Workload analysis based on event logs

Jia Zheng Alice Lin

University of Twente
Faculty of Behaviour Management and Social Sciences

Supervisory committee:
First supervisor: dr. M. de Visser, University of Twente
Second supervisor: dr. M.L. Ehrenhard
Faculty of Behavioural, Management and Social sciences (BMS)
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
Date 30-06-2021

Acknowledgement

This thesis would not have been possible without the inspiration and support of a number of wonderful individuals - my thanks and appreciation to all of them for being part of this journey and making this thesis possible. I am forever thankful to my supervisor Dr. M de Visser. Without his patients, enthusiasm, encouragement and support this thesis would hardly have been completed. I express my warmest gratitude to my second supervisor dr. M.L. Ehrenhard and writing advisor E.M. Kampen for their advice and expertise.

I would also thank my friend Digdem and my partner Hao you both represent my past and future. Finally, my deep and sincere gratitude to my family for their unconditionally love, help and support. I am grateful to my wonderful brother Alex for always being there for me as a friend. I am forever indebted to my mother for giving me the opportunities and experience that have made who I am today. This journey would have not been possible without them and I dedicate this milestone to my supervisor, advisors, friend, partner and family.

Keywords

Process mining, event logs, process discovery, business process, business process mapping, business productivity

Confidentiality note: Due to the potentially sensitive nature of information, the financial institution at which the research was conducted is kept anonymous. Divisions, departments or specific processes are also not named for this reason.

1. Introduction

In the last decade, data has become an essential type of source in various organizations (Robert Jacobs and Weston, 2006) . However, the explosive growth of business and databases has begun to far outpace our ability to interpret and digest the data (Soibelman & Kim, 2002). The increase in data collection comes from various reasons. One of the main reasons is that business invest in information systems (Syed et al., 2019), also known as enterprise resource planning (ERP) systems These systems support organizations with day-to-day business activities such as accounting, human resource, supply chain, etc. (Robert Jacobs and Weston, 2006). Due to heavily investing in ERP systems, human activities have reduced and are taken over by computers (Mabert, Soni & Venkataramanan, 2003). Computers provide support in daily business activities, but they also play a significant role in collect and store data to a database (Davenport, 1998). Event logs of information systems are such data sources that contain an unexploited reservoir of knowledge about the way employees manage every-day business transactions (De Weerdt, De Backer, Vanthienen & Basesens, 2012). Analyzing these event logs is a promising way of acquiring insights into the semantics of real business processes (Dustar et al., 2005). However, the problem is that many organizations neglect to explore and use their data (van der Aalst, 2012). This is also the case for the one of the financial institution that operates in the Netherlands.

The financial institution is a global audit, accounting and consulting firm. One of its sub-departments operates in the Netherlands. The department optimizes administrative processes, whereas new information systems replace many low-value tasks. The current problem the firm faces is time-to-time high workload peak(s) which led to increased work pressure. Employees claim to work overtime more often than that they were used to. The department assumes that work pressure is high due to the invoice booking process. The invoice booking process is one of the departments' primary task, thus it is the most time-taking activity (In the company summary, the paper will outline the current invoice booking process.) The reasoning behind this assumption is that the workload is not evenly distributed. By exploring the data from the ERP system, the department can analyze the real cause.

This financial institution is one of many organizations that has not taken the opportunity to manage the information embedded in their operating system, analyses data in a way that enhance their understanding and then take changes as an action and response to new evidence and insights.

The data that the firm collects comes from various systems. For this particular research, ERP systems are the sources of data. Pajić & Bečejski-Vujaklija (2016) describe the use of ERP system's data as a bundle of digital traces also known as event log. Event logs could be used for exploring or mapping a business process and answer business questions relating to certain processes' productivity (van der Aalst, 2005). The value of mapping a business process based on event logs is the accurate representation Hysoja et al., (2005) argue that lack of understanding its business process results to poorer business outcomes. The value of answering business question based on event logs provides business' insights based on facts. Therefore, this research has formulated the following research question: "*What insights can be derived from analyzing the invoice booking process based on event logs from an accounting software system?*"

The research domain that is concerned with knowledge discovery from event logs is called process mining (van der Aalst et al., 2007). Process mining emerged as a new research field that focuses on analyzing processes using event data (van der Aalst, 2002). It can be situated at the intersection of the fields of data mining and Business Process Management

(BPM) (van der Weerdt et al., 2012). Furthermore, we point to an important terminology issue concerning process mining and process discover. Process mining describes a family of techniques for extracting knowledge from event logs, whereas process discovery only deals with extracting control-flow model and analyze the productivity. This paper will focus on the process discovery method to explore and analyze the invoice booking process. The outcome of this study could provide the organization a guideline in analyzing business processes based on event logs. It also provides the organization an overview of productivity level within the department.

1.1 Company background information

The department of the financial institution focuses on optimizing the design of their client's financial administration. The department's objective is to work smarter, easier and more efficiently. In doing so, the department provides advice of useful digital tools and extensive support to their customer. During the last few years, the department has invested in an Enterprise Resource Planning (ERP) system and accounting software for the invoicing process. Technology has enabled the automation of invoice processing and has significantly improved the efficiency to execute this process. For instance, the company has improved the feature of scanning and detection. The system here scans the invoice and detects all relevant information that is needed to be registered for the general ledger. In the past, the account payable clerk had to manually enter the information to the general ledger. One of the benefits from automation of invoice processing is the reduction of human error.

The current invoice booking process starts with the clients (a business owner) sending their invoices to the department. Then receives a paper invoice, PDF, or other electronic means. Once the invoice arrives the accounts payable clerk will only need to scan the invoice into the ERP system. The ERP system will capture the necessary invoice data and finally register this into the company's general ledger. These data may include the supplier or clients name, the supplier or clients code, the purchase amount, VAT amount, the description of the purchase, etc. After the data is included, the invoice is coded for the correct account, cost centre or project. To illustrate, a company based on the Netherlands purchases notebooks for a total of 100 euros, including 21% VAT. The cost center is "office supply", and purchase expense is divided into 79 euros' purchase expense and 21 euro's VAT. Finally, it is ready for review and approval from the responsible person or budget owner. Once, the review is accepted, the information will be registered in the company's general ledger.

The ERP system has produced tremendous labour savings in the invoice booking process, reduced human errors and reduction in duplicate invoice payments, which has led to an increase in service quality. Above all, over the years, the ERP system has generated a large amount of data, also known as event log. An event log is a logbook that has captured and registered all activities from an information system. Although the invoicing process has improved a lot, the company's next step wants to answer business questions based on the databases from their ERP system. Based on two meetings conducted in November and December 2019 the management of the financial institution expressed its workload concerns that were raised by complaints from staff within in the department. Staff claims to face time-to-time high peaks of workload which lead to overwork and high work pressure. The department assumes that the invoice booking process is the main cause. The reason behind this notion is that the workload in unfair distributed amongst employees, but also that productivity are uneven spread during the time such as weeks, months, and years. These statements are based on gut feeling and employees' indications.

Since this financial institution has no practical experience and theoretical knowledge of extracting value from event logs. The firm has recruited master students to conduct this research. The students will work 16 hours a week, half a year prior to the start of the research. From here practical knowledge of the invoice booking process will be gained. The reasoning behind this is that the student has practical knowledge of the process and tasks, but also the capability to conduct high qualitative research.

1.2 Management summary

Digitalization has impacted business processes for many organizations, including accounting firms. One of the changes caused by digitalization is the replacement of manual processes with process automation software systems. These systems support employees with daily tasks. A prominent example of this department is the scan and detect system used in the invoice booking process. The accounting software system supports the department with automating financial administration. Usually, a financial administrator manually enters each data one by one. This ERP system automatically scans and detects information from the invoice and is finally booked in the financial ledger. Despite the automatization, the department is still facing a high workload. The complaints the department receives are from financial administrator's whose daily work is to book invoices.

The complaints lead to several management questions. This first management question is: *"What is the current invoice booking business process?"* The second management question is: "*Are there are bottleneck process mining technique can detect in the invoice booking process?"* The third question is: *"Is there a discrepancy in productivity divided among employees?"* The final question is: *"Is there a discrepancy in productivity level between certain periods?"*

To answer the above-stated questions, this research will use the technique of process mining to evaluate event logs from 2017, 2018 and 2019. The invoice booking process during these three years has remained the same. However, the analysis has found an increase in productivity. This means that more invoices were processed during these years via the ERP system. The bottleneck in the invoice booking process is located in the non-main variant path. The activity authorization is the cause of the bottleneck. The analysis has found a discrepancy in productivity among employees within the invoice booking process and in a specific time period.

To conclude, this paper will focus on the possibilities event logs provide for organizations in the financial industry. The research starts with useful literature to understand the basics of process mining techniques and the benefits of process mining. Then the paper describes the methodology for this particular event log. Finally, the conclusion and the managerial implications of the results are highlighted in the final section. In the end, a direction for the management is suggested for future research.

Table of Contents

2. Problem analysis

The objective of this paper is to look for opportunities in event logs that can provide the firm interesting insights. Event logs is from nature a great source to analyze a business process and also to analyze the productivity from various stakeholders. But analyzing event logs is not an approach many organizations choose. One of the reasons lays in preparing data. The size of the event log plays a significant role of the complexity. The other reason is that analyzing event logs with process mining could be seen a new and rising technique. The last reason is that organization are not aware of the possibilities event logs can provide to answering operational business questions. Many organizations do not obtain the knowledge and experience in analyzing and visualizing these datasets. Business' rather use classic techniques including employee interviews and scorecards methods to find bottlenecks and analyze productivity. However, process mining techniques can provide organizations answers to various business questions.

One of the problems the firm is currently facing are the complaints from employees related high workload. Based on the interview with one of the managers in the department, they state that employees feel high work pressure because of work that are unevenly distributed. Work could be unevenly distributed among employees. In this case, one person has more work compared to another peer. Work could be also unevenly distributed through the year. This means that the department experience high workload during certain period of the week, month, quarter or years. The department assumes that the high workload comes from the invoice booking process, this particular process takes up the most time. There are various speculations where the issue could be related to.

This paper provides the management theoretical background about process mining techniques and the use of event logs. To make the theory practical executable, the theory is accompanied by examples. Based on the knowledge gained from the theory and the dataset provided by the organization the methodology was created. Each practical step from data preparation to data analysis are explained. Then, the results of this paper are presented in combination with tables and graphs. Finally, the interpretation of the results are discussed in the conclusion.

3 Literature review: knowledge gap

This chapter focuses on describing existing literature in business processes and its relationship to process mining discovery technique. Gumaer (1996) argues that today's one of the notable sources of competitive advantage is the ability to understand its business process and continuously improve its efficiency. A strong understanding of the subject leads to select an appropriate method that can successfully analyze event logs for the busines.

This first paragraph starts with literature related to business processes. It will create a basic understanding of business process concepts and management, but it also provides an overview of two different business process modelling languages. Since this financial firm and many more organizations create business models based on interviews between different actors involved in the process. It is interesting to consider other methods. Process mining bridges the gap between business process modeling and business intelligence (van der Aalst, 2011). Therefore, the next paragraph describes the possibilities of a new research field that focuses on the analyses of processes based on event logs (Van Der Aalst, 2012). The theory is named process mining. Process mining and its techniques uses event logs to evaluate and analyze a business process. Event logs include specific attributes in order to successfully analyze the data. For this reason, the paper has included the basic concepts and knowledge of event logs including some practical examples on how an event log resembles.

3.1 Business processes

Burratin (2015) mention that activities and tasks firms are required to perform in order to achieve business objectives are become more complex. An approach to simplify the business division of operations, management decided to divide the business activities into smaller entities (Motahari-Nezhad et al., 2010). The performed activities are often repetitive and have several individuals involved. In these events, it is very helpful to identify a standard procedure that everyone can follow. A business process is the definition of such a standard procedure (Aguilar-Savén, 2004). Since there are several definitions of the business process, the two most influential ones will be mentioned in this research. The first presented by Champy and Hammer (1994) state that a business process is: A collection of activities that takes one or more kinds of input and creates an output that is of value to the customer. A business process has a goal and is affected by events occurring in the external world or in other processes (p.11). The second definition of business process is presented by Ould (2007). He describes that business process is seen as something that: (a) contains purposeful activities; (b) is performed collaboratively by a group of people and/or machines; (c) crosses functional boundaries: (d) is invariable driven by the outside world (Ould, 2007).

3.1.1 Business process management

Closely related to Busines Process is Business Process Management (BPM) (van der Aalst, ter Hofstede & Weske, 2003). Business Process Management is the art and science of observing how work of activities is performed in an organization (Dumas et al., 2018). But it is also supporting business processes by using techniques, methods, and software to enact, design, control, and analyze operational processes involving organizations, humans, applications, documents, and other sources of information (van der Aalst et al., 2003). It has received a great deal of attention in recent years due to its potential for substantially increasing productivity and saving costs (Bussler, Jablonski & Schuster 1996). But also, companies have to sustain competitive advantage and outstanding performance in rapidly changing enjoinments. The objective of BPM is to increase the visibility of activities that allows identification of issues that may arise, but also fields of potential improvements and optimization (Weske, 2007). By way of grouping activities (tasks) in 'sections' and grouping the persons in 'roles,' it could define duties more clearly. BPM heavily relies on business process models, which is an overflow of notations that exists to model operational business processes (Weske, 2007). These notations have in common that processes are explained in terms of activities. The ordering of these activities is modelled by describing causal dependencies (van der Aalst, 2016). The next paragraph will focus on the essential concepts of process modeling using the Business Process Management Notation language.

3.1.2 Business process management

Busines process modeling came to light due to the need of a creating better understanding of business process in organization. The objective of business process modelling is to communicate a wide variety of information to different audiences. Process model consists of a set of activity model and execution constraints between them. It is frequently illustrated with activities and events that are associated with control flow. Processes can be modelled with different languages with the help of different tools. This section briefly describes the business process model languages that are organizations uses to construct a process model or flowchart (describing a business process) from event logs. The paper will introduce two models that are often used in process mining. The two types are business process modeling notation and Petri Net (source 7).

3.1.2.1 BPMN

Business Process Model and Notation (BPMN) specifies business processes in a graphical representation. It defines a Busines Process Diagram (BPD), which is a flow chart method that illustrates the steps of a business process from end to end. In order to define a clear business, process the BPD categorize notations so the elements can be easily recognize (White, 2004). The four basic categized notations are flow objects, connecting objects, swim lanes and artifacts. The Flow objects are consisting of three core elements, see table 1.0.

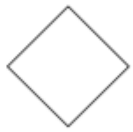
Event	Represented by circles and explain that something “happens. There are three types of events based on when it affects the flow: Start, Intermediate and End	
Activity	Represented by a rounded corner rectangle, it stands for work that a business preforms	
Gateway	Represented by a diamond shape and is used to control divergence and convergence of a sequence flow. It will determine decisions, merging and joining paths. (for example, decisions with yes or no and)	

Table 1.0: BPMN four basic notations

The flow objects are connected together with arrows in a diagram, also known as Connect Objects (White, 2004) BPD includes three connectors which are explained and shown in table 1.1.

Sequence Flow	Represented as a solid line with a solid arrowhead. It is used to show the order (sequence) that activities will be performed in a process.	
Message Flow	Represented by a dashed line with and open arrowhead. It is used to show the flow of message between two separate business entities or business roles that send and receive the message.	
Association	Represented by a dotted line with a line arrowhead. It is used to associate text, data and other artifacts with flow objects.	

Table 1.1: BPMN three connectors

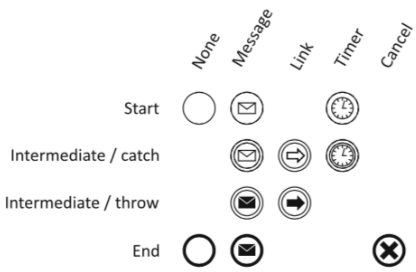
The concept of swim lanes is a mechanism to organize activities into separate visual lanes. It illustrates various functional capabilities or responsibilities. BPMN differentiate two main constructs’ (White, 2004) See table 1.2 for the visualization and explanation.

Pool	Represents a participant in a process	
Lane	Represents two participants in a process	

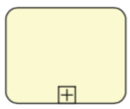
Table 1.2: BPMN swim lanes

3.1.2.2 Petri net

The second type of business process modelling this research discuss is the Petri net. The graphical language was proposed in 1962 by Carl Adam Petri (Petri, 1966). The Petri net is a bipartite graph where the graph consists two types of nodes. A petri net consists of transition, places and arcs. Transitions represent activities that can be executed, and places represents the states that the process can reach. Arcs runs from a place to a transition is also called the input places of transition. Arcs that run from transition to a place are called the output places. Arrows can only connect between components of a different type. For instance, an arrow never joins two circles or rectangles. (Peterson, 1977). A simple example of a Petri net is illustrated in figure 1.1 and 1.2.



(**d**) Events



(**e**) Sub-process

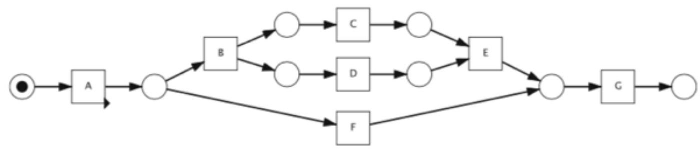


Figure 1.1 (left): Petri net basic notation, Figure 1.2. (right) complete petri net

2.2 Process mining

Many organizations create business models based on interviews between different actors involved in the process. It is interesting to consider other methods. Process mining bridges the gap between business process modeling and business intelligence (van der Aalst, 2011). With process mining it will be possible to detect or identify problems based on facts and not on intuitions or conjectures.

Process mining and its techniques can play an important role in addressing the issue of monitoring operational businesses. There are multiple definitions of process mining. Reinkemeyer (2020) describes the term process mining as "a process management technique that allows for the analysis of business processes based on event logs" (p.3). Whereas Burattin, 2015 refers process mining to a general to take, as input, some event data, and perform a fact-based analysis of process executions. While a variety of definitions of the term process mining have been suggested, this paper will use the definition from van der Aalst (2016). Van der Aalst isalso known as the founding father of process mining. The process mining expert published his first book and defines process mining as: "an emerging discipline providing comprehensive sets of tools to provide fact-based insights and to support process improvements" (p.7). This definition has been broadened to include sets of tools for analyzing and improving business processes. The objective of process mining is to extract an explicit process model from event logs, whereas the challenge is to create a business process model by observing events recorded by some enterprise system (van der Aalst, Weijters & Maruster, 2004)

3.2.1 Process mining perspectives

Dustdar et al. (2005) distinguishes in his study three different perspectivists in process mining. The perspectives are: (1) process perspective, (2) organizational perspective and (3) case perspective.

Process perspective focuses on the ordering of activities. The objective here is to select the correct paths within the process. These paths can be expresses in terms of a process model, which has been described in earlier chapters in this study (BPMN and Petri). The perspective focuses on the "How?" question (Turner, Tiwari, Olaiya & Xu, 2012).

The organizational perspective focuses on the sources within a process, an example could be the people and roles that are involved in a process and how they are related to each other. This approach can be used to find relationships between people in a process in terms of a social network (Dustdar et al., 2005). The organizational perspective focuses on the "Who?" questions (Turner, Tiwari, Olaiya & Xu, 2012).

The case perspective focuses on the characteristics of a case. The perspective takes into account the attributes of a case. An example could be an interaction of another department within the process. The case perspective is concerned with the "What?" question (Turner, Tiwari, Olaiya & Xu, 2012)..

Van der Aalst et al., (2011 W.M. Adrainsyah) adds a fourth process mining perspective in the process mining manifesto, the time perspective. This perspective focuses on the timing of the events. It could lead to discover bottlenecks, monitor resources and predicting the productivity level of a certain task.

3.2.2 Process mining techniques

Besides the three process mining perspectives identified by Dustdar er al. (2005) Van der Aalst (2004) distinguished three main types of process mining techniques.

The techniques are: (1) process discovery, (2) conformance, and (3) enhancement. The discovery technique takes an event log and creates a Petri net, which translates the behaviour recorded in the log. The conformance technique focuses on an existing process model, which compares with an event log of the same process. The technique is used to check if reality, as recorded in the log, conforms to the model and vice versa. The last technique is enhancement. Here the idea is to extend or improve an existing process model. One of the enhancements is repair, where processes are corrected based on a comparison between existing the analysis. These three types of mining can identify different business process perspectives (van der Aalst, 2004). In order to scope the research in more depth, this research will focus on the process discovery technique. The reasoning behind this decision is that process discovery focuses on analyzing business processes based on an event log. The other techniques focus on comparing and improving process models.

3.2.3 Process Mining: Process Discovery

The previous paragraph provided an overview of three different process mining techniques and the research has scoped the focus on the process discovery method. Within the process discovery method there are three classes namely: (1) interview-based discovery, (2) workshop-based discovery and (3) evidence-based discovery.

The interview-based discovery aims at interviewing domain experts to study how the process is executed and the workshop-based discovery method is similar to the interview-based discovery. However, the workshop-based method offers the possibility to provide a profound understanding of a business process from workshops whereas multiple participants are involved at the same time. While the interview-based is conducted with an interviewee and the interviewer. Both methods have to rely on the interpretations and descriptions of the domain experts who are involved with the process. Notably, domain experts might have different perceptions and ideas of how a process operated which leads to the risk that their description might be partially incorrect. Moreover, interview-based discovery requires several feedback iterations and workshop sessions can be difficult to schedule, because various domain experts at the same time a needed. On the other hand, both methods can provide rich insights into the process. Domain experts involved in interviews and workshop are a great valuable resource to clarify reasons why a process is set up as it is.

The evidence-based discovery typically provides the best level of objective and can be distinguished into three sub methods namely document analysis, observation and automated process discovery. First, the document analysis exploits documents that are available and can be related to an existing business process. This includes firms' policies, organization charts, employment plans, work instructions, and handbooks. Although the document analysis provides structured information and are independent from stakeholder availability, documents can be outdated. Then the observation method focuses on following individual cases in order to understand how the process work. It can be executed based on a role play, whereas for example the customer role triggers the process execution. Furthermore, it provides rich insights, but are potentially intrusive and stakeholders are likely to behave not as usual.

Finally, the automated process discovery (ADP) is a method that uses event logs retrieved from information systems. The captured event data can generate a formal model of a business process. Despite that event logs can provide valuable information of the business process; data may not be always available or only available in certain business process parts. So far, the purpose of this research is to understand the business process based on event logs. For this reason, this research will further scope the subject and focus on the automated process discovery.

3.2.4 Automated Process Discovery – geef een paar voorbeelden het is te technisch

The automated Process Discover is one of the most widely investigated process mining operations. The technique takes as input an event log and produces the output a process model that visualizes the behaviour of the log (Dumas et al., 2018). There is a wide range of ADP techniques that are supported by commercial process mining tools and open-source process mining toolset (Wen, van der Aalst, Wang & Sun, 2007). (The next paragraph will discuss different process mining tools.)

Dependency graph is one of the ADP technique. It produces a simple albeit rather than an incomplete representation of how activities in the process follow each other. Each node in a graph represents one event (tasks or activity). An arc between these activities directly follows relations. An arc exists between two event classes A and B. The graph may be annotated with an integer, which indicated time spend on an activity A directly follows by B. In some process mining tools, it is also possible to indicate the time has taken between the two activities. For example, the time between activity A and B is, on average, 2 min and 17 seconds. Owing to their simplicity, dependency graphs are supported by two open-source mining toolsets which are: ProM and Apromore. Both tools provide visual signals to enhance the comprehensibility of dependency graphs (Mans, van der Aalst & Vanwersch, 2015).

2.2.5 Process mining software tools

There are various types of process mining tools on the market. Augusto (2019) summarized 35 primary studies in the field of automated process discovery. His study evaluates process mining methods based on studies published in 2011 or later. Earlier studies have been reviewed and evaluated by de Weerdt, De Backer, Vanthienen and Baesens (2012). The 35 studies use three process mining framework tools: Standalone, ProM and Apromore. Tabel 2 summarizes the information on those tools.

	Standalone	***ProM***	***Apromore***
Accessibility	Commercial	Open-Source	Open-Source
Operating system	IOS and Microsoft	Microsoft	Mac and Microsoft
Model language	Petri net, Declare, WoMan, Directed acyclic graphs, BPMN	Declare, Process trees, Petri nets, BPMN	BPMN, Causal nets

Table 2: Process mining tools

3.3. Event Logs

Each dataset consists of different attributes and information. This depends on how and where the data is collected but also what the objective was to collect the data. The source of event logs comes from information systems. Event logs focuses on the entire population for a certain business process. Usually, a dataset contains data based on a sample size from a population. Another difference between event log and normal dataset is origin. Event logs are automatically saved data, whereas a dataset that is prepared for a research is collected by the researcher. It includes specific attributes in order to successfully analyze the data. The following paragraph focuses on the basic concepts and knowledge of event logs including some practical examples on how an event log resembles.

3.3.1. Description event logs

Chiu and Jans (2019) describe that the event log is a bundle of digital traces that are automatically and chronologically records the actions in the system. Dumas et al. (2018) describe that event log is a collection of timestamped event records. Every event recorded describes us something about the execution of a task, process, or it describes us other relevant events that could have occurred in the context of a given case in the processes (Jans et al., 2010). Moreover, these logs include information about the activity executed for each case, including where it was executed and by whom. Certain systems also contain information about the user entered for each activity. Unfortunately, this data is not actively explored by the organization to analyze the underlying processes (Jans et al., 2011).

Chiu and Jans (2019) and Reinkemeyer (2020) suggest four attributes that are essential for analyzing the information from event logs in the system. These attributes are (1) activity (event), (2) Process Instance (3) Resource, and (4) Timestamp.

The activity or event represents the recorded transaction and example could be: “received” and “booked” in the invoice booking process. Process instance is the unique case (ID number) that is used throughout the process, an example could be the invoice number, or identify case number in the booking process. Resource is responsible for the activity; this is also called as originator or action owner. The resource is often a person who conducts the activity, for example Louis entered the invoice, then he tagged the invoice, added relevant description and finally booked in the invoice. The timestamp refers to when the event had occurred often outlined in year-month-date and time (Jans et al., 2011). In practice there may be other event variables, such as for example the costs and domain-specific data such as supplier and customer data.

Event logs can be very useful to analyze a business process or business productivity, however it is important that all four attributes are included. Furthermore, event logs could be delivered in various formats. Depending on the business question, it is often required to prepare the data. For example, event logs could be delivered without aa unique case ID, in this case the data analyst adds this attribute.

Description

3.3.2 Practical explanations and examples

The following section briefly describes how event logs are generated and prepared. To begin with, this research will use an invoice as an example (see Fig.2). Figure 2 illustrates information that is entered into an ERP system, which automatically translate the input in an event log. On the left-hand side column (Input data) shows the information about the invoice which have been entered by an individual. The system recognizes the information that has been entered and record all the necessary data such as: invoice number, the supplier, the invoice date, ledger number, description.

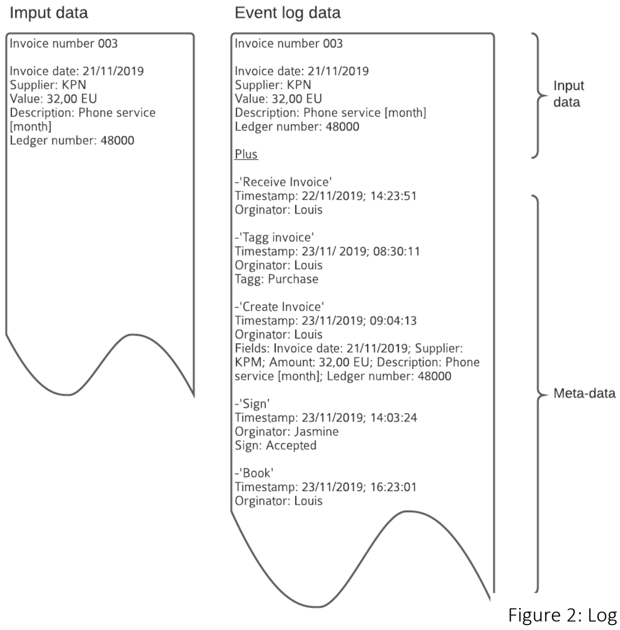


Figure 2: Log

The right-hand side column (Event Log data) represents information that are stored in the event log from the same invoice. Each and every input will be accurately recorded by the information system (Jans et al., 2013). Meta-data enables the firm to reconstruct the record of a transaction by identifying the relationships between various transactions and actions stored in the database. The various actions stored in the database could be used to identify changes in transactions, errors and sequences of processes. Figure 2 seems to look like a ledger of information and based on the information a story can be told. The KPN invoice of 32 euro's is received on the 22/11/2019 at 14:23:51 and delivered by Louis. The next day Louis tag and create an invoice. The next day Jasmine signs and approve the process and finally Louis book the invoice to the company's general ledger.

3.3.3 Case variants

While the example illustrates one of the booking processes, there are various booking combination possible. Depending on several reasons business processes could differ from each other. Event logs are captured as a list in a tabular format. Simple event logs are usually represented as tables and stored in a Comma-Separated-Values (CSV) format. Despite that, in more complicated event logs, where the activities obtain data attributes (e.g. invoice number, amount of invoice payment/receivables, supplier and buyer), a flat CSV file is not an adequate representation. A more extensible file format for exchanging and storing event log is the eXtensible Event Stream (XES) format standardize by the IEEE task Force on process mining.

Table 3 shows an in-depth example of a variant and four corresponding process instances. All processes that have identical routing will be grouped into the same variant. Process instance that are distinct by other routing will be classified into different variant. By examining case variants, firms are able to distinguish between standard and non-standard routings that occurred in the business process (Dumas book). The standard variant is the process instances that have paths that are aligned with the firm's standard business processes. While the non-standard variant has paths that deviate from the standard business process. In regard of the standard variant and non-standard variant, firms could evaluate employees task efficiency, but also detect bottlenecks and tasks patterns.

Using Table 3 as an example, it demonstrates variant one and two. A standard procedure of invoice process is "Receive invoice – Tagg invoice- Create Invoice- Book invoice".

Whereas a non-standard procedure of invoice process is “Receive Invoice – Tag invoice - Create Invoice- Sign Invoice - Book invoice”. Therebetween the firm could argue why the “Sign Invoice” is missing in the standard variant procedure. This means that the standard procedure of process is finalized without the need of an authorizer approval. Since this routing saves time, it also indicates that the variant allows one resource to take full control of the invoicing process. Overall event logs in many forms of format could provide a business a in depth look into their business processes, however the format should be an adequate representation alongside firms have to overcome challenges for log data extraction.

TABLE 3
Example of an event log

Variant	CaseID	Timestamp	Activity	Resource
1	2019110012	05-11-2019 14:08:56	Receive invoice	Joyce
1	2019110012	07-11-2019 11:53:01	Tag invoice	MZ 870911
1	2019110012	07-11-2019 11:55:38	Create invoice	MZ 8709
1	2019110012	08-11-2019 09:30:01	Sign/Authorize	Joyce
1	2019110012	11-11-2019 08:21:20	Book	MZ 87091
2	2019110008	02-09-2019 16:00:01	Receive invoice	MZ 15879
2	2019110008	02-09-2019 16:10:51	Tag invoice	MZ 15879
2	2019110008	02-09-2019 16:32:18	Create invoice	MZ 15879
2	2019110008	02-09-2019 16:32:59	Book	MZ 15879
2	2019110008	08-10-2019 11:53:10	Receive invoice	Marit
2	2019110008	10-10-2019 15:08:11	Tag invoice	MZ 77135
2	2019110008	10-10-2019 15:17:23	Create invoice	MZ 77135
2	2019110008	10-10-2019 15:18:24	Book	MZ 77135
1	2019110032	16-10-2019 12:01:59	Receive invoice	MZ 73625
1	2019110032	16-10-2019 12:25:08	Tag invoice	MZ 73625
1	2019110032	16-10-2019 12:26:01	Create invoice	MZ 73625
1	2019110032	20-10-2019 10:53:57	Sign/Authorize	Anthony
1	2019110032	21-10-2019 09:01:22	Book	MZ 73625

3.3.4 Event log extraction challenges

Dumas book (2018) identifies four major challenges for log data extraction. The first challenge is the correlation challenge, which refers to the issue of identifying the case an event belongs. Some enterprise systems lack an explicit notion of process defined. In other words, recorded digital traces are correlated among each other, however the traces need a unique number in order to be identified. The second challenge is the timestamