.g., *"I can do things with the guy in real-time and, usually, it triggers questions. And some questions I can try to understand live with the process mining tool, some questions will need to be worked on after the interactive session."* (p34). When access to stakeholders is limited, analysts tend to focus on time performance and conformance questions, as they are the *"most common perspectives that one can look at while doing process mining without having additional information about the context"* (p36).

T **Tools (S3, S9, S10, S11, S12).** Tools can influence data and process understanding, as well as what analyses can be planned and executed. Here, we distinguish between tools allowing (raw) data access and manipulation, tools supporting visual analyses and interactive exploration, and tools supporting custom analyses. Tools supporting raw data access and manipulation allow analysts to *"open the log file"* (p11) and *"understand, like the sort of data structure first"* (p14). Such tools seem to better support strategies, such as S11 and S12, where data manipulation and querying is needed, e.g., *"I wanted to delve in and do the subset analysis, but it was taking me a long time to go from generating a subset to producing the XES, to loading it and looking at it and say, 'oh, that's not quite right'. Then I went to [...] because it does all this processing out-of-the-box"* (p31). However, these tools are not necessarily process mining tools. Analysts report preferring tools that allow creating visualizations and animations when they engage in exploration or *"want to do a quick filter or get a quick idea"* (p19), e.g., *"I feel that the animation button [...] is key for me to understand, have better clarity on the identification of the process flow"* (p20). During planning, analysts choose which tools they will use. The possibility of creating custom analyses seems to be one of the criteria considered when making this choice, e.g., *"I wanted a way of taking very complicated data and just pushing a button to get my PowerPoint"* (p31).

5 Discussion and Conclusion

From interviews with 37 practitioners and academics working with process mining, we have derived 16 strategies related to the analysis stage and have examined recurring challenges and factors affecting their use in practice. These findings contribute to our understanding of the work practices of individual process analysts, with a specific focus on the analysis stage and its needs.

Our participants reported some challenges related to analysis strategies. One is the lack of domain knowledge, which affects both domain and data understanding (S2, S3) as well as the validation of results (S17), and is often connected to the limited S availability of business stakeholders (e.g., domain experts) and, also, the R analyst's role. While this challenge is renowned in the field [7,14], our participants stress the need for more concrete advice to work interactively and involve domain experts *"because only they know exactly what is there"* (p33). In this direction, our characterization of analysts into generalists and specialists could inspire research on how to organize work at the group level [3]), considering roles and skill sets of process mining teams. Indeed, the lack of clearly defined roles and responsibilities seem to be a cause for collaborative tensions [20].

Our results also show that the generation and refinement of Q analysis questions is a persistent challenge of the planning phase and also an influential factor for the analysis phase. Indeed, many participants reported that *"the analysis should always follow the question"* (p36) but that identifying the "right" question remains difficult. In their work, Emamjome et al. [9] have discovered that the thoroughness of question formulation as the first phase of process mining projects has improved over the years, indicating S increased interactions with business stakeholders. Still, our interviewees remark that *"specific methodologies that can help to define more effectively a research question"*(p36) are missing. Analysts rely on their experience to deal with different kinds of questions or lack thereof, sometimes using the pre-defined analyses available in some tools as a starting point. We think that pre-defined analyses could inspire the development of guidance for the less experienced. For example, they could be used as a starting point to develop questions catalogs or checklists adapted to specific use cases or event log features. This could be realized by collecting and refining frequently posed questions such as those identified by Mans et al. for healthcare processes [13] or the domain problems in [11]).

Regarding the analysis phase, our participants mentioned the lack of support of process mining tools for identifying causality being a challenge, limiting them in developing recommendations for solving stakeholders' problems (which relates to C.17 in [14]). The research community has picked up on this challenge, and recent work on causal machine learning provides promising results towards addressing it [2]. Still, more research is needed to support the identification of root causes and the consequent development of practical recommendations.

Last but not least, from our analysis, we have discovered some factors affecting the practical use of strategies (cf. Sect. 4.2). We believe that further investigation of these factors and, potentially, other individual and organizational factors could enrich our understanding of process mining practices and help us explain which strategies are suitable in a given context and why. Such an understanding could then be used to inform the development of concrete process mining guidance covering several aspects, e.g., analysis questions.

