UNIVERSITY OF TARTU

Institute of Computer Science

Innovation and Technology Management Curriculum

Nadiia Danylyshyn

Analysis templates for identifying improvement opportunities using Apromore

Master's Thesis (20 ECTS)

Supervisors:

Fredrik Milani

Katsiaryna Lashkevich

Tartu 2022

Analysis templates for identifying improvement opportunities using Apromore

Abstract:

One of the crucial assets that organizations possess is their business processes, which require continuous improvements to stay efficient. For companies to remain efficient, it is necessary to analyze processes, identify and address improvement opportunities. However, currently, there are no comprehensive guidelines for process analysts that will allow them to identify improvement opportunities from the event logs. The aim of this research is to develop analysis templates for identifying process improvement opportunities using Apromore. To fulfill this aim, 22 improvement opportunities that can be found using manual analysis of processes in Apromore were discovered from the analysis of academic papers. In order to understand how these improvement opportunities can be detected, content analysis of 129 BPIC submission reports was performed in addition to eliciting instructions from the other materials. Based on these, 21 templates were developed and evaluated. Therefore, the contribution of this thesis is templates that will help students and junior process analysts to easier identify improvement opportunities in the processes by manual analysis of event logs in Apromore.

Keywords:

analysis template, improvement opportunity, process mining, process analysis, Apromore

CERCS: P170 Computer science, numerical analysis, systems, control

Analüüsivormid Apromore’i kasutamisel täiustusvõimaluste väljaselgitamiseks

Lühikokkuvõte:

Organisatsioonide üks kriitilise tähtsusega väärtusi on nende äriprotsess, mis nõuab tõhusana püsimiseks pidevat täiustamist. Selleks, et ettevõte püsiks tõhusana, tuleb analüüsida protsesse, välja selgitada täiustusvõimalused ja neid kasutada. Samas aga praegu ei ole olemas protsessianalüütikutele igakülgseid juhiseid, mille abil sündmuselogide põhjal täiustusvõimalusi välja selgitada. Selle uuringu eesmärk on välja töötada analüüsivormid Apromore’i kasutamisel täiustusvõimaluste väljaselgitamiseks. Selle eesmärgi saavutamiseks tehti akadeemiliste uurimustööde analüüsi põhjal kindlaks 22 täiustusvõimalust, mida on võimalik Apromore’i protsesside manuaalsel analüüsil leida. Selleks, et mõista, kuidas neid täiustusvõimalusi on võimalik tuvastada, viidi lisaks muudest materjalidest juhiste saamisele läbi 129 BPIC saatenõuderaporti sisuanalüüs. Nende põhjal töötati välja ja hinnati 21 vormi. Seega on käesoleva lõputöö panus need vormid, mis aitavad tudengitel ja noortel protsessianalüütikutel protsessides Apromore’i sündmuselogide manuaalse analüüsi teel hõlpsamini välja selgitada täiustusvõimalusi.

Võtmesõnad:

analüüsivorm, täiustusvõimalus, protsessikaeve, protsessianalüüs, Apromore

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Table of Contents

1. Introduction

Business processes are one of the most valuable assets of any organization. Continuous improvement of the business processes is required if the organization aims to be efficient (Dumas et al., 2013). The ability to quickly align processes with new technological, economic, organizational, and other changes allows companies to remain competitive (Davenport, 1993). For instance, the elimination of manual processes leads to faster customer service with a lower error rate, which triggers an increase in customer satisfaction (Lewis, 2021).

One of the most effective and widely recognized methodologies for continuous improvement of the company's process is Business Process Management (BPM) (Ongena and Ravesteyn, 2019). BPM is "a body of methods, techniques, and tools to identify, discover, analyze, redesign (improvement), execute and monitor business processes in order to optimize their performance" (Dumas et al., 2013). As BPM is characterized as a continuous process, it is often represented as the BPM lifecycle, which consists of the following phases: process identification, process discovery, process analysis, process redesign, process implementation, and process monitoring (Dumas et al., 2018). The process analysis phase aims at measuring process efficiency, identification of process improvement opportunities, investigating dependencies that influence processes, and others (Association of Business Process Management Professionals, 2009).

Process mining, a bridge between traditional BPM and data mining, is a data-driven approach that focuses on deriving process insights from event logs (Van Der Aalst, 2012). Among other applications, process mining allows analysts to find opportunities for process improvements from event data (Elkoumy and Dumas, 2022).

Current literature on the topic is mostly focused on developing new high-level approaches and frameworks on how to use process mining to find improvement opportunities in the specific processes of the specific industries (for example, Cho, Song, and Yoo (2014). As these methodologies and frameworks are usually abstract and are not industry-agnostic, as well as they do not account for a set of specific improvement opportunities, they are challenging to use by junior process analysts for the identification of process improvement opportunities. Some research papers propose detailed algorithms on how to identify concrete improvement opportunities (for instance, Awad, Zaki, and Di Francescomarino (2016), but they are not tailored for specific process mining tools thus, they require process analysts to have advanced knowledge and skills to apply proposed algorithms in practice. Currently, there are no standard guidelines, techniques, or methods developed that will be industry- and/or tool-agnostic, and that will provide detailed step-by-step instructions on what should be done with an event log to manually find different opportunities for process improvements and what patterns identify those opportunities. This thesis tries to contribute to fulfilling this research gap.

The thesis aims to develop analysis templates for identifying process improvement opportunities using Apromore. The following research questions (RQ) were formulated in order to achieve the aim:

- RQ1: *What improvement opportunities can be detected from event logs?*
- RQ2: *How can improvement opportunities be detected from event logs?*

To answer the first research question, an analysis of improvement opportunities outlined in several research papers was performed to extract improvement opportunities that can be detected from event logs and can be presented in the form of guidelines. To answer the second research question, content analysis of BPI Challenge submission reports alongside analysis of additional materials were performed that allowed to identify patterns for identification of different improvement opportunities. The contribution of the thesis is 21 templates that provide step-by-step instructions on how to find various improvement opportunities from the event log using Apromore. The templates were developed based on the outputs of research questions. The first version of the templates was evaluated using interview and survey methods. Based on the ideas collected during interviews, templates were improved. The target audience of the templates are people with basic to good knowledge of BPM and elementary to intermediate skills with Apromore, for example, students and junior process analysts. These users will benefit the most from the templates because it will be easier for them to understand what types of improvement opportunities can be identified from event logs, what is the meaning of each improvement opportunity, how to detect these improvement opportunities, and what are the redesign possibilities by using the templates.

The remainder of this thesis is structured as follows. Chapter 2 introduces the concepts that compose the theoretical background, such as business processes, business process management, and process mining. Chapter 3 provides an overview of related work and formulates the research gap. Chapter 4 describes the applied research methodology. In Chapter 5, results are presented alongside developed templates for visual process analysis. Templates evaluation and improvements are described in Chapter 6. Chapter 7 defines the study's limitations. Conclusions and implications for future research are drawn in Chapter 8.

2. Background

This section provides an overview of the central concepts used in the thesis. It covers the notion of business processes, followed by the definition of business process management and its lifecycle as well as a process improvement opportunity. It continues with the description of different aspects of process mining.

2.1. Business Processes

Within every enterprise, there are sets of activities that are structured in a logical way and done by specific actors and aim at achieving the goals of the enterprise. Such a series of steps are called business processes. In the literature, business processes are defined in different ways. According to Dumas et al. (2013), a business process is "a collection of inter-related events, activities, and decision points that involve a number of actors and objects, which collectively lead to an outcome that is of value to at least one customer". Hammer and Champy (1993) defined a business process as "a collection of activities that takes one or more inputs and creates an output that is of value to the customer". Laguna and Marklund (2018) provided the following definition of business process "a network of connected activities and buffers with well-defined boundaries and precedence relationships, which use resources to transform inputs into outputs for the purpose of satisfying customer requirements". From these definitions, it can be derived that business processes have a defined set of components, which are a set of related activities, inputs, events, and decisions that lead to the outcome that should bring value to the customer. Additionally, processes typically involve actors, which can be human beings or organizations, and objects, which can be equipment, products, electronic records, and others. (Dumas et al., 2018). The relationships between all the components are depicted in Figure 1.

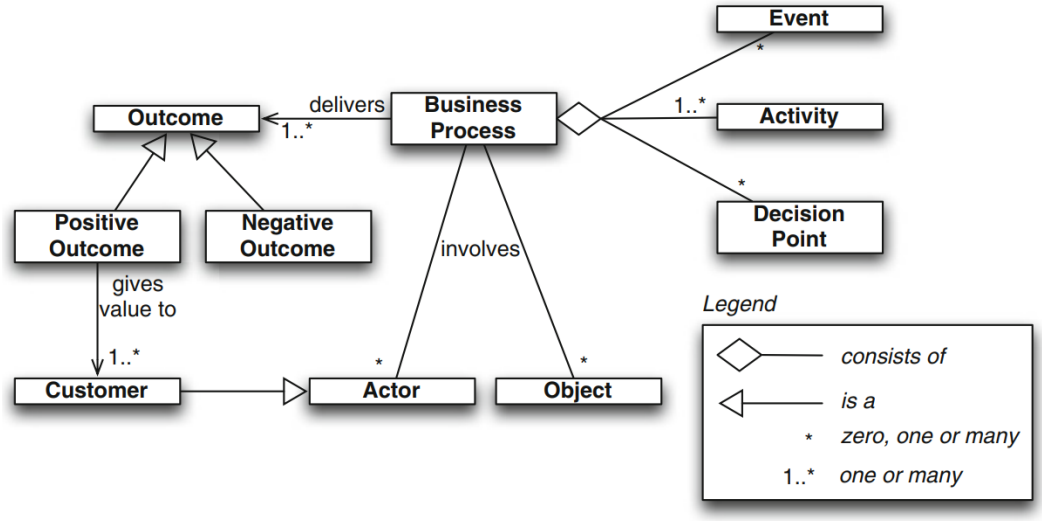


Figure 1. Relationships between components of business processes

Source: Dumas et al., 2018

It is critical for organizations to know their processes, manage them, and constantly look for improvement opportunities to be efficient and competitive (Dumas et al., 2018). The efficiency of the processes can be evaluated by to what extent their outcomes meet customer requirements.

2.2. Business Process Management

The discipline that deals with business processes is called Business Process Management (BPM), and it can be defined as "a body of methods, techniques and tools to identify, discover, analyze, redesign (improvement), execute and monitor business processes in order to optimize their performance" (Dumas et al., 2013). BPM lifecycle (see Figure 2) should be treated as a continuous cycle to keep processes efficient. They must be continuously reviewed and adapted to changes in the internal or external environment of the organization (Dumas et al., 2018).

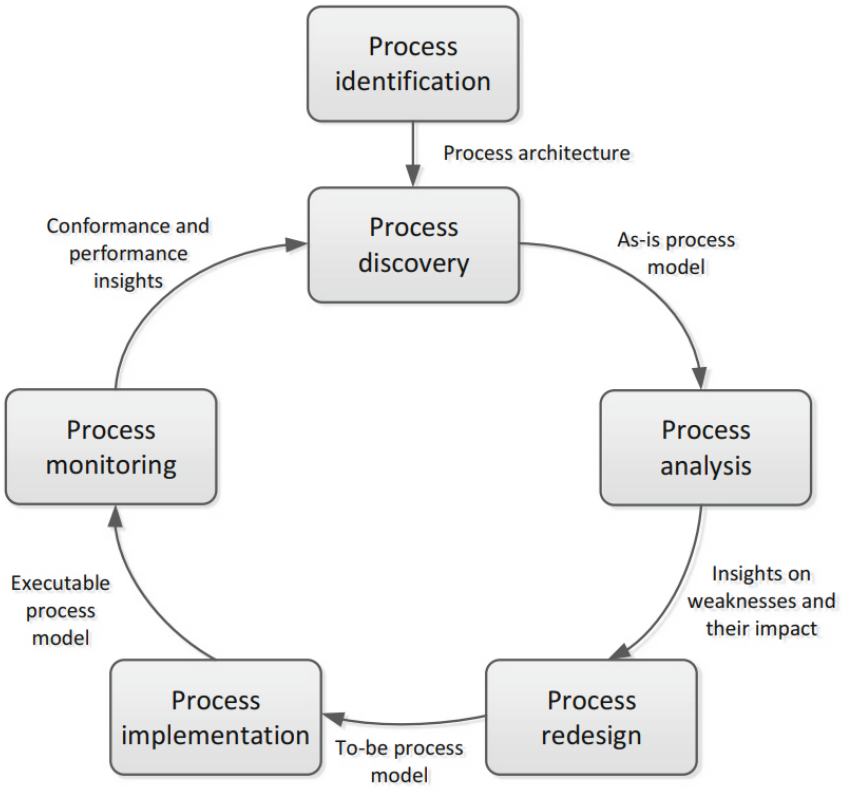


Figure 2. BPM lifecycle

Source: Dumas et al., 2018

The first phase of the BPM lifecycle, process identification, aims at defining process architecture that serves as an outlook of the main processes within the organization and their connections. Process discovery is the second step, the outcome of which should be identified current state and "as-is" process models. During the next step, process analysis, issues in the "as-is" models should be found, documented, and prioritized. In the following step, process redesign, change possibilities are provided and evaluated, "to-be" process model serves as the output of this phase. Process implementation is the fifth phase covering required changes to shift from the "as-is" to the "to-be" state. Process monitoring is the last step during which redesigned processes are assessed by analyzing the relevant data. (Dumas et al., 2018).

The aim of business process analysis is to measure process efficiency and effectiveness that can be communicated to decision-makers and other stakeholders to allow them to make management decisions (Zur Muehlen and Shapiro, 2015). Depending on the task, different objectives might be set for this phase, as well as different tools and techniques might be used to reach the objectives. Identifying process improvement opportunities is one of the objectives

of the process analysis phase, alongside researching and understanding the organization's strategy and objectives regarding the particular process, preparing stakeholder's analysis, analyzing business rules and dependencies that affect the process, investigating performance metrics that are utilized for process monitoring, among others (Association of Business Process Management Professionals, 2009). When it comes to the techniques that can be used for process analysis, they include interviewing the process' participants, observing the process in real-time, conducting information flow analysis, cost analysis, root-cause analysis, SWOT analysis (Association of Business Process Management Professionals, 2009).

In order to keep business processes efficient, it is needed to analyze them to detect and implement improvement opportunities continuously. It can be noticed when observing the process performance from the time perspective (see Figure 3). Most of the processes tend to lose their performance with time if there are no improvement efforts. Moreover, an organization can lose its position in the market without process improvements due to competition and high customer expectations. (Andersen, 2007)

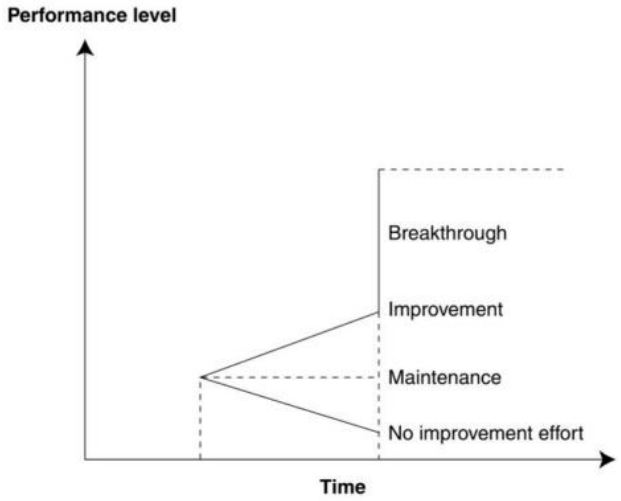


Figure 3. Performance level of processes along the time

Source: Andersen, 2007

From the BPM lifecycle perspective, the process analysis phase is a starting point for improvement. BPM offers a broad toolkit for process analysis focused on the improvement opportunities identification: waste analysis, root cause analysis, flow analysis, and others. (Dumas et al., 2018). Using these and other techniques, it is possible to identify weaknesses in the processes or areas for improvement. A business process improvement opportunity can be defined as a pattern in the process that has the potential to be improved.

BPM is becoming more and more critical for organizations because it provides insights about their business processes, and helps to manage and improve business processes by providing a straightforward approach.

2.3. Process Mining

Modern information systems collect valuable information in the form of event logs that can be translated into actions that will improve processes. Traditional BPM did not meet the expectations of professionals as it is mainly managed manually and involves building handmade models. Data mining techniques, such as regression or clustering, make it possible

to define patterns in data efficiently. However, they are not process-oriented, meaning that they are not able to derive needed insights from event logs. Consequently, process mining emerged intending to fill the gap between BPM and data mining and provide a toolkit for efficiently analyzing process data. (Van Der Aalst, 2012)

Process mining aims to collect enterprise data (event records) from IT systems to discover and improve actual business processes (Van Der Aalst et al., 2011). Event records capture the execution of a particular activity by a particular resource at a particular point in time (Buijs et al., 2011). A set of timestamped records forms an event log, where each event should have at least a case identifier, event class, and end timestamp (Dumas et al., 2013). Process mining techniques can be applied to understand the event logs better and extract insights. Process mining can be defined as "a family of techniques to analyze the performance and conformance of business processes based on event logs produced during their execution" (Dumas et al., 2013).

The literature distinguishes three different types of process mining. The first one is discovery, which uses the event log of the business process as an input for generating a process model. This type aims to create a model without any additional information. The second type is conformance checking, which involves a deep analysis of the process and checks if the process model corresponds to the event log. The last type is an enhancement, which extends or improves the process model with additional information from event logs. Therefore, enhancement can be applied with different goals, for instance, to better demonstrate the actual process or to extend the process model in order to get bottlenecks, frequencies, throughput times, and other insights. (Van Der Aalst, 2012; Van Der Aalst, 2016)

One of the ways to describe the typical process mining project lifecycle is the L* life-cycle model developed by Van Der Aalst (2011) (see Figure 4) based on the CRISP-DM, which depicts the life of a data mining project. This model consists of five stages. At the zero stage, the goal of the process mining project should be defined, and activities planned in detail. The first stage is about data extraction. Here all the relevant event data, objectives, and questions are taken from IT systems, specialists, and others. At the next stage, a control-flow model should be created, and the event log connected, so that event log records correspond to the model's activities. The third stage requires the development of the integrated process model, which means adding extra perspectives to the model. For instance, timestamps of the event log can be used to calculate the duration of activities. This integrated process model can be used for analyzing the current state or defying wastes. The last stage is about operational support activities (detect, predict, recommend). By taking "premortem" data, it is possible to, for example, predict flow time or propose recommendations. (Van Der Aalst, 2011)

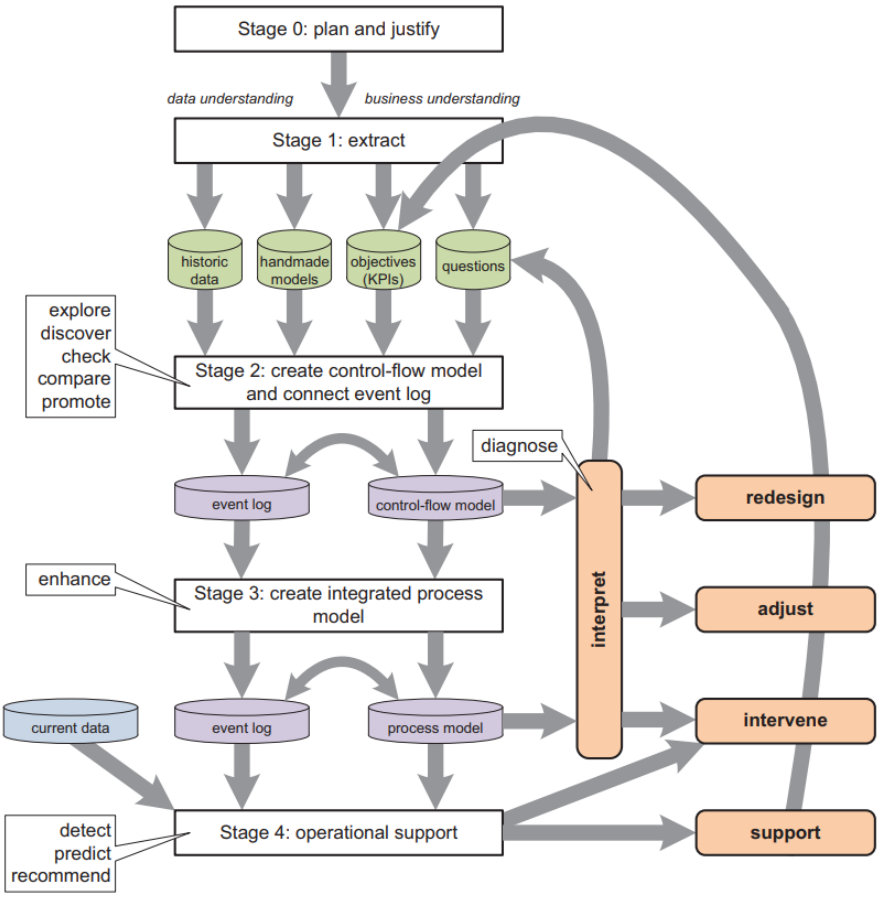


Figure 4. L* life-cycle model

Source: Van Der Aalst, 2011

Event logs can be analyzed from the following perspectives: control-flow, organizational, data, and performance (Van Der Aalst et al., 2007). Control-flow perspective is focused on the analysis of activities and the order in which they are performed; organizational perspective is concerned with the relationships between resources that are performing a particular process; data perspective deals with data objects that can be inputs and outputs of activities; performance perspective is focused on measuring different performance indicators of the processes (Rebuge and Ferreira, 2012). For each of the perspectives, a variety of techniques were developed. For instance, Heuristic Miner is an example of a control-flow technique (Weijters, Van Der Aalst, and De Medeiros, 2006), Organizational Model Miner – an organizational perspective technique (Song and Van der Aalst, 2008), Petri Net Models – a performance technique (Ou-Yang and Winarjo, 2011).

Successful implementation of process mining projects heavily depends on software tools utilized. Among the most widely used tools are ProM, Celonis, and Disco. ProM is an open-source software based on Java and developed by Wil Van der Aalst and his research group (Celik & Akçetin, 2018). ProM supports most process mining techniques and allows the development and use of plug-ins (Van der Aalst, 2009). Celonis and Disco are commercial products for process mining that have more user-friendly interfaces than ProM, making them better for beginners (Celik & Akçetin, 2018). In terms of functionality of these tools, several features are available in ProM but not yet developed for Celonis and Disco; there are several studies that provide a comparative analysis of these tools (e.g., Celik & Akçetin, 2018; Kebede

& Dumas, 2015). Another popular process mining tool is Apromore, which provides both commercial and academic licenses. Apromore has a collaborative repository workspace, which allows generating process maps and BPMN models, conducting performance analysis, compliance checking, and conformance checking, among other features ("Key Features", n.d.).

Considering the applications of process mining, healthcare, information, and communications technologies, manufacturing, education, and finance are the sectors for which most of the research studies are done. To provide several examples of use cases, they are customer satisfaction assessment, performance and bottleneck analysis, resource allocation, patterns analysis, event-based predictions, and optimization of spaghetti processes. (dos Santos Garcia et al., 2019)

Process mining tools are used more and more by organizations because these tools enable the derivation of more sophisticated grounds for decision-making (Wixom and Watson, 2010). Significantly, this concerns the opportunities for process performance improvements that are triggering the attention of the companies (Davenport and Spanyi, 2019). Such international corporations as BMW, SAP, Ernst & Young, and Airbus are using process mining (Market guide for process mining, 2019). The main focus areas of the process mining that these companies are utilizing are process improvements, auditing, processes automation, digital transformation, and IT operations (Grisold, Mendling, Otto, and vom Brocke, 2020).

To summarize, business processes form the foundation of organizations. In order to be competitive and profitable, organizations should continuously inspect their processes and find possibilities for improvements. One of the conventional ways to approach it is business process management. To overcome the limitations of BPM, process mining techniques are currently used more and more. Process mining aims to discover and improve business processes based on event logs.

3. Related Work

This section outlines an overview of related work concerning the topic of this thesis and determines the research gap.

The current scientific literature does not provide comprehensive guidelines that can be used for identifying improvement opportunities using process mining tools. Hence related works on broader topics will be considered. They can be clustered into three groups. The first one will discuss papers concerning methods and frameworks that can be utilized for defining process improvement opportunities in general. The second group is about studies that propose algorithms to tackle concrete improvement opportunities for a specific industry. The third one comprises the studies that describe different improvement opportunities but do not provide an answer on how to identify them from event logs.

To start with the first group, in their paper, Malinova, Gross, and Mendling (2022) investigate, classify, and define similarities and differences between methods that are used for process improvements. The authors used a multi-method research design that incorporated systematic literature review, process mining, and statistical analysis techniques. This study considered all of the methods proposed since the early 1990s when the topic of process improvements started to be actively researched and methods developed. Based on the literature review, 90 methods were elicited; on their basis, eight process models were generated in order to define similarities and differences. As the final step, logistic regression and hierarchical cluster analysis were performed to investigate correlation and co-occurrence, respectively. As a result of the research, a process improvement framework that categorizes 264 activities was developed, as well as the hypothesis that there are dissimilarities in proposed process improvement methods were supported. The paper of Malinova, Gross, and Mendling (2022) provides a comprehensive overview of methods that can be applied for process improvements and focuses on the whole spectrum of methods, from strategic benchmarking to training and support. It is done by collecting and analyzing methods analogous to this thesis's approach. The study provides a guideline for defining improvement opportunities, but it leaves the choice of methods and steps to the experience and intuition of business or process analysts. On the contrary, this thesis focuses only on data-driven identification of improvement opportunities using event logs, as well as it aims to develop templates that individuals can utilize regardless of their industry knowledge, experience in process mining and analysis, and skills in using process mining tools and Apromore in particular.

Most of the related work presents a developed framework for improvement opportunities identification using process mining in accompany with verifying this framework on a particular real-life case scenario. One such paper is the case study of Son et al. (2014) which considers how process mining can be applied to the manufacturing industry based on the real-life data of Samsung Electro-Mechanic company. Like many companies in the industry, Samsung Electro-Mechanic adopted a manufacturing execution system (MES) to be able to collect data about the process and control them. The paper's contribution is the developed process mining framework for analyzing manufacturing processes. To validate the practical use of the proposed framework, the authors of the case study analyzed the event log that was extracted from MES. The developed method is based on four stages: data preparation, pre-processing, process mining & analysis, and evaluation & interpretation. The third stage is divided into two perspectives: process and resource. Process perspective suggests using such process mining techniques as

process discovery, conformance checking, bottleneck analysis, whereas resource perspective includes resource network analysis, organizational model, and resource performance analysis. This paper provides the end-to-end high-level framework, which is industry-specific and process mining tool-agnostic. Whereas the thesis aims to create detailed instructions that can be applied for event logs from any industry and as the prerequisite for using the templates developed is to have prepared and pre-processed, they do not represent end-to-end processes.

Măruşter and van Beest (2009), in their study, developed a methodology for process redesign based on the combination of bottom-up process mining and simulation techniques. The proposed approach consists of four steps: define the relevant performance criteria, mine as-is process, simulate as-is and to-be processes, and compare as-is and to-be. The authors also tested the model on three case studies that are representatives of different industries, and have different processes, goals, and objectives. The first company was from the gas industry, and its business processes were analyzed; the second company represented governmental institutions and their workflow processes; the third organization was from the agriculture sector, and its decision support system was analyzed. The applied methodology showed different results on each step of the proposed approach for each of the case companies, which can be explained by the different nature of examined processes, and it witnesses the applicability of the methodology to real-life examples. This academic paper provides high-level steps for identification of some improvement opportunities and process redesign that can be used for any industry and with any process mining tool, but in the paper detailly presents the application of the methodology on three industries and using the ProM tool. At the same time, this thesis's scope is different as it does not cover detailed instructions for redesigning and attempts to create very detailed step-by-step instructions for most of the improvement opportunities that can be identified by manual analysis from event logs using Apromore.

In their research work, Gupta, Serebrenik, and Jalote (2017) developed a framework to identify improvement opportunities in the ticket handling process, which is a crucial part of software maintenance. This framework is tailored specifically for the ticket handling process, and the authors do not elaborate on its applicability to other types of processes. The applied research methodology is the following: firstly, the authors conducted interviews and surveys with practitioners to understand the common challenges and formulate the generic problem, then problem statements were classified, and an importance analysis survey was conducted with practitioners to select the area that required more support in identification of inefficiencies. The company that took part in this case study was Infosys, which is an international software company providing outsourcing services. As an outcome of the study, a framework for analyzing software repositories for IT tickets handling using process mining was proposed. The framework combines three steps: data extraction, event log formation, and multi-perspective process mining. The last step includes process discovery, bottleneck analysis, reopen analysis, loop analysis, conformance analysis, and social network analysis. The validity of this framework was tested on multiple open-source studies. This study develops a framework for the identification of several improvement opportunities in a specific process that is meant to be analyzed automatically using machine learning algorithms. At the same time, this thesis is focused on developing step-by-step instructions for manual analysis using a specific process mining tool that will cover most of the possible improvement opportunities.

Cho, Song, and Yoo (2014) researched the outpatient process of the healthcare industry and developed a methodology for its analysis using process mining. The approach is based on three

phases: data integration and preparation, data analysis, and discussion with industry experts. The data analysis step consists of process discovery, delta analysis, pattern analysis, performance analysis, and simulation. This methodology was validated on the case study done with Seoul National University Bundang Hospital. It is worth mentioning that the nature of processes in healthcare is often spaghetti-like, which makes it challenging to define patterns and analyze such data. The presented methodology gives suggestions on how to overcome this issue. This research paper provides an end-to-end industry-specific approach for event log analysis. On the contrary, this thesis presents industry-agnostic templates tailored to identify specific improvement opportunities.

All of the aforementioned papers provide methodologies for finding improvement opportunities using process mining. However, all of them are either designed specifically for a particular process in one industry or require deep domain knowledge to modify the method to another sector. Moreover, these studies provide only a general principle or skeleton of how the analysis should be conducted, leaving for the process analyst the decision on how exactly the method will be applied, which makes it difficult to use for not proficient in process analysis and data mining people. One more limitation of such studies is that for the data analysis step they provide a list of techniques that can be utilized, but they neglect the guidelines about in what cases what tool should be used, how precisely use those tools, and more importantly, what patters identify that improvement opportunity is found. Lastly, all of the papers present general approaches meaning that it is hard to know beforehand what improvement opportunity will be found; this can also be treated as a weakness.

Another group of related works concerns papers that propose algorithms for identifying concrete improvement opportunities using process mining. Ganesha, Dhanush, and SM (2017) proposed an algorithm on how to reduce waiting time by optimizing resource usage in the healthcare processes using Fuzzy Miner. One more example can be the work of Awad, Zaki, and Di Francescomarino (2016), where authors propose an approach on how to identify and resolve the problem of human multitasking at software development processes by means of serialization. The limitation of such papers is that they are focused only on one process improvement opportunity and often are dependent on some specific sector or even process.

One more category of papers is works that describe different improvement opportunities without connection to a specific process mining tool or industry. The research paper of Lashkevich and Milani (2020) provides an overview of 80 improvement opportunities, including types of improvement opportunities, their definitions, examples, and relevant redesign possibilities. However, the work does not cover the ways to find improvement opportunities. Sharma (2021) defines the list of business process wastes that can be determined from the event logs and answers on how to do it. The paper proposes algorithms for identifying wastes but does not give tool-specific instructions. Moreover, the author focuses on wastes only while improvement opportunities is a broader term. In their study, Reijers and Mansar (2005) present the best approaches to process redesign; this includes only a general idea about what type of problems could potentially be solved with these redesigns and a non-detailed description of redesigns. Even though the paper is not describing improvement opportunities, descriptions of the redesign possibilities can be used to identify for which improvement opportunities these redesigns can be applied. Nevertheless, it does not cover how they can be identified from event logs.

To conclude, the related work papers consider different methods, techniques, and frameworks that can be utilized for process improvement opportunities identification using process mining tools. However, they are not providing explicit industry- and/or process-agnostic guidelines that non-experienced process analysts and domain experts can use to identify improvement opportunities in the processes. This forms a research gap that this thesis aims to fulfill.

4. Methodology

In this section, the research methodology used to address the research questions is described. Figure 5 gives an outlook of the research process. The process is comprised of the following steps: define research questions, analyze RQ1, analyze RQ2, develop templates, evaluate and improve templates, and communicate results. Firstly, research questions were formulated to attempt to address the research gap. The second step focused on RQ1 analysis, where the list of improvement opportunities was composed based on an analysis of academic literature. The next step attempts to address RQ2, patterns for improvement opportunities were formulated relying on the results of content analysis. In the fourth step, outcomes of the second and third steps were used to develop the templates for improvement opportunities identification from event logs using Apromore. In the next step, templates were evaluated using interviews and survey analysis, and then templates were revised based on elicited suggestions. As the last step, improved templates were communicated by publishing them online.

Steps	1. Define research questions	2. Analyze RQ1	3. Analyze RQ2	4. Develop templates	5. Evaluate and improve templates	6. Communicate templates
Inputs	Research Gap	• RQ1 • Lashkevich and Milani (2020) • Sharma (2021) • Reijers and Mansar (2005)	• RQ2 • 2011-2020 BPIC submission reports • Dumas (2021a) • Dumas (2021b) • List of improvement opportunities	List of improvement opportunities Patterns for improvement opportunities identification	21 templates for identification of improvement opportunities using Apromore	Improved templates
Activities	1) Formulate research questions	1) Elicit all the improvement opportunities from reasearch papers 2) Evaluate and list improvement opportunities that can be identified using process mining tools	1) Conduct content analysis of BPIC submissions to define possible patterns for improvement opportunities identification 2) Extract instructions for improvement opportunities identification from Dumas (2021a), and Dumas (2021b)	1) Create structure of the templates 2) Develop templates	1) Define interview procedure 2) Prepare pre- and post-study questions, survey 3) Find interviewees and conduct interviews 4) Analyze interview results 5) Elicit possible impovements for the templates 6) Improve templates	1) Publish thesis and templates
Outputs	• RQ1 • RQ2	List of improvement opportunities that can be identified using Apromore	Patterns for improvement opportunities identification	21 templates for identification of improvement opportunities using Apromore	Improved templates	Templates are shared

Figure 5. Research methodology

The following subsections describe each step of the research methodology in the corresponding order.

4.1. Define Research Questions

This subsection describes the formulated research questions, and it covers the first step of the research methodology (see Figure 5).

The current work aims to develop templates for identification process improvement opportunities using Apromore. The templates are targeted to be tool-specific because this is a precondition for creating detailed step-by-step instructions. As the functionalities of the process mining tools differ, tool-agnostic guidelines would provide only a general approach but not precise instructions. Since the target audience of the templates is students and junior process analysts with limited experience with improvement opportunities identification, having an exhaustive guideline is more beneficial for them. In this research, the process mining tool Apromore is selected due to several reasons. Firstly, Apromore is a commercial product which is preferred to use by organizations over open-source solutions, because commercial solutions are more reliable and stable in terms of functionalities and performance. Secondly, unlike other widely used process mining tools, Apromore is available for free in full version for academic

purposes, which allows the author to utilize all the functionality needed to develop templates and it will be more convenient for students to use the templates later.

To achieve the aim, the following research questions were defined:

RQ1: *What improvement opportunities can be detected from event logs?* It refers to the identification of what improvement opportunities can be found using manual analysis of processes by examining the outputs of process mining tools (for example, process maps, BPMN models, dashboards).

RQ2: *How can improvement opportunities be detected from event logs?* This research question aims to determine the patterns for improvement opportunities identification. This will allow us to formulate the instructions that process analysts should follow to identify improvement opportunities using the process mining tool Apromore.

4.2. Analyse RQ1

In this subsection, methodology for RQ1 analysis is presented. The subsection refers to the second step of Figure 5.

The aim of this step is to analyze academic literature in order to discover improvement opportunities that can be identified from the event logs. For that, firstly, data sources were defined. Secondly, the improvement opportunities mentioned in the selected literature were gathered. The third phase was the elimination of those improvement opportunities that could not be identified from an event log or identification of which is too sophisticated to be described via the step-by-step instructions.

For eliciting the initial set of improvement opportunities, three data sources were selected: Lashkevich and Milani (2020), Sharma (2021), and Reijers and Mansar (2005). Each of these papers provides descriptions of several improvement opportunities. The main data source was Lashkevich and Milani (2020) because this is the systematic literature review (SLR) study devoted to improvement opportunities. Research papers of Sharma (2021), and Reijers and Mansar (2005) were used as additional sources because Sharma (2021) focused only on the wastes and Reijers and Mansar (2005) presented several redesign possibilities.

The next phase after defining the data sources was to list all the improvement opportunities covered in those research papers. Lashkevich and Milani (2020) paper describes 80 improvement opportunities, the work of Sharma (2021) contains 11 improvement opportunities and Reijers and Mansar (2005) provides an overview of 9 groups of redesign possibilities, descriptions of the redesigns incorporate the descriptions of the improvement opportunities they can be applied to.

The final phase was to compose the final list. In order to do it, improvement opportunities were eliminated based on the following criteria:

- **It is a duplicate.** If the same improvement opportunity was presented in two or more papers, only one was included;
- **It is impossible to identify the improvement opportunity using manual analysis of event logs in Apromore.** If for improvement opportunity identification, information that is usually not included into event log is required or identification is not possible with the default functionality of Apromore, such improvement opportunity was

excluded. For instance, improvement opportunity “lack of expertise”, which defines human resources with deficient skills and/or knowledge for performing the process; usually such information as knowledge sufficiency is not included in the event log as it is rather a subjective evaluation.

- **It is difficult to identify the improvement opportunity using manual analysis of event logs in Apromore.** When the identification of the improvement opportunity does not follow a straightforward pattern but is a rather exploratory process, it is difficult to describe it in the form of step-by-step guidelines. For example, improvement opportunity “unavailability of input in time”, which describes the situation when resources that are inputs for a certain activity are not available at the moment they are needed; with the basic functionality of Apromore, by manually analyzing a generated process map, it is difficult to determine and measure the situations when input for some activity was unavailable.

4.3. Analyse RQ2

This subsection presents the process of RQ2 analysis. This subsection corresponds to the third step of the research methodology (see Figure 5).

To address the second research question, *How can improvement opportunities be detected from event logs?*, the author followed the content analysis approach because it allows deriving deeper insights of the investigated works to get exhaustive input for developing templates for visual process analysis. For that, guidelines for qualitative content analysis with inductive category development were adopted and followed (Mayring, 2000). Figure 6 presents the chosen research methodology algorithm.

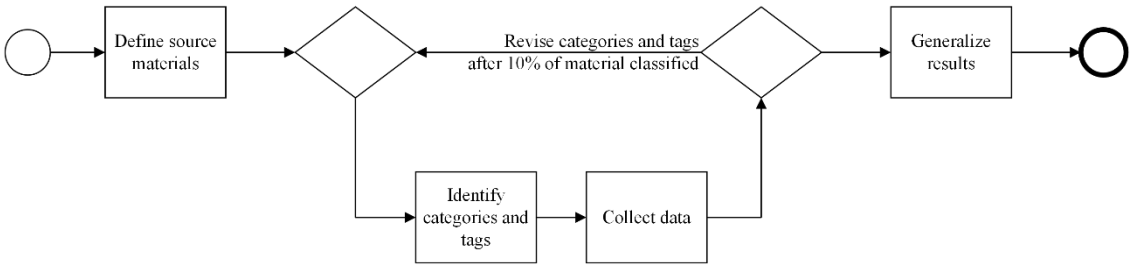


Figure 6. Content analysis approach

Source: compiled by the author, based on Mayring, 2000

The data source materials used for content analysis was BPI Challenge (BPIC) submission reports. In addition, materials developed by Dumas (2021a) and Dumas (2021b) were included in the study but as they are already presented in the form of instructions, content analysis was not applied, only improvement opportunities and corresponding instructions were elicited and analyzed.

BPIC is an annual contest for students, academics, and industry professionals that challenges analyzing real-life event logs using any tools and methods for answering questions that process owners proposed. At this time, BPIC was conducted ten times (2011-2020), and overall, there are 129 submitted reports published. This data source was chosen because each year, participants are working on data from different industries and answering different sets of questions, which allows collecting heterogeneous data and then create industry-agnostic

templates. The second reason is that participants are not limited in tools, methods, techniques to use, which provides a better overview of possible ways to tackle the problem. Thirdly, the audience of the BPIC is broad and has diverse experience; students, researchers, and people from industry from all over the world are eligible to participate; this factor is also essential for collecting data. The fourth reason is that per each provided event-log and set of questions, there are at least several submissions, which allows comparting the ways and approaches that can be used for identification of improvement opportunities.

Dumas (2021b) created templates for performance mining, and Dumas (2021a) describes templates for conformance checking, and variant analysis. These templates are not precisely designed for improvement opportunities identification, they cover different topics of performance mining, conformance checking, and variant analysis, from which it is possible to define some improvement opportunities. It is worth mentioning that templates' structure contains only the name of the category (for example, Workload & demand analysis), high-level steps that should be taken as well as more detailed step-by-step instruction. Also, this data source was chosen because templates are designed for Apromore which allows to identify best practices for improvement opportunities identification using this specific process mining tool.

The next step after defining source materials was identifying the categories and tags. They were derived from research questions. As an additional category, the step-by-step description of how authors identified a certain improvement opportunity was derived from the papers to have a better overview of the sequence of steps applied.

Then BPIC submissions were examined; for each improvement opportunity found in each paper, a set of tags and comments was assigned. As there was a threat that the initial set of categories and tags may be incomplete, after analyzing 10% of the material, categories and tags were revised and modified. Table 1 summarizes the final structure of categories and tags that were used.

Table 1. Categories and tags for content analysis

Category	Tag
General information	• ID of the paper • Title • Author(s) • Publication year • Industry of the event log • Process name
RQ1	• Improvement opportunity type
RQ2	• Event log attributes used (for example, activities, resources) • Process mining tool used • Features of process mining tool used • Data manipulations applied (for example, filtering) • Other tools used (for example, Excel) • Patterns that indicate improvement opportunity

Step-by-step description	• Description of the steps applied

After data from all the BPIC submissions were collected, results were grouped based on improvement opportunity. Some of the improvement opportunities were excluded as they were process-specific, and generalizations were not possible (for instance, not user-friendly website). Moreover, similar improvements were considered as one type of improvement (for example, waiting time, waiting time caused by internal resources, and waiting for external input from customer).

This was followed by generalizing findings for each type of improvement opportunity from the perspective of what is needed to define improvement (event log attributes, process mining tool functionality, step-by-step descriptions). It is worth noting that usually for one improvement opportunity generalization of instructions, step-by-step descriptions from several industries or processes were available and therefore taken into consideration. This makes the findings' generalizations and afterwards the templates developed industry-agnostic, so that they can be used for different types of event logs.

When it comes to the analysis of Dumas (2021a) and Dumas (2021b), templates presented in these materials were investigated to define improvement opportunities. Then detailed step-by-step instructions were extracted for each of the identified improvement opportunities.

As guidelines for not all the improvement opportunities elicited in RQ1 were available in BPIC submission reports, Dumas (2021a), and Dumas (2021b), the rest of the guidelines were created based on the patterns from elicited guidelines.

4.3.1. Content Analysis on the Example of Bottleneck

This subsubsection provides an example of how content analysis was done for the bottleneck improvement opportunity. After all the BPIC submission reports were tagged and categorized, grouping by improvement opportunities was done, as it was described above. Table 2 summarizes the main insights (data used, and step-by-step description) obtained from the BPIC papers. The full extract contained the whole set of aforementioned tags.

Table 2. Summary of content analysis for the bottleneck improvement opportunity

ID	Submission title	Data used	Step-by-step description
4	Mining Process Performance from Event Logs (Adriansyah & Buijs, 2012)	• Activity • Start timestamp • End timestamp	1. Make performance analysis 2. Identify activities that have severely long average duration 3. Calculate frequency, min duration, max duration, average duration and 99% confidence interval for all cases and separately for cases with different end points
70	Analysis and Prediction of Undesired Outcomes (Brils, Elsen, Priester & Slooff, 2018)	• Activity • Start timestamp • End timestamp	1. Filter cases with late payments (if needed, in Python add a case attribute for late cases) 2. Generate statistics on the duration of cases 3. Compare the process map with

			cases with late payments and a general process map
89	Automated Machine Learning in a Process Mining Context (Kas, Post & Wiewel, 2020)	• Activity • Start timestamp • End timestamp	1. Use the replay function in the Disco
94	Business Process Intelligence Challenge 2020: Analysis and evaluation of a travel process (Klein, et al., 2020)	• Activity • Start timestamp • End timestamp	1. Calculate the duration of an activity for each case 2. Calculate statistics over all cases including mean, median, min, max and std 3. Select activities with the highest bottlenecks
98	Process Mining in Finance Sector: Trip Compensation Process in TU/e Business Process Intelligence Challenge 2020 (Khodaev, Kalugin, Korneeva, Savintseva & Balashova, 2020)	• Activity • Start timestamp • End timestamp	1. Calculate mean and median and total transition time between the statuses 2. Define transition with highest total time
103	Business Process Intelligence Challenge 2020: Investigation of Business Trips Arrangement Process at the Eindhoven University of Technology (TU/e) (Pakileva, Skvortsova, Zakoryuchkin, Tsaplin, & Zarubin, 2020)	• Activity • Start timestamp • End timestamp	1. Create a filter based on case duration to filter out the cases duration of which exceeds the average duration of cases in this process significantly 2. Create a model 3. Find the longest transition and % of cases with this duration

To sum up, most of the BPIC submissions described the procedure of bottleneck identification as the following:

- Find activities with high processing time/high waiting times between activities;
- Identify the highest processing/waiting time.

In his work, Dumas (2021a) provided guidelines for bottleneck variant analysis. His approach was the following (Dumas, 2021a):

- Filter event log to keep one process variant. Repeat for the second variant;
- Compare generated process maps to identify arcs with high waiting time;
- Compare activities' processing times;
- From a resource perspective, find handoffs;
- Visually define bottlenecks using multi-log animation.

Dumas (2021b) presented a slightly different approach in the analysis template:

- Using a process map, find bottlenecks from activities perspective;
- Using a process map, find bottlenecks from resources perspective;
- In the slider, select slowest or fastest nodes/arcs based on the need;
- Using dashboard, define processing times of activities;
- Detect bottlenecks using animation plug-in.

In terms of data needed to identify this improvement opportunity, all the solutions presented in BPIC reports used activities, start timestamps, and end timestamps. In addition to those, Dumas (2021a), and Dumas (2021b) used resources. This analysis was later used to develop Bottleneck template.

4.4. Develop Templates

In this subsection the procedure of templates development is outlined, which refers to the fifth step of the research methodology described in Figure 5.

Outputs from the last two subsections were used to develop the templates. Based on the patterns for improvement opportunities identification derived from BPIC submissions and other information available about improvements in the research papers used in Subsection 4.2, templates were developed. To provide the end user with enough information about improvement opportunities templates also cover definition, examples, minimum data needed, redesign possibilities and references in addition to the guideline itself. Overall, 21 templates were created that covered 22 improvement opportunities selected. High human resource utilization and Low human resource utilization templates were combined into one template because of high similarity of steps. Some of the templates also gave the end user two options on how he/she can identify a particular improvement opportunity, for example, using process map visualization and dashboard. Additionally, templates were grouped for easier navigation.

4.4.1. Templates Development on the Example of Bottleneck

In this subsubsection the process of template development is described, taking as an example bottleneck.

Information for Bottleneck template development was derived from five sources. Based on the data from Lashkevich and Milani (2020) and Sharma (2021), definition of bottleneck was formulated. Examples of bottlenecks were composed based on these papers as well as other research papers provided in references. Redesign possibilities were taken from Lashkevich and Milani (2020).

As it was presented in subsubsection 4.3.1., guideline section was developed based on the information extracted from BPIC submissions, Dumas (2021a), and Dumas (2021b). Taking into consideration all the sources, the author proposed the next steps to detect bottlenecks:

- Generate process map;
- Using activities perspective, find activity bottlenecks, waiting time bottlenecks, resource-capacity bottlenecks;
- Switch to resources perspective and identify resource bottlenecks, waiting time bottlenecks from resources perspective.

In addition to the steps' description, screenshots from Apromore and explanations were provided to help the end user to understand this improvement opportunity better and quicky identify it. The final version of the Bottleneck template can be found in Appendix I.

4.5. Evaluate and Improve Templates

This subsection corresponds to the fifth step of Figure 5, and it describes the methodology chosen to evaluate the first version of the templates, to list suggestions that can be used to improve templates, prioritize these suggestions and implement them.

In general, evaluation was done by conducting interviews, during which participants were asked to use several templates to identify improvement opportunities in the provided event log. After templates testing, interviewees were asked to answer several questions and complete a survey. Later, all the outputs of the interviews were analyzed, and templates were improved based on the results of analysis.

The target audience of the templates are students and junior process analysts that have basic to good knowledge of BPM and elementary to intermediate Apromore skills. Considering this, the selected target audience for templates evaluation were university students that finished Business Process Management and/or Business Process Mining courses and/or with BPM and Apromore skills obtained in other ways.

In total, 12 participants took part in the study. They were divided into two groups based on the level of experience with BPM and Apromore:

- Group 1: 6 people have basic knowledge of BPM and elementary Apromore skills (students of "Business Process Management" course at the University of Tartu, and/or students with less than 1 year experience with BPM and Apromore);
- Group 2: 6 people have good knowledge of BPM and intermediate Apromore skills (students of "Business Process Mining" course at the University of Tartu, and/or students with more than 1 year experience with BPM and Apromore).

To evaluate the developed templates, the Technology Acceptance Model was adapted and used. This model states that users have the intention to use the system when they think that it is useful and easy to use. According to it, the model describes relationships between three elements: perceived usefulness, perceived ease of use, and intention to use. (Mertens, Pugliese, and Recker, 2017)

Intention to use was excluded from the model and was not evaluated because study participants were students and some of them will not be extensively involved in identification of improvement opportunities using process mining tools in the future. As this parameter was not relevant for all interviewees, it was eliminated in order to prevent biased results.

Moreover, additional dimension, possible improvements, was included in the study because the goal is not only to assess them but also to understand in what ways templates could be improved. So, the formulated evaluation goals (EG) of the study were:

- EG1: Assessment of the usefulness of the templates for identifying improvement opportunities using Apromore;
- EG2: Assessment of the understandability (ease of use) of the templates for identifying improvement opportunities using Apromore;

- EG3: Determination of possible improvements of the templates.

Overall, the study procedure followed the next steps (see Supplementary Material A):

1) Describing the aim of the study to the interviewee;
2) Asking participant about his/her experience with BPM;
3) Introducing templates to the interviewee;
4) Asking participant to complete study tasks;
5) Asking interviewee post-interview questions;
6) Asking interviewee to complete the post-survey.

The study task was to find three improvement opportunities using the file with templates and event log provided. Not all the templates were tested during the study, because:

- The aim of the study was to evaluate overall usefulness and understandability of templates but not usefulness and understandability of each of the templates;
- Testing templates is a time-consuming task, interviews were limited by 1 hour;
- As information can be perceived differently by different people, it was important to test the same template several times in both the groups.

Taking into account the constraints described above, all the templates were clustered based on similarity of guidelines. For instance, small activities and large activities templates were grouped into one cluster, because from the perspective of guidelines section they are akin. The differences are only that in the first case, the end user should visually identify activities with the shortest processing time, and in the second case – activities with the longest processing time. Also, in the small activities template, there is one additional step about checking for manual activities.

Then, one template was selected from each cluster considering the presence of particular improvement opportunity in the event log. As it is difficult to have a single log that contains all the improvement opportunities, two logs were used: Loan Origination and Refund Process. Loan Origination is the real-life event log used during 2017 BPIC; Refund Process is a synthetic log. Table 3 captures clusters of the templates, template selected for the study and corresponding event log.

Table 3. Templates clustering for evaluation phase

Cluster ID	Cluster composition	Selected template	Event log
Cl1	Small activities Large activities	Large activities	Loan Origination
Cl2	Activity variants	Activity variants	Loan Origination
Cl3	Similar activities Independent sequential activities Controls performed by internal resources	Similar activities	Refund Process
Cl4	Frequent handovers Ping pong behavior	Frequent handovers	Refund Process
Cl5	High/low resource utilization	High/low resource utilization	Refund Process

Cl6	Rework Knock-out Workaround	Rework	Loan Origination
C7	Highest waiting times Cases with the highest waiting times Bottleneck Manual time-consuming fragment	Bottleneck	Loan Origination
C8	Manual process High process complexity Similar process variants	High process complexity	Loan Origination
C9	Overprocessing Overproduction	Overprocessing	Loan Origination

In total, 9 templates were tested during the study. As the author was asking participants to find three types of improvement opportunities per interview, these 9 templates were grouped by 3 groups in a way that in one group could be templates tested by the same event log in order not to ask interviewees to upload and learn two event logs. Taking into account the assumption that when working with the first template, it takes more time for the participant to get familiar with the templates and event log that bring more comments regarding understandability and ease of use, two orders in which the author asked participant to test templates were used to overcome this bias. Table 4 provides the final templates grouping.

Table 4. Groups of templates used during the study

Group number	Templates Group 1	Templates Group 2	Templates Group 3
List of templates selected (order 1)	• Large activities • Activity variants • High process complexity	• Rework • Bottleneck • Overprocessing	• Similar activities • Frequent handovers • High/low resource utilization
Regrouped templates (order 2)	• Activity variants • High process complexity • Large activities	• Bottleneck • Overprocessing • Rework	• Frequent handovers • High/low resource utilization • Similar activities
Event log	Loan Origination	Loan Origination	Refund Process

Participants' questions, comments, and suggestions mentioned during the templates testing were collected. In addition, in order to get general comments, post-study questions were asked after the testing phase. These questions were focused on understanding what sections or aspects

of the templates were the most and the least useful, and helpful; are there any parts that interviewees found redundant; what section or aspect could be improved and how.

After discussing post-study questions, participants were asked to complete a post-survey. The survey was created in MS Forms to collect responses online. It consisted of three sections: questions about Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and demographic. A detailed list of PU and PEOU questions is provided in Table 5.

Table 5. PU and PEOU post-survey questions

ID	Question
PU1	Using Templates would enable me to identify types of improvement opportunities more quickly
PU2	Using Templates would make it easier to understand the meaning of improvement opportunities (based on definition and examples sections)
PU3	Using Templates would enable me to define what data I need to identify improvement opportunities more quickly
PU4	Using Templates would enable me to identify improvement opportunities more quickly
PU5	Using Templates would enhance my effectiveness in identifying improvement opportunities
PU6	Using Templates would enhance my effectiveness in identifying relevant redesign possibility for each improvement opportunity
PU7	Using Templates would enable me to find relevant academic papers for each opportunity more quickly
PU8	I would find Templates overall useful
PEOU1	Learning how to use Templates would be easy for me
PEOU2	I would find it easy to use Templates to identify improvement opportunities and redesigns
PEOU3	Structure of Templates is clear and understandable
PEOU4	It would be easy for me to become skillful at using Templates
PEOU5	I would find Templates easy to use

PU and PEOU questions were collected using a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). After responses were collected, mean, standard deviation, median, and variance values were calculated and analyzed. It should be mentioned that post-survey was used to evaluate EG1 (assessment of the usefulness of templates), and EG2 (assessment of the ease of use of templates) only, it was not used to support EG3 (determination of templates' improvements).

Demographic questions covered gender, age, years and type of experience with BPM and process mining tools participant have used. Results of demographic findings are presented in Supplementary Material A.

Table 6 gives an overview of the interviewees, their experience and templates that they were testing.

Table 6. Participants

ID	Experience with BPM (in years)	Group (1 – elementary skills, 2 – advanced skills)	Templates group (see Table 4)	Order of the templates (see Table 4)
I01	0,5	1	3	1
I02	0,5	1	1	1
I03	3	2	2	1
I04	0,5	1	2	1
I05	5,5	2	1	1
I06	0,5	1	3	2
I07	3	2	3	1
I08	1	2	2	2
I09	1,5	1	1	2
I10	3	2	3	2
I11	2	2	1	2
I12	0,8	1	2	2

All interviews, except the interview with I03, were conducted online using MS Teams or Skype. Interview with I03 was conducted in person with audio and screen recording. Overall, the duration of each interview (including introduction, templates, testing, post-study questions, and post-survey) was approximately one hour. Templates testing and post-study questions parts of the interview were recorded. The duration of the recorded part varied from 17 to 41 minutes.

All the recordings were transcribed using the in-built transcription function of OneDrive. Automatic transcription files were reviewed, and transcriptions were manually corrected, where it was needed.

Then interviews were analyzed using affinity diagram method. From the interviews, participants' phrases regarding templates (questions, suggestions, comments) that later were clustered on code-, theme-, and evaluation goal-levels were elicited. As a result, 21 codes were classified by 5 themes, which were divided between 3 evaluation goals (EG) (see Table 7). In this study, EG1 corresponds to the first goal of the study (assessment of the usefulness of templates), EG2 – to the second goal (assessment of the ease of use of templates), and EG3 – to the mentioned suggestions for templates improvements.

Table 7. Codes, themes, and evaluation goals by which quotes were classified

Codes	Themes	Evaluation goals
• Overall templates • Table of contents • Type of IO section • Definition section • Examples section • Minimum data needed section • Guideline section • Redesign possibilities section • References section	Content	EG1: assess usefulness of the templates
• Usefulness	Usefulness	

• Not useful • Incompleteness of information		
• Understandability of the content • Understandability of the structure • Issue with understandability	Understandability	EG2: assess ease of use of the templates
• Ease of use • Issue during use	Ease of use	
• Content improvement • Structure improvement • Screenshots improvement • Adding additional section	Improvement	EG3: determine possible improvements of the templates

The next step after clustering quotes was to analyze results to derive findings and list possible improvements. After listing the suggestions, they were prioritized. Prioritization was done by three categories: *must*, *could*, and *won't*. *Must* priority was used for the ideas that were mentioned by more than one person and/or those that require low effort to implement and/or if the implementation will bring a lot of value to end users (for example, adding an introduction was prioritized as *must* because this idea was mentioned by five interviewees and this will bring value to users as there can be find useful information that will make work with templates easier). *Could* label was used for suggestions that require medium effort to implement and/or will not bring much value to the end users. *Won't* type of prioritization was applied to those ideas that are not feasible to implement during this study due to time constraints (improvements that need medium or high effort to be implemented) and/or those that are out of the scope of current research and/or those that require and/or ideas that will not bring value to the end users.

All of the suggestions marked as *must* were implemented as well as some of suggestions from *could* category, based on time availability. The rest of the suggestions are subject to review and implementation as a part of future work.

4.6. Communicate Templates

The current subsection is devoted to the sixth step of the research methodology (see Figure 5), and it outlines the way templates will be communicated.

The file with the complete set of templates is provided as a Supplementary Material B. Everybody interested will have the opportunity to access them.

5. Results

This section presents the results of analysis and templates development. The first subsection describes the results of RQ1 analysis. The second subsection outlines the results of RQ2 analysis. The next subsection provides a general overview of the developed templates including the data sources based on which each template was created and an overview of the structure of the file with templates as well as templates themselves. The fourth subsection detailly presents the Bottleneck template as an example.

5.1. Results of RQ1 Analysis

After analysis of the improvement opportunities described in Lashkevich and Milani (2020), Sharma (2021), and Reijers and Mansar (2005) and applying the exclusion criteria described in subsection 4.2., the list of 22 improvement opportunities that can be identified by manual analysis of event logs in Apromore was compiled. These improvement opportunities are listed in Table 8 alongside their definitions.

Table 8. Improvement opportunities and their definitions

Improvement opportunity	Definition
Small activities	An activity that has one or several functional procedures and short processing time
Large activities	An activity that has many functional procedures and long processing time
Activity variants	An activity that has abnormal variation in process time and/or waiting time based on the resource type or case attribute
Similar activities	Two or more activities that have similar procedures
High resource utilization	High percentage of non-occupied resources in a process over the total number of resources
Low resource utilization	High/low percentage of non-occupied resources in a process over the total number of resources
Controls performed by internal resources	An activity that has one or several functional procedures and short processing time
Independent sequential activities	Activities that are executed in a sequence while they are not dependent on each other in terms of inputs, outputs and, resources
Frequent handovers	Transferring the case from one resource to another (as part of the business process or for the purpose of check, control, verification, etc.)
Ping pong behavior	Transferring the case between two consecutive activities from one resource to another
Rework	An activity or a fragment of the process that is repeated within one case
Knock-out	An activity that checks if the case should move forward with the process or if it should be rejected
Workaround	Temporary process deviation from the standardized business process
Highest waiting time in the business process	The longest average time that cases spend in an idle mode (waiting for further processing) between two activities

Cases with the highest waiting times	Cases that have the highest total time in an idle mode (waiting for further processing)
Bottleneck	A situation when the number of cases arriving exceeds the number of cases that can be handled which leads to queues or case build-ups
Manual time-consuming fragment	Business process fragment that is partially or fully executed manually
Manual process	Process most activities of which are executed manually
High process complexity	Process with high number of loops and/or branches (decision points)
Similar process variants	Two or more process variants that are similar in terms of case execution for different types of case attributes or by different resources
Overprocessing	Part of the business process, the output of which provides value neither to the client nor the business
Overproduction	Executed business process instance, the output of which is not later used (output is higher than demand)

Source: compiled by the author, based on Lashkevich and Milani (2020)

5.2. Results of RQ2 Analysis

After content analysis of BPIC submission reports and elimination of process-specific improvement opportunities that were not detected during RQ1 analysis, general instructions for six improvement opportunities were compiled (the full version of content analysis file can be accessed here: https://bit.ly/3SAaZzF). Table 9 provides the list of those improvement opportunities and number of papers based on which general guidelines were formulated.

Table 9. Improvement opportunities, for which guidelines were formulated based on the findings from BPIC reports and the number of papers that describe these improvement opportunities

Improvement opportunity	Number of BPIC papers
Handover	1
Ping pong behavior	11
Rework	24
Waiting time	24
Bottleneck	6
Overprocessing	22

From Dumas (2021b) were extracted instructions for identification of bottlenecks, rework, high and low resource utilization, and overprocessing. From Dumas (2021a) additional guidelines for bottlenecks and rework were identified.

Instructions for the rest of the improvement opportunities were developed based on the patterns identified from the abovementioned sources. Overall, data sources that are available for each of the selected improvement opportunities are summarized in Table 10.

Table 10. Data source availability for the templates

Template title	BPIC submissions	Dumas (2021a), Dumas (2021b)	Lashkevich and Milani (2020)	Sharma (2021)	Reijers and Mansar (2005)
Small activities			✓		
Large activities			✓		
Activity variants			✓		✓
Similar activities			✓		
High/low resource utilization		✓	✓		
Controls performed by internal resources					✓
Independent sequential activities			✓		✓
Frequent handovers	✓		✓	✓	
Ping pong behavior	✓			✓	
Rework	✓	✓	✓	✓	
Knock-out			✓		
Workaround			✓		
Highest waiting time in the business process	✓		✓	✓	
Cases with the highest waiting times				✓	
Bottleneck	✓	✓	✓	✓	
Manual time-consuming fragment			✓		
Manual process			✓		
High process complexity			✓		
Similar process variants					
Overprocessing	✓	✓	✓	✓	
Overproduction			✓	✓	

5.3. General Overview of the Developed Templates

After analysis of academic literature focused on identifying the list of improvement opportunities that can be discovered using event logs and process mining tools as well as content analysis that aimed at defining the algorithm patterns that can be applied, 21 templates (Low human resource utilization and High human resource utilization were combined into one template because of high similarity of instructions) were developed (see Supplementary Material B).

The structure of the final version of the file with templates is the following:

- Table of contents;
- Introduction;
- 21 templates;

- List of references.

To start with the table of contents, templates were grouped by the relation to the same process element. The groups provide general understanding of what the template will be focusing on. For instance, if an improvement opportunity relates to the waiting time, such as high waiting time between the activities, cases with high total waiting time, and bottlenecks, and manual time-consuming fragments, they are categorized as waiting time-related improvement opportunities. The list of groups and their composition is the following:

- **Activity-related improvement opportunities**:
  - Small activities;
  - Large activities;
  - Activity variants;
  - Similar activities;
- **Resource-related improvement opportunities**:
  - High/low resource utilization;
- **Activity- and resource-related improvement opportunities**:
  - Controls performed by internal resources;
  - Independent sequential activities;
  - Frequent handovers;
  - Ping pong behavior;
- **Control flow-related improvement opportunities:**
  - Rework;
  - Knock-out;
  - Workaround;
- **Waiting time-related improvement opportunities:**
  - Highest waiting time in the business process;
  - Cases with the highest waiting times;
  - Bottleneck;
  - Manual time-consuming fragment;
- **Process-related improvement opportunities:**
  - Manual process;
  - High process complexity;
  - Similar process variants;
- **Process wastes-related improvement opportunities:**
  - Overprocessing;
  - Overproduction.

The next part of the file with templates is an introduction, it covers the background information, description of the sections, brief instructions on how to use templates, and some general instructions about Apromore functionality.

After the introduction, 21 templates are provided. All the templates follow the same structure:

- **Improvement opportunity (IO)** – name of the improvement opportunity;
- **Definition** – short description of the improvement opportunity;
- **Examples** – short scenarios that demonstrate the improvement opportunity

- **Minimum data needed** – list of minimum data in the event log (e.g., activities, end timestamps) that are required to identify improvement opportunity;
- **Guideline on how to identify this IO** – step-by-step instruction on how to identify improvement opportunity in the process. The section starts with expected output that describes the main findings after completing certain step(s). It is followed by the instruction itself, which consists of step description (what you should do), Apromore example (screenshot with example of the output after completing the step), and explanation (some hints of what you should look for in the step and the description of the result of the step). The last part is output which describes the overall result after completing all the steps. This section allows the user to find the improvement opportunity but not the reasons why this improvement opportunity occurred;
- **Redesign possibilities** – generic ideas on how processes can be redesigned to address the improvement opportunity, based on Lashkevich and Milani (2020);
- **References** – list of references. This covers references to academic literature and/or references to BPIC submissions. Academic papers can cover case studies that address improvement opportunity, description of new method on how to define the improvement opportunity, description of the improvement opportunity. References to BPIC submissions provide articles where the particular improvement opportunity was identified with a short description of the process. References are not Apromore-specific, they depict improvement opportunity either for some other process mining tool (e.g., Celonis, ProM, Disco) or without linkage to particular process mining tool.

It was crucial not only to provide the guidelines but also to give the user better understanding of the improvement opportunity itself, that is why templates have the improvement opportunity name, definition, and examples sections. In order to check if the event log is suitable for identifying particular improvement opportunities before starting to perform steps, minimum data needed section was included into the structure. Guidelines section is the core part of the template, where step column is the instruction for the analyst what actions he/she should take and expected output, output, Apromore example, and explanation subsections are meant to provide additional support and guidance for the end user. Redesign possibilities were included in the templates to provide analysts with a set of ideas for the next steps that can be taken to improve the process. Lastly, references are meant to be a list of additional resources users can investigate to learn more about the improvement opportunity.

The last part of the file is the list of references, where all the materials based on which templates were created are collected as well as papers to which references were provided in the References section of the templates.

The general approach to use the templates is to start with determining the goal, if the process analyst's task is to identify all improvement opportunities in the process, it is recommended to use all the templates; if the task is to identify some specific improvement opportunities, process analyst can use only those templates. Before using the templates, especially for the first time, it is advised to read through the introduction. When it comes to the work with the template itself, firstly the end user needs to get familiar with the name, description and examples of improvement opportunity in order to make sure that he/she understands the meaning of the improvement opportunity. Secondly, the process analyst should check if the event log he/she is going to use has the required data with the help of Minimum data needed section. Once it is confirmed that there is all the required data in the log, the user should get familiar with the

expected outcome section and decide if he/she needs to perform several or all of the steps based on the task. When the scope is determined, process analyst needs to perform the selected steps, the user may find helpful to check Apromore example column with the screenshot of example of the performed step and explanation column. After steps are done, the process analyst should check his/her output with the output section. Additionally, in the redesign possibilities section, the end user may find ideas on how to redesign the process to make it more efficient. Moreover, if the analyst is interested in some extra information, he/she can check the materials provided in the resources section.

5.4. Overview of the Bottleneck Template

Bottleneck template (see Appendix I) can be examined as an example. Following the structure outlined in the previous subsection, template starts with the name of the improvement opportunity – Bottleneck (see 1. Improvement opportunity (IO) of Appendix I). The next section (see 2. Description of Appendix I) presents the description of the improvement opportunity of bottleneck that is "A situation when the number of cases arriving exceeds the number of cases that can be handled which leads to queues or case build-ups". Examples of this improvement opportunity from the healthcare process are "In the hospital, patients need to wait to make the computed tomography scan due to lack of personnel which causes queues" and "In the pharmacy, customers need to wait in the queue to get served because pharmacists are busy with serving other customers". The Examples section provides a description of how this improvement opportunity can manifest itself in a real-life scenario. To identify this IO using the process mining tool, an event log must have at least Activities, resources, start timestamps, and end timestamps (see 4. Minimum data needed of Appendix I). Using this information, an analyst can see if the event log they work on has all the required data to identify this improvement opportunity. The next section (see 5. Guideline on how to identify this IO of Appendix I) consists of three subsections: expected output, steps, and output. Expected output gives the end user an understanding of what will be the general outcomes of steps before executing them. In the bottleneck template, expected output is "Activity bottlenecks are found using *Steps 1-2*, resource bottlenecks are found using *Steps 3-4*". For example, if an analyst's task is to identify resource bottlenecks, from expected output, he/she knows that it can be done by performing Steps 3-4 only. Steps subsection divided by four columns: step number, step, Apromore example, and explanation. Step number is the ID of the step. Column step gives instruction an analyst needs to perform. Apromore example column gives a screenshot from Apromore which displays an example of a performed step. The explanation column contains additional information that may help the end user as well as the description of the result of the step. In the bottleneck template, the first step is "1) Open event log in the process discoverer. 2) In the Visualization settings, select Duration overlay and choose Average. In the View section, choose the Activities perspective", Apromore example gives the screenshot of configured view section, and the explanation is "The bottleneck is identified based on the long time needed to process the case or the long waiting time between activities. Result of the step: generated process map based on average duration and activities perspective". The last subsection of the guideline is output which gives an overview of the final outcome after completing all steps. In the bottleneck template, output is "List of bottlenecks based on activities and resources". The next section of the template is redesign possibilities (see 6. Redesign possibilities of Appendix I), which provides an overview of the possible redesign for the specific improvement opportunity. For the bottleneck template the redesign possibilities presented are "implement a technological solution to minimize constraints in the process;

organize separate process paths for different types of orders; add more resources to the business process, if bottlenecks are caused by a lack of resources; implement a scheduling system for clients to evenly distribute the workload, if bottlenecks are caused by queues during peak hours; implement a resource scheduling system that will allow having more resources during peak hours and fewer resources during periods with low demand; introduce a buffer queue; use incentives to shift customers from high-demand hours to low-demand hours; allow customers to execute some parts of the process by themselves". The last section (see 7. References of Appendix I) provides references to some BPIC submission reports where the process of identifying bottlenecks is better described and to several academic papers which describe other ways to identify this improvement opportunity.

6. Evaluation

In this section the results of templates evaluation and improvement phase (see Figure 5) are described. Following the order of stages of validation step presented in Section 4, firstly, the results of the templates testing phase of the interview combined with interviewees answers to post-interview questions are elaborated. Secondly, implemented improvements are described. Thirdly, the results of post-survey are presented.

6.1. Interviews Findings

This subsection outlines the results of conducted interviews. It includes both comments from templates testing phase of the study as well as post-interview questions stage. As it was described in Section 4, comments were analyzed using affinity diagram method, based on which elicited quotes from the interviews were clustered on code-, theme-, and evaluation goal-levels (EG) (see Table 7). Subsubsection 6.1.1. covers EG1 and EG2, while subsubsection 6.1.2. describes the results associated with EG3.

6.1.1. Comments About Usefulness and Ease of Use

The aim of the first evaluation goal is to evaluate usefulness of the templates. This goal comprises of two themes, content, which clustered participants' statements based on the section of the templates they relate to, and usefulness, which collected statements connected with usefulness of the templates or their sections.

The majority of the participants mentioned guideline section as useful (I02, I03, I04, I05, I06, I08, I09, I10, I11). As this section was the main part interviewees were working with while identifying improvement opportunities, the section brings value to end users and helps them to find improvement opportunities. One interviewee said: "*This step-by-step guideline is very useful. Especially, I can imagine that it is very useful for novice users."* (I11). It is worth mentioning that usually participants indicated specific parts of the guideline section as the most useful. I02, I03, I04, I05, I08, I09, I10, and I11 found screenshots of the Apromore example part useful, for example, "*The images are a good help because only with text it is more difficult to visualize it and understand it."* (I05).

Sections that also were mentioned by participants as useful are examples (I01, I06, I10, I12), minimum data needed (I01, I06, I10), definition (I02, I11), type of IO (I02) and references (I06). Speaking about the examples section, one interviewee said: "*I really like the example, since you give the examples. So, it was pretty nice to understand the overall if the process can be generic one, but you can actually read through the example to get to know that. So, this is an example, it does make a lot more sense."* (I06). On the contrary, I05 reflected: "*I wasn't paying much attention to the example. […] I think the example is useful. It depends also on the experience of the analyst.*". Since I05 is from group 2 and I06 – from group 1, it indicates that people with advanced knowledge and experience with BPM and process mining tools understand the meaning of improvement opportunities without examples, but for people with limited knowledge and experience it is helpful to have examples to get better understanding.

Regarding the minimum data needed section, one of the participants said: "*The minimum data is also a good thing. Because when you have the data that needs to be there, the user can actually see that this is the minimum data that we require for this thing.*" (I06). At the same time, I11 indicated: "*I don't imagine yet how I could use it [minimum data needed]. Perhaps if*

I were to upload the event log first, it would help me, but I don't know about that yet because I haven't used it.". As for the purpose of the study participants were provided with event log that satisfies minimum data needed requirement, interviewees did not have a chance to check if this section will help them when they will be using templates with different event logs.

Speaking about the definition section, one interviewee said: "*The short definition of the improvement opportunity that we're looking at [is useful] because the name can be not very self-explanatory, so this helps.*" (I11). This section also provides value to the end users. The same applies to the type of IO section *"[Type of IO section] it's pretty helpful to understand what you actually doing*" (I02), and references section "*References is a good thing. So, if somebody is doing the research project, they can actually refer to what exactly, they can use these references specifically.*" (I06). But I10 brings the argument that usefulness of the references part largely depends on in which context templates will be used: "*For academic work, references are awesome, great, everyone loves them. If you're going to put it into a company, maybe 10% actually care.*".

Answering the question about are there some redundant sections or parts of sections, most of the participants could not mention anything, but I08 said: "*I think the explanation part [is redundant], I almost skipped. I didn't look at that at all.*", this statement was also supported by I05 and I12. I08 and I05 are participants with advanced knowledge of BPM and Apromore, so it can be assumed that this section does not bring much value to experienced users. At the same time, novice users I01 and I04 mentioned that this part was useful for them as well as other participants with little experience were using this part more extensively based on the observations during the study. Additionally, regarding the explanation part, I01 mentioned that it can be improved to bring more value: "*[Explanation part] doesn't give me an insight how do I read it. That's what I said. So, those total frequency means that how many cases like the person does and the duration, like how much time it spends. Kind of human translation and what it means because this is just the result, what I see on the picture. But to actually understand what I see, that would be cool to give an explanation*".

Some of the interviewees (I01, I06) also said that the templates overall are useful. I01 said: "*Overall it's a good instruction, I wish I had it during my exam.*". Interestingly, the only section that was not mentioned by participants neither as useful nor as not useful is redesign possibilities.

Speaking about the second evaluation goal, it aims to evaluate the ease of use of the templates by considering Understandability and Ease of use themes. The objective is to define aspects that were understandable and easy to use as well as those with which participants experienced some issues.

To start with understandability, it can be analyzed from the following perspectives: understandability of the content, understandability of the structure, and issues with understandability. Also, it is worth mentioning that there were general comments for all templates overall and specific statements about particular templates.

Speaking about understandability of the content, most of the comments were about the guideline section (I01, I10, I12). For example, "*The steps are clear to me at least. Especially once you understand how everything is laid out and how to follow them."* (I10). Moreover, one of the interviewees said that the templates overall are understandable: "*Overall, I liked the templates,*

and they seem very easy and clearly written, so it should be a great support and help for people who are especially new to different process mining tools." (I09). Additionally, there was one statement regarding other sections of the templates: "*The improvement opportunity, the definition, the minimum data needed, I think all clear*." (I07).

From the perspective of the structure, three participants mentioned that it is understandable (I01, I04, I09). One of the interviewees said: "*What I especially loved about it is the structure of the template. That I know they go, I know what I should expect, and I know what kind of steps I should perform and clicking what buttons. So, I would say that the structure is the strongest part of this templates.*" (I09).

But there are also several issues with understandability identified, all of them were mentioned one time. One of the participants said that the explanation part of guideline section is not understandable and should be improved: "*Add more human explanation of what they actually see on the screen. Because if the person who doesn't have much experience and they just follow the instruction, I guess they wouldn't be able to recognize what the actual outcome there.*" (I01). Another issue is with having expected output and output in the guideline section: "*You have two outputs; you have the expected output and then in the beginning of the template have always this output there in the end. They are kind of similar, I guess. Maybe so. I'm not sure. But this might be just one. Or if they are different than they actually add value for them then leave it as it is. This might be a bit confusing.*" (I01). Another interviewee mentioned that it is not understandable in what order templates are listed and proposed to classify them based on abstraction level: "*this is certainly useful that you can just find which one [template in the table of contents] you want and go back to it. But how they are listed? […] I think maybe they could be broken down into more abstract and less abstract [templates]*." (I11). One more issue was regarding understandability of the examples as they are mostly created based on some event log and if the end user is not familiar with it, there may be some confusion: "*[Examples section] for example, here in the insurance claims handling process activity assess claim is a large activity. I worked with these event logs before, so I know that insurance claims is and what is assess claim. But if I were to look at it first, I am not sure that it would be very self-explanatory to me. [...] There is such big variation between these examples that I think maybe it could be useful to have some very generic process which is known to everybody like go shopping for bread [...] And in the examples this could be used like unified one example for each template so that the person wouldn't have to look at it.*" (I11). Another interviewee said that there is variation between proposed redesign possibilities, and it is not understandable which redesign should be applied in particular scenarios: "*I would say that sometimes there the redesign possibilities were not so connected*." (I12).

Additionally, there were mentioned issues with understandability of concrete templates. I01 experienced difficulties with High/low resource utilization template and described two issues: "*How do I know if [Resource] 25 is still the lowest? It has 11 [cases], but it has much more than [Resource] 29. It is still like light blue*" and "*So, the first view [Step 1] was like how many times the resource does the task and this one [Step 3] is how much time it takes to do it? ... I would really appreciate if it was written somewhere*". So, I01 had a problem with defining the threshold and not understandable explanations of the guideline section. Another template that brought many issues is Activity variants, all of them are regarding step part of the guideline section. An example could be: "*I was confused with the text, how it's written in the step. Because it says that I need to analyze both obtained distributions and I need to check for outliers and*

then list this activity and I was confused because here what we're looking for is a graph which shows the average processing time for cases and then somehow I completely forgot that I already mode here a filter for this particular activity and this just didn't work." (I11).

The other theme that needs to be considered is ease of use. Some of the participants said that the templates are easy to use (I01, I12). For example, "*From finding problems [improvement opportunities], I think it does the job really well. It really takes my hand and shows me exactly what I have to do and what I have to find.*" (I12). But there are also several issues during the use mentioned: too small screenshots (I01), need to detect improvement opportunities based on filtering some data, not just visually (I04), and confusing explanation part of the guidelines section, where explanations are in the same column as the result (I12). The corresponding statements are: "*No, in that case [I didn't pay attention to the screenshots], they are pretty small here.*" (I01), "*I was thinking it will be easier for me to just go and filter out stuff somehow here.*" (I04), and "*I don't think it's like about adding something, it was just organizing it a little different because you have the result and a little explanation. They were a bit confusing to understand. What am I reading here? I didn't know what to look for. [...] I'm just thinking it would have been probably clearer to have this result part separate*." (I12).

6.1.2. Comments About Possible Improvements

The last evaluation goal was to define the ways in which the templates can be improved and be more useful and easier to use. This goal is represented by Improvement theme, which was coded as content improvement, structure improvement, screenshots improvement, adding additional section.

In terms of content improvements, several participants (I01, I07) suggested removing either expected output or output: "*What I would suggest is minimize the text on expected output. Just put the table with the step, examples and the explanation, that's it*." (I07). Another idea is to provide an explanation about why each of the improvement opportunity types is dangerous for the process efficiency: "*I'm missing the part that tells me if I have those similar activities [for example], what is kind of the default danger. Why should I think that having similar activities is something that makes my process unoptimized."* (I10). And in the same way describe the benefits of process redesign: "*Redesign possibilities, the same thing, what is the expected benefit of actually doing the redesign."* (I10).

Regarding the improvement of examples section, I05 and I07 proposed to have bigger examples: "*It's maybe better to develop a bit the example, so that you get a better idea before even going forward in these steps.*" (I05); I11 had an idea to have one comprehensive scenario, based on which all of the examples be made: "*There is such big variation between these examples that I think maybe it could be useful to have some very generic process which is known to everybody like go shopping for bread [...] And in the examples this could be used like unified one example for each template so that the person wouldn't have to look at it*.".

Speaking about the reference part, I09 proposed to have hyperlinks to papers: "*Maybe you can also add the hyperlink to the resources*."; I11 offered to provide not references to academic literature but to resources where improvement opportunities are described in more detail: "*I think that maybe instead of the reference to a particular paper, what would be useful specifically for me is rather a reference to a place where this improvement opportunities may be described in more detail but not in the entire paper.*".

I03 suggested to add an option to identify improvement opportunity using quantitative representation, not only visually, where it is possible: *"[I would improve] the quantitative thingy, if you can provide the quantitative for them to compare with the numbers, that would be better than just visualization.*". One more idea is to provide definitions about Apromore tabs and/or other functionality: "*It would be good to have more definitions regarding [...] a section [of Apromore, e.g., View section] like what you can do there*" (I02). Another suggestion is to develop more explanation part of the guideline section, so that it will have more understandable text: "*Add more human explanation of what they actually see on the screen. Because if the person who doesn't have much experience and they just follow the instruction, I guess they wouldn't be able to recognize what the actual outcome there*" (I01). Additionally, there were a few minor corrections in concrete templates, for example, improve spacing in Frequent handovers template (I01); change word "activity statistics section" in Activity variants template (I05, I11), rephrase word "extreme cases" in Activity variants template (I11).

Considering structure improvements, a few of the interviewees (I10, I11, I12) proposed different ways to categorize templates in the table of contents to increase the ease of use: "*maybe organize or group some of them [templates in the table of contents] together. One way to do it would be based on the minimum data needed. […] Another one would be based on, let's say perspective."* (I10). Also, there was an idea to merge type of IO and definition sections: "*Depending on how long the definitions are, if they are all of this length then you could even merge them with the improvement opportunity. Have small activities, space, dash, space followed by a definition, if you really want*." (I10); and to make result of the step as an additional column in the guideline section: "*[…] you have the result and a little explanation. They were a bit confusing to understand. [...] I'm just thinking it would have been probably clearer to have this result part separate.*" (I12).

From the perspective of screenshots improvements, I04 and I08 suggested adding more images: "*I think for me it's mostly some of the steps might have been more helpful if to add like more pictures to visualize what exactly is going on.*" (I04). Another idea was to increase the size of pictures: "*No, in that case [I didn't pay attention to the screenshots], they are pretty small here*" (I01). Moreover, I12 offered to add screenshots to the redesign possibilities: "*If you're telling me here a redesign possibility, if you could somehow also show me how that should look like visually speaking*.".

The last code from Improvement theme is adding additional section. Many of the participants (I01, I02, I04, I07, I10, I11) suggested to add some introduction or preface at the beginning of templates file, where can be included background information, generic description of some features of Apromore, recommendations on how to use the templates. For example, "*I am considering actually having some kind of introduction or preface. What are the templates about? And how do you use them? And maybe you'll need something to make notes [pen and paper or a document file].*" (I10). Additionally, I06 proposed to add unique identifier to each template: "*I think there can be one thing that you can add to it. […] it's better to have ID numbers associated with it because when you are developing the templates and the project is going to be bigger and bigger […] it is good to have identification number*.".

6.2. Templates Improvements

All of the suggestions proposed by study participants are collected in Appendix II. Some of the ideas were concerning one particular template, and some were general, applicable to all

templates. These suggestions were prioritized based on whether they should be addressed during this study (detailed information on the prioritization principles were provided in Subsection 4.5.). In Appendix II, improvements that should be implemented are marked in green, those that are nice to be implemented – in amber, those that should not be implemented during this study – in red. Additionally, the level of effort required to implement the change is evaluated based on the author's experience; it is presented in the last column of Appendix II, along with other comments that provide reasoning for the prioritization.

Overall, six suggestions were implemented. The current file with Templates is an updated version where these improvements are in place. The rest of the ideas are either subject to further implementation as a part of future work or they are suggested not to be implemented.

6.3. Post-survey Findings

This subsection summarizes the results of the post-survey separately for Group 1 and Group 2 in order to identify similarities and differences between the groups in terms of perceived usefulness and perceived ease of use as well as to understand how participants perceive templates in general.

As it was defined in the previous sections, the target audience of the templates are people with limited to medium experience with BPM, and Apromore. So, both of the groups selected for the study are considered as target end users. Participants were asked to complete a post-survey that comprised of eight questions focused on assessment of perceived usefulness and five questions were for assessment of perceived ease of use. Tables 11 and 12 provide the summary of the post-survey results.

Table 11. Post-survey statistics for group 1 (statements were evaluated from 1 (Strongly Disagree) to 5 (Strongly Agree))

Statement number	**Interviewee ID (group 1, elementary skills)**									
	01	**02**	**04**	**06**	**09**	**12**	Mean	St.dev.	Median	Variance
PU1	4	5	4	4	5	5	4,50	0,55	4,5	0,30
PU2	3	5	5	4	5	5	4,50	0,84	5	0,70
PU3	4	4	5	5	5	5	4,67	0,52	5	0,27
PU4	4	4	5	4	4	5	4,33	0,52	4	0,27
PU5	4	5	5	4	5	5	4,67	0,52	5	0,27
PU6	3	5	4	4	4	4	4,00	0,63	4	0,40
PU7	2	5	4	5	4	5	4,17	1,17	4,5	1,37
PU8	5	5	4	5	4	5	4,67	0,52	5	0,27
PEOU1	5	5	5	3	4	5	4,50	0,84	5	0,70
PEOU2	4	5	5	4	4	5	4,50	0,55	4,5	0,30
PEOU3	5	4	5	5	5	5	4,83	0,41	5	0,17
PEOU4	5	4	5	5	4	5	4,67	0,52	5	0,27
PEOU5	4	4	5	5	5	5	4,67	0,52	5	0,27
Usefulness (EG1)							**4,44**	**0,66**	**4,63**	**0,48**

Ease of Use (EG2)							**4,63**	**0,57**	**4,90**	**0,34**

Table 12. Post-survey statistics for group 2 (statements were evaluated from 1 (Strongly Disagree) to 5 (Strongly Agree))

Statement number	**Interviewee ID (group 2, advanced skills)**									
	03	**05**	**07**	**08**	**10**	**11**	Mean	St.dev.	Median	Variance
PU1	4	5	4	5	5	5	4,67	0,52	5	0,27
PU2	3	4	4	4	3	4	3,67	0,52	4	0,27
PU3	4	5	4	3	4	3	3,83	0,75	4	0,57
PU4	3	4	3	5	5	4	4,00	0,89	4	0,80
PU5	5	5	4	5	5	5	4,83	0,41	5	0,17
PU6	3	5	3	4	4	5	4,00	0,89	4	0,80
PU7	4	5	5	5	5	3	4,50	0,84	5	0,70
PU8	4	5	5	5	5	5	4,83	0,41	5	0,17
PEOU1	4	5	5	5	5	4	4,67	0,52	5	0,27
PEOU2	3	5	5	5	4	4	4,33	0,82	4,5	0,67
PEOU3	5	5	5	4	5	3	4,50	0,84	5	0,70
PEOU4	5	5	5	5	5	5	5,00	0,00	5	0,00
PEOU5	5	4	5	5	5	5	4,83	0,41	5	0,17
Usefulness (EG1)							**4,29**	**0,65**	**4,50**	**0,47**
Ease of Use (EG2)							**4,67**	**0,52**	**4,90**	**0,36**

The majority of the questions participants from both groups answered with 4 (Agree) and 5 (Strongly Agree). From the perceived usefulness statements, Group 1 rated the lowest in terms of mean grade (4,00), PU6 (Using Templates would enhance my effectiveness in identifying relevant redesign possibility for each improvement opportunity), which can be explained by the fact that participants were not asked to interact with redesign possibilities during the study. The lowest mean grade for Group 2 (3,67) in terms of perceived usefulness is PU2 (Using Templates would make it easier to understand the meaning of improvement opportunities (based on definition and examples sections), this is a surprising finding as most of the participants of this group were familiar with the improvement opportunities. Such a low result can be explained by the fact that Group 2 was paying less attention to the definition and examples sections than Group 1, based on the observations. Speaking about the highly ranked statements, in Group 1 they are PU3 (Using Templates would enable me to define what data I need to identify improvement opportunities more quickly), PU5 (Using Templates would enhance my effectiveness in identifying improvement opportunities), and PU8 (I would find Templates overall useful), all are 4,67. For Group 2, they are PU5 (Using Templates would enhance my effectiveness in identifying improvement opportunities), and PU8 (I would find Templates

overall useful), both 4,83. It is interesting to notice that the highest marks in both groups are for the usefulness of the templates in terms of identifying improvement opportunities as well as the usefulness of templates overall, this indicates that study participants see the benefit in the templates as a tool for improvement opportunities identification.

From the perspective of perceived ease of use, statements that have the lowest result in Group 1 are PEOU1 (Learning how to use Templates would be easy for me), and PEOU2 (I would find it easy to use Templates to identify improvement opportunities and redesigns), both 4,50, this indicates that participants from this group experienced some issues with usability of templates. For Group 2, the statement that was evaluated the lowest is POEU2 as well (4,33). The statement with the highest mark in Group 1 is PEOU3 (Structure of Templates is clear and understandable) and in group 2 it is PEOU4 (It would be easy for me to become skillful at using Templates).

Overall, the results for Usefulness and Ease of Use are similar for both groups. In terms of EG1 (Usefulness), group 1 rated it at 4,44 and group 2 rated it at 4,29. For EG2 (Ease of Use), group 1 marked it at 4,63 and group 2 – 4,67.

To sum up, there is no significant variation between people with elementary skills in BPM and Apromore and those who are more experienced in terms of perceived usefulness and ease of use of the templates. Both groups rated these criteria on a high level which indicates that templates are helpful in identifying improvement opportunities using Apromore and it is easy to start using the templates.

7. Limitations

This section outlines the limitations of the research. As different methods were applied for different steps of the study, limitations will be discussed separately for RQ1 analysis, RQ2 analysis, templates development, and templates evaluation steps.

To start with the limitations connected to the results of RQ1 analysis, the findings of this step are grounded in three academic papers which is a limited amount and may indicate that not all the relevant sources were included in the study. This was mitigated by selecting two studies that follow SLR procedure, the first article analyzed 150 papers, the second one – 187 papers, which gives a solid base that most of the possible improvement opportunities were covered in these sources. The second limitation of this step is that not all the improvement opportunities that can be found by manual analysis of event logs using Apromore were listed. This was mitigated by continuous review of the intermediary results by the supervisors who have extensive experience in the field.

There are several limitations to the content analysis conducted to answer RQ2. Firstly, there was a bias of data extraction, which was minimized by conducting analysis by 10% of papers by the author and one of the supervisors in parallel to ensure that the valid set of tags and categories is selected as well as that the correct data is extracted. Secondly, the threat of misinterpretation of results and incorrect generalizations of patterns (Bengtsson, 2016) was mitigated by regular meetings with supervisors where the results were discussed.

The limitations connected to the templates' development is that templates provide inaccurate or not optimal ways to identify improvement opportunities, because of author's lack of knowledge and skills in the field. This was mitigated in two ways: all the templates were reviewed several times by supervisors who have a lot of experience, all their suggestions were incorporated; also, after templates were developed, they were tested as a part of evaluation process, all the interviewees were able to identify improvement opportunities.

Lastly, there are limitations to the setting of the evaluation processes. In order to ensure the comparability of results study participants were selected from the students of the same university who have experience in completing the same university courses as well as none of the study participant is working as a process analyst. It can be expected that people with different academic and professional backgrounds (for example, processes analysts that did not complete any university courses on BPM and Apromore) will provide different results to the study. This threat was minimized by including into the study participants that learned BPM and Apromore by other means than completing Business Process Management and/or Business Process Mining courses at the University of Tartu as well as by including several participants that have work experience as analysts in different fields. Additionally, as the aim of evaluation was to measure the usefulness and understandability of the templates, only 9 of 21 templates were tested to ensure that one template was used by several participants. This may result in templates that were not tested be not as useful or understandable as the ones selected. Moreover, coding of the interviews was done solely by the author which might lead to interpreter bias. As an attempt to mitigate it, the author closely followed the established procedure as well as the findings were several times discussed with the supervisors.

To conclude, this thesis is subject to several limitations. Most of the identified limitations were mitigated in order not to obstruct the research results.

8. Conclusion

This thesis aimed to develop templates for improvement opportunities identification from event logs using Apromore. To fulfil this aim, two research questions were attempted to answer: what improvement opportunities can be detected from event logs by manual analysis using Apromore, and how can it be done.

In order to analyze what improvement opportunities can be identified using Apromore, improvement opportunities mentioned in three academic papers were investigated. As a result, 22 improvement opportunities were elicited. To understand how these improvement opportunities can be found in the event log using process mining tools, content analysis of 129 reports of BPIC as well as analysis of templates proposed by Dumas (2021a) and Dumas (2021b). From these sources, patterns for identification of 6 improvement opportunities were collected and generalized. Based on an elicited list of improvement opportunities and patterns, 21 templates for identifying improvement opportunities using Apromore were developed. To evaluate the usefulness and understandability of the templates, 12 interviews were conducted. Interviewees were invited to test three templates, after which to answer a set of questions and fill in a post-survey. Based on the observations during the testing phase and analysis of the participants' answers, a list of suggestions about how templates could be improved was extracted and prioritized. Ideas that were marked as high priority were implemented. Usefulness and understandability were evaluated from the interviews and post-survey. Overall, during the interviews, participants mentioned that templates are useful and easy to use. The same was derived from the survey results. The group with limited experience rated usefulness at 4,44 out of 5 and the more experienced group rated it at 4,29. In terms of ease of use, the first group rated it at 4,63 out of 5 and the second group – 4,67.

There are several directions for the future work. Firstly, the developed templates can be improved by implementing the rest of the proposed suggestions during the interviews. Secondly, ways for templates automatization can be researched and possibilities for such automatization developed. Having a plug-in or other type of solution that automatically identifies improvement opportunities in the processes will help process analysts to analyze event logs faster and more efficiently.

List of references

1. Adriansyah, A., & Buijs, J. C. (2012, September). Mining process performance from event logs. In *International Conference on Business Process Management* (pp. 1-46). Springer, Berlin, Heidelberg.

2. Andersen, B. (2007). Business process improvement toolbox. Quality Press.

3. Association of Business Process Management Professionals. (2009). *Guide to the business process management common body of knowledge (BPM CBOK): ABPMP BPM CBOK-[version 2.0-second release]*. ABPMP.

4. Awad, A., Zaki, N. M., & Di Francescomarino, C. (2016). Analyzing and repairing overlapping work items in process logs. *Information and Software Technology*, *80*, 110-123.

5. Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus open*, 2, 8-14.

6. Brils, J.H., Elsen, N.A., Priester, J., & Slooff, T.A. (2018). Business Process Intelligence Challenge 2018: Analysis and Prediction of Undesired Outcomes.

7. Buijs, J. C., van Dongen, B. F., & van der Aalst, W. M. (2011). Towards cross-organizational process mining in collections of process models and their executions. In *International Conference on Business Process Management* (pp. 2-13). Springer, Berlin, Heidelberg.

8. Celik, U., & Akçetin, E. (2018). Process mining tools comparison. *Online Academic Journal of Information Technology, 9*, 97-104.

9. Cho, M., Song, M., & Yoo, S. (2014). A systematic methodology for outpatient process analysis based on process mining. In *Asia-Pacific Conference on Business Process Management*, 31-42. Springer, Cham.

10. Davenport, T. H. (1993). *Process innovation: reengineering work through information technology*. Harvard Business Press.

11. Davenport, T.H., & Spanyi, A. (2019). *What process mining is, and why companies should do it*. Harvard Business Review. Retrieved from: https://hbr.org/2019/04/what-process-mining-is-and-why-companies-should-do-it

12. dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., ... & Scalabrin, E. E. (2019). Process mining techniques and applications–a systematic mapping study. *Expert Systems with Applications, 133*, 260-295.

13. Dumas, M. (2021a). *Business Process Mining. Lecture 5: Conformance Checking.* Retrieved from: https://courses.cs.ut.ee/LTAT.05.025/2021_spring/uploads/Main/Lecture5.pdf

14. Dumas, M. (2021b). *Business Process Mining Lecture 6: Process Performance Mining*. Retrieved from: https://courses.cs.ut.ee/LTAT.05.025/2021_spring/uploads/Main/Lecture6.pdf

15. Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management.* (1st ed.). Heidelberg: Springer.

16. Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of business process management*. (2nd ed.) Springer Berlin Heidelberg.

17. Elkoumy, G., & Dumas, M. (2022). Libra: High-Utility Anonymization of Event Logs for Process Mining via Subsampling. *arXiv preprint arXiv:2206.13050*.

18. Ganesha, K., Dhanush, S., & SM, S. R. (2017). An approach to fuzzy process mining to reduce patient waiting time in a hospital. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*,1-6. IEEE.

19. Grisold, T., Mendling, J., Otto, M., & vom Brocke, J. (2020). Adoption, use and management of process mining in practice. *Business Process Management Journal*.

20. Gupta, M., Serebrenik, A., & Jalote, P. (2017). Improving Software Maintenance Using Process Mining and Predictive Analytics. *2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 681-686.

21. Hammer, M., & Champy, J. (1993). *Rengineering the corporation: A manifesto for business revolution*. New York: HarperColins.

22. Kas, S., Post, R., & Wiewel, S. (2020). Automated Machine Learning in a Process Mining Context. In *International Conference for Process Mining (ICPM)* (pp. 1-25).

23. Kebede, M., & Dumas, M. (2015). Comparative evaluation of process mining tools. *University of Tartu*.

24. Key Features. (n.d.). Retrieved from: https://apromore.com/key-features/

25. Khodaev, D., Kalugin, V., Korneeva, E., Savintseva, M., & Balashova, M. (2020). Process Mining in Finance Sector: Trip Compensation Process in TU/e Business Process Intelligence Challenge 2020.

26. Klein, S., Lahann, J., Mayer, L., Neu, D., Pfeiffer, P., Rebmann, A., ... & Fettke, P. (2020). Business Process Intelligence Challenge 2020: Analysis and evaluation of a travel process. In *10th Business Process Intelligence Challenge at the Int. Conf. on Process Mining (ICPM).*

27. Laguna, M., & Marklund, J. (2018). *Business process modeling, simulation and design.* Chapman and Hall/CRC.

28. Lashkevich, K., & Milani, F. P. (2020). Business process improvement opportunities: a framework to support business process redesign.

29. Lewis, L. (2021). *5 Business Process Improvement Examples*. Retrieved from: https://www.processmaker.com/blog/5-business-process-improvement-examples/

30. Malinova, M., Gross, S., & Mendling, J. (2022). A study into the contingencies of process improvement methods. *Information Systems, 104*, 101880.

31. Market guide for process mining. (2019). Retrieved from: https://www.gartner.com/en/documents/3939836/market-guide-for-process-mining

32. Măruşter, L., & van Beest, N. R. (2009). Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems, 21*(3), 267-297.

33. Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1*(2). https://doi.org/10.17169/fqs-1.2.1089

34. Mertens, W., Pugliese, A., & Recker, J. (2017). Models with Latent Concepts and Multiple Relationships: Structural Equation Modeling. In *Quantitative Data Analysis* (pp. 37-59). Springer, Cham.

35. Ongena, G., & Ravesteyn, P. (2019). Business process management maturity and performance: A multi group analysis of sectors and organization sizes. *Business Process Management Journal.*

36. Ou-Yang, C., & Winarjo, H. (2011). Petri-net integration–An approach to support multi-agent process mining. *Expert systems with applications, 38*(4), 4039-4051.

37. Pakileva, A., Skvortsova, E., Zakoryuchkin, N., Tsaplin, S., & Zarubin, V. (2020). Business Process Intelligence Challenge 2020: Investigating Business Trips Arrangement Process at the Eindhoven University of Technology (TU/e).

38. Rebuge, Á., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information systems, 37*(2), 99-116.

39. Reijers, H. A., & Mansar, S. L. (2005). Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega, 33*(4), 283-306.

40. Sharma, A. S. (2021). Waste Identification from Event Logs.

41. Son, S., Yahya, B. N., Song, M., Choi, S., Hyeon, J., Lee, B., ... & Sung, N. (2014). Process mining for manufacturing process analysis: a case study. In *Proceeding of 2nd Asia Pacific Conference on Business Process Management, Brisbane, Australia*.

42. Song, M., & Van der Aalst, W. M. (2008). Towards comprehensive support for organizational mining. *Decision support systems, 46*(1), 300-317.

43. Van Der Aalst, W. (2012). Process mining. *Communications of the ACM, 55*(8), 76-83.

44. Van Der Aalst, W. (2016). *Process mining*. Springer, Berlin, Heidelberg.

45. Van Der Aalst, W. M. (2011, April). Process mining: discovering and improving Spaghetti and Lasagna processes. In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 1-7. IEEE.

46. Van Der Aalst, W. M., Reijers, H. A., Weijters, A. J., van Dongen, B. F., De Medeiros, A. A., Song, M., & Verbeek, H. M. W. (2007). Business process mining: An industrial application. *Information Systems, 32*(5), 713-732.

47. Van der Aalst, W. M., van Dongen, B. F., Günther, C. W., Rozinat, A., Verbeek, E., & Weijters, T. (2009). ProM: The process mining toolkit. *BPM (Demos), 489*(31), 2.

48. Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Wynn, M. (2011). Process mining manifesto. In *International conference on business process management* (pp. 169-194). Springer, Berlin, Heidelberg.

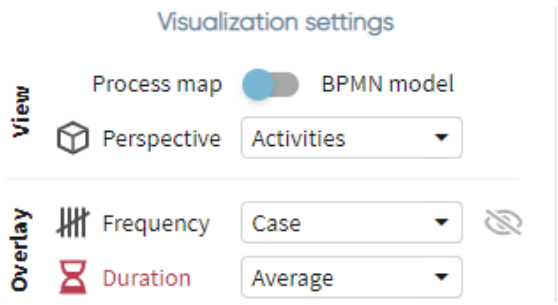
49. Weijters, A. J. M. M., van Der Aalst, W. M., & De Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP, 166(July 2017)*, 1-34.

50. Wixom, B., & Watson, H. (2010). The BI-based organization. *International Journal of Business Intelligence Research (IJBIR), 1*(1), 13-28.

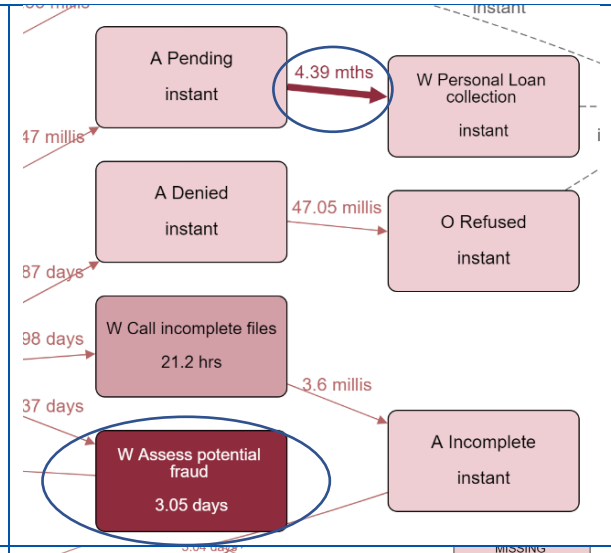
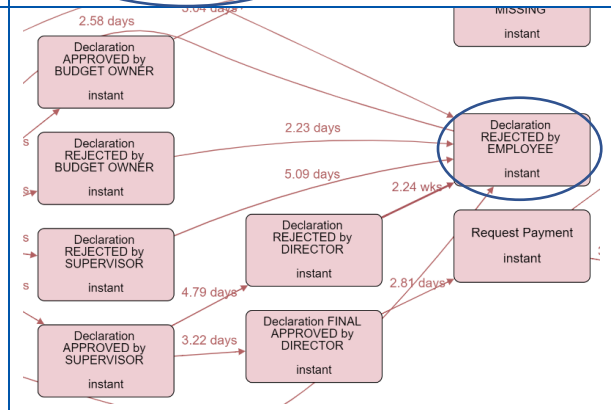
51. Zur Muehlen, M., & Shapiro, R. (2015). Business process analytics. In *Handbook on Business Process Management 2* (pp. 243-263). Springer, Berlin, Heidelberg.

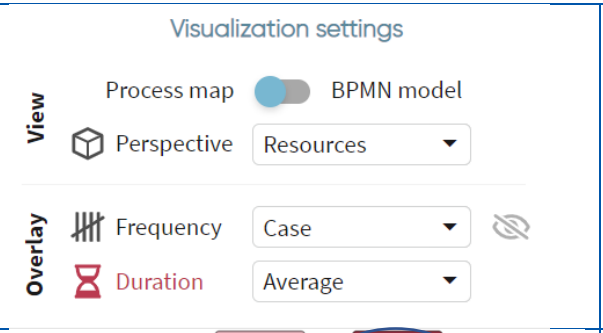
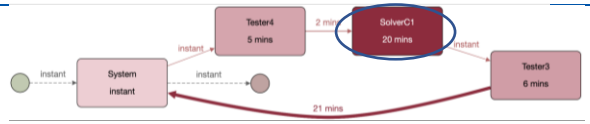
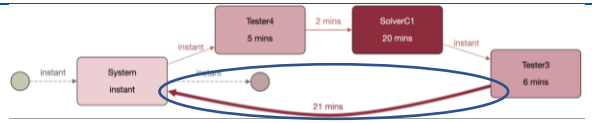
Appendix

I. Bottleneck template

<table>
<tr><td>1. Improvement opportunity (IO)</td><td colspan="3">Bottleneck</td></tr>
<tr><td>2. Definition</td><td colspan="3">A situation when the number of cases arriving exceeds the number of cases that can be handled which leads to queues or case build-ups</td></tr>
<tr><td>3. Examples</td><td colspan="3">In the healthcare process:
- In the hospital, patients need to wait to make the computed tomography scan due to lack of personnel which causes queues
- In the pharmacy, customers need to wait in the queue to get served because pharmacists are busy with serving other customers</td></tr>
<tr><td>4. Minimum data needed</td><td colspan="3">Activities, resources, start timestamps, end timestamps</td></tr>
<tr><td rowspan="3">5. Guideline on how to identify this IO</td><td colspan="3">Expected output: Activity bottlenecks are found using Steps 1-2, resource bottlenecks are found using Steps 3-4.</td></tr>
<tr><td>#</td><td>Step</td><td>Apromore example</td><td>Explanation</td></tr>
<tr><td>1</td><td>1) Open event log in the process discoverer
2) In the Visualization settings, select Duration overlay and choose Average. In the View section, choose the Activities perspective</td><td> 
</td><td>The bottleneck is identified based on the long time needed to process the case or the long waiting time between activities

Result of the step: generated process map based on average duration and activities perspective</td></tr>
</table>

	2	From the process map, 1) Find **activity bottlenecks**: activities with the longest processing time. List these activities. 2) Find **waiting time bottlenecks**: arcs with the longest duration. List activity pairs.		To find **activity bottleneck**, define activities with the longest duration. To find **waiting time bottlenecks**, find the longest waiting times (the thicker the arrow between activities, the higher is waiting time). On the example screenshot, they are circled in blue.
		3) Find **resource-capacity bottlenecks**: activities with the highest number of incoming arcs with long arc duration. List these activities.		To find **resource-capacity bottlenecks**, define activities all/most incoming arcs of which have long waiting time. On the example screenshot, it is circled in blue. **Result of the step:** list of activity bottlenecks, waiting time bottlenecks, and resource-capacity bottlenecks

	3	In the *Visualization settings*, select *Duration overlay* and choose *Average*. In the *View* section, choose the *Resources* perspective		**Result of the step:** generated process map based on average duration and resources perspective
	4	From the process map, 1) Find **resource bottlenecks**: resources with the highest processing time from the resource perspective. List these resources		To find **resource bottleneck**, define resources with the longest duration. On the example screenshot, it is circled in blue.
		2) Find **waiting time bottlenecks from the resource perspective**: arcs with the longest duration. List these pairs of resources		To find **waiting time bottlenecks from the resource perspective**, find the longest waiting times (the thicker the arrow between resources, the higher is waiting time). On the example screenshot, it is circled in blue. **Result of the step:** list of resource bottlenecks, and waiting time bottlenecks from the resource perspective
	Output: List of bottlenecks based on activities and resources			

6. Redesign possibilities	- Implement a technological solution to minimize constraints in the process - Organize separate process paths for different types of orders - Add more resources to the business process, if bottlenecks are caused by a lack of resources - Implement a scheduling system for clients to evenly distribute the workload, if bottlenecks are caused by queues during peak hours - Implement a resource scheduling system that will allow having more resources during peak hours and fewer resources during periods with low demand - Introduce a buffer queue - Use incentives to shift customers from high-demand hours to low-demand hours - Allow customers to execute some parts of the process by themselves
7. References **7.1. References to BPIC submissions**	- Adriansyah, A., & Buijs, J.C.A.M. (2012). *Mining Process Performance from Event Logs.* Retrieved from: https://www.win.tue.nl/bpi/lib/exe/fetch.php?media=2012:adriansyah.pdf - Brils, J.H.H., van den Elsen, N.A.F., de Priester, J., & Slooff, T.A. (2018). *Business Process Intelligence Challenge 2018: Analysis and Prediction of Undesired Outcomes.* Retrieved from: https://www.win.tue.nl/bpi/lib/exe/fetch.php?media=2018:bpi2018_paper_15.pdf - Pakileva, A., Skvortsova, E., Zakoryuchkin, N., Tsaplin, S., & Zarubin, V. (2020). *Business Process Intelligence Challenge 2020: Investigating Business Trips Arrangement Process at the Eindhoven University of Technology (TU/e).* Retrieved from: https://icpmconference.org/2020/wp-content/uploads/sites/4/2020/10/ICPM_2020_paper_124.pdf
7.2. References to the academic literature	- Premchaiswadi, W., & Porouhan, P. (2015). Process modeling and bottleneck mining in online peer-review systems. *SpringerPlus*, *4*(1), 1-18. - Caballero-Hernández, J. A., Dodero, J. M., Ruiz-Rube, I., Palomo-Duarte, M., Argudo, J. F., & Domínguez-Jiménez, J. J. (2018, September). Discovering bottlenecks in a computer science degree through process mining techniques. In *2018 International Symposium on Computers in Education (SIIE)* (pp. 1-6). IEEE.

II. Template improvements elicited from interviews, with prioritization

Specific template / general	Improvement description	Occurrence	Was implemented during this study?	Comments on prioritization
General	Remove either expected output or output	2	No	Low-effort improvement; it should not be implemented as both expected output and output bring value to the templates and provide slightly different information
General	Add explanation on how to define the threshold (for example, what activities are the lowest)	1	Partially	Medium-effort improvement; it is difficult to describe how to define the threshold because it differs based on the event logs and aims of the research. In the introduction was added a note that this should be defined by the analyst based on his/her needs
General	Enlarge screenshots	1	No	Medium-effort improvement, based on the observations, several study participants had problems due to the small size of the screenshots. Enlarging was not implemented due to time constraints as this will require changing the outline of the templates. It can be implemented as a part of future work
General	Add information about arcs and nodes	1	Yes	Low-effort improvement, this information was added to the introduction section
General	Add introduction	5	Yes	Medium-effort improvement; it was implemented because many participants said that it will be beneficial to have an introduction before the templates

General	Add explanations about the functionality of all the tabs and windows of Apromore	1	No	Medium-effort improvement, it should not be implemented because the prerequisite of using the templates is basic knowledge about all tabs and windows of Apromore. Templates are not intended to be a guideline on how to use Apromore itself
General	Add more screenshots	2	No	Medium-effort improvement; it was not implemented as currently the majority of the steps have screenshots and because of time constraints. As a part of future work, screenshots can be reviewed and added, where it is relevant
General	Add grouping in the table of contents	3	Yes	Medium-effort improvement; it was added because several study participants mentioned that they are expecting to have grouping. Also, based on the observations during the interviewees, most of the participants were struggling to find the template needed (when table of contents was not grouped)
General	Visualize redesign possibilities	1	No	High-effort improvement; it should not be added as detailed information about redesigns is out of the initial scope of the study
General	Provide bigger, more self-explanatory examples	2	No	High-effort improvement; it was not added due to time constraints. It can be implemented as a part of future work
General	Change highlightings in the screenshots (for example,	1	No	Medium-effort improvement; it should not be implemented as this issue was mentioned

	highlight only buttons)			by only one interviewee and none of the rest of the interviewees experienced any problems with interpretation of highlightings based on the observations
General	Add hyperlinks to the resources	1	No	Low-effort improvement; it was not implemented due to time constraints and low prioritization as the design of the list of resources is not the core value of the templates. It can be implemented as a part of future work
General	Instead of academic papers, provide links to some documents where IOs are described in more detail	1	No	High-effort improvement; it was not implemented because such documents were not identified during the literature review. But as part of improvement after the interviews, links to the sources based on which templates were created were added to the file. These links provide more information on improvement opportunities
General	Add explanations about what is the danger of IOs, and why they should be redesigned	1	No	High-effort improvement, it was not implemented because such information was not in the original scope of the study, but it can be implemented as a future work
General	Add explanations about what are the benefits of process redesigns	1	No	High-effort improvement; it should not be added as detailed information about redesigns is out of the initial scope of the study
General	Add more explanations about how to interpret each step	1	No	Medium-effort improvement; it was not implemented due to time constraints. Can be added as a part of future work

General	Add ID numbers to templates	1	No	Low-effort improvement; it will not be implemented because of the lack of argumentation. Potentially, can be added when the number of templates is higher or there will be practical need
General	Make result of the step as a separate column	1	No	Medium-effort improvement; it will not be implemented because of the lack of space in the templates. Implementation will require full restructuring of the file. Plus, this was mentioned only by one interviewee
Frequent handovers	Improve spacing in Step 3	1	No	Low-effort improvement; it will not be implemented because this step is already divided into sub-steps that allows easier navigation. but the same spacing is applied to all the templates. Plus, this was mentioned only by one interviewee, the rest were able to identify improvement opportunity without difficulties caused by the non-convenient spacing based on the observations. Additionally, changing the way instructions are written in one template will instigate the change in the rest of the templates to keep templates harmonized
Similar process variants	Formulate instructions in symbols, not sentences	1	No	High-effort improvement, changing the way instructions are written in one template will instigate the change in the rest of the templates to keep templates harmonized. Plus, this

				was mentioned only by one interviewee, others were able to find improvement opportunities using instructions written in sentences
Rework	Add another option to identify rework (using Rework filter)	1	No	Medium-effort improvement; it was not implemented due to time constraints. Can be added as a part of future work
Rework	Check self-loops using BPMN model	1	No	Medium-effort improvement; it was not implemented due to time constraints. Can be added as a part of future work
Bottleneck	Add option to find bottlenecks using dashboard	1	No	Medium-effort improvement; it was not implemented due to time constraints. Can be added as a part of future work
High process complexity	Use numeration for sub-steps	1	Yes	Low-effort improvement, it was implemented to be aligned with the rest of the templates
Activity variants	Add information about how to open dashboard	1	Yes	Low-effort improvement, information about that was added to all the templates that require opening the dashboard. Additionally, it was described in the introduction. This suggestion was prioritized because many participants struggled with opening the dashboard (based on the observations)
Activity variants	Replace word "activity statistics section"	2	Yes	Low-effort improvement, as two interviewees mentioned it, additional explanation was added about how to find the required table
Activity variants	Add instructions on how to define the reason why	1	No	High-effort improvement, adding such an explanation to one template will trigger the

	resources have high duration			need to add it to all the templates in order to keep them harmonized. Such addition to the templates was not in the original scope but it can be implemented as a future work
Activity variants	For steps that require repetition, mention it at the beginning, not the end of the step	1	No	Low-effort improvement; it will not be implemented because it was mentioned by only one study participant, based on the observations the current description was understandable to other interviewees
Activity variants	Replace phrase "extreme cases"	1	Yes	Low-effort improvement; it was implemented to increase the understandability of templates as this phrase can be ambiguous

Note: Suggestions for templates' improvements are prioritized and marked by different colors. Green – ideas that should be implemented, amber – suggestions that are nice to be implemented, red – ideas that should not be implemented.

III. License

Non-exclusive licence to reproduce thesis and make thesis public

I, Nadiia Danylyshyn,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

of my thesis

Analysis templates for identifying improvement opportunities using Apromore,

supervised by Fredrik Milani and Katsiaryna Lashkevich.

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **08.08.2022**