Limitations. Since they are based on retrospective interviews, our findings are subject to validity threats typical of interview studies [16]. First, the participants' behavior could have been influenced by the interviewer's presence and behavior (reactivity). To mitigate this risk, we asked questions using the wording defined in the interview guide, guaranteed anonymity of the answers, and recorded only the audio of the interviews, i.e., the videos of interviewees and interviewer were turned off. Second, the participants' answers could have been biased, e.g., by the study setting (respondent bias). To mitigate this risk, we developed our interview guide, considering questions about individual work practices, i.e., focusing on "general" actions, which did not require participants to name specific companies or share sensitive information. Also, participants had no prior knowledge of the study and the interview questions. Third, the data collection and analysis could have been biased by the researcher's personal opinions and interpretations (researcher bias). To mitigate this risk, we conducted all the interviews following the prepared guide, which had been pilot tested with researchers external to the author team as advised in [5] and we periodically met to review and discuss the coding process. In addition, the interviewer did not have any personal or working relationships with the participants.

Regarding the generalizability of our results, we would like to note that the set of strategies presented in this paper was derived from explicit statements of our participants, who may not have reflected upon all strategies they apply in practice. Thus, this set may not be complete. We cannot exclude that other strategies can emerge in different settings, for example, when anchored to different tasks, interviewing specific user groups, or focusing on certain application domains. Still, our sample included participants with varied backgrounds and experience levels who use process mining in different sectors, including healthcare, food processing, and insurance. Moreover, we considered only strategies that were reported by at least 4 participants, suggesting that most of them are relevant to different contexts.

We foresee several directions for future work. As a first direction, we plan to extend our findings by triangulating them with behavioral data. In particular, we aim to analyze the behavioral data collected during the process mining task to allow for a finer-grained characterization of strategies, including concrete steps and specific process mining techniques used to implement these steps. Such an analysis will also let us look deeper into factors and, potentially, explain how they affect the use of the strategies. In addition, we envision investigating strategies applied in other stages of a process mining project, such as the *extraction* and *data processing*. Last but not least, we hope that the strategies we identified in this paper will inspire research on developing actionable support for process mining practitioners.

Acknowledgment. We would like to thank our participants for taking the time to participate in the study and sharing their valuable experience. We also thank N. Soltani for her help with the transcriptions.

Funding. This work is part of the ProMiSE project, funded by the Swiss National Science Foundation under Grant No.: 200021_197032. P. Soffer was partly funded by the Israel Science Foundation as part of the Eco4PPM project, grant agreement 2005/21.

References

1. Van der Aalst, W.M.: Data science in action. In: Process Mining: Data Science in Action, pp. 3–23. Springer (2016)
2. Bozorgi, Z.D., Teinemaa, I., Dumas, M., La Rosa, M., Polyvyanyy, A.: Process mining meets causal machine learning: Discovering causal rules from event logs. In: 2nd Int. Conf. on Process Mining (ICPM). pp. 129–136 (2020). https://doi.org/10.1109/ICPM49681.2020.00028
3. vom Brocke, J., Jans, M., Mendling, J., Reijers, H.A.: A five-level framework for research on process mining. Bus. Inf. Syst. Eng. (2021). https://doi.org/10.1007/s12599-021-00718-8
4. Carlson, M.P., Bloom, I.: The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. Educ. Stud. Math. **58**(1), 45–75 (2005). https://doi.org/10.1007/s10649-005-0808-x
5. Chenail, R.: Interviewing the investigator: Strategies for addressing instrumentation and researcher bias in qualitative research. Qual. Rep. **16**(1), 255–262 (2011). https://doi.org/10.46743/2160-3715/2011.1051
6. De Leoni, M., Mannhardt, F.: Road traffic fine management process. Eindhoven University ofTechnology. Dataset (2015)
7. van Eck, M.L., Lu, X., Leemans, S.J., Van Der Aalst, W.M.: PM^2: A process mining project methodology. In: Int. Conf. on Advanced Information Systems Engineering (CAiSE). LNCS, vol. 9097, pp. 297–313. Springer (2015). https://doi.org/10.1007/978-3-319-19069-3_19

8. Eggers, J., Hein, A., Böhm, M., Krcmar, H.: No longer out of sight, no longer out of mind? how organizations engage with process mining-induced transparency to achieve increased process awareness. Bus. Inf. Sys. Eng. **63**(5), 491–510 (2021). https://doi.org/10.1007/s12599-021-00715-x
9. Emamjome, F., Andrews, R., ter Hofstede, A.H.M.: A case study lens on process mining in practice. In: On the Move to Meaningful Internet Systems: OTM 2019 Conferences. pp. 127–145. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33246-4_8
10. Grisold, T., Mendling, J., Otto, M., vom Brocke, J.: Adoption, use and management of process mining in practice. Bus. Process Manag. J. **27** (2020). https://doi.org/10.1108/BPMJ-03-2020-0112
11. Klinkmüller, C., Müller, R., Weber, I.: Mining process mining practices: An exploratory characterization of information needs in process analytics. In: Int. Conf. on Business Process Management (BPM). pp. 322–337. Springer (2019). https://doi.org/10.1007/978-3-030-26619-6_21
12. Koorn, J.J., Beerepoot, I., Dani, V.S., Lu, X., Weerd, I.v.d., Leopold, H., Reijers, H.A.: Bringing rigor to the qualitative evaluation of process mining findings: An analysis and a proposal. In: 3rd Int. Conf. on Process Mining (ICPM). pp. 120–127. IEEE (2021). https://doi.org/10.1109/ICPM53251.2021.9576877
13. Mans, R.S., van der Aalst, W.M., Vanwersch, R., Moleman, A.: Process mining in healthcare: Data challenges when answering frequently posed questions. In: Int. Work. on Knowledge Representation for Health Care, pp. 140–153. Springer (2012). https://doi.org/10.1007/978-3-642-36438-9_10
14. Martin, N., Fischer, D.A., Kerpedzhiev, G., Goel, K., Leemans, S.J., Röglinger, M., van der Aalst, W.M., Dumas, M., La Rosa, M., Wynn, M.T.: Opportunities and challenges for process mining in organisations: Results of a Delphi study. Bus. Inf. Syst. Eng. **63** (2021). https://doi.org/10.1007/s12599-021-00720-0
15. Nurcan, S., Etien, A., Kaabi, R., Zoukar, I., Rolland, C.: A strategy driven business process modelling approach. Business Process Management Journal (2005). https://doi.org/10.1108/14637150510630828
16. Padgett, D.K.: Qualitative methods in social work research, vol. 36. Sage (2016)
17. Ralph, P., Ali, N.b., Baltes, S., Bianculli, D., Diaz, J., Dittrich, Y., Ernst, N., Felderer, M., Feldt, R., Filieri, A., et al.: Empirical standards for software engineering research. arXiv preprint arXiv:2010.03525 (2021)
18. Saldaña, J.: The coding manual for qualitative researchers. Sage (2015)
19. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., Jinks, C.: Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant **52**(4), 1893–1907 (2018). https://doi.org/10.1007/s11135-017-0574-8
20. Syed, R., Leemans, S.J.J., Eden, R., Buijs, J.A.C.M.: Process mining adoption. In: BPM Forum. LNBIP, vol. 392, pp. 229–245. Springer (2020). https://doi.org/10.1007/978-3-030-58638-6_14
21. Thiede, M., Fuerstenau, D., Barquet, A.P.B.: How is process mining technology used by organizations? A systematic literature review of empirical studies. Bus. Process Manag. J. **24**(4), 900–922 (7 2018). https://doi.org/10.1108/BPMJ-06-2017-0148
22. Zerbato, F., Soffer, P., Weber, B.: Initial insights into exploratory process mining practices. In: BPM Forum. LNBIP, vol. 427, pp. 145–161. Springer (2021). https://doi.org/10.1007/978-3-030-85440-9_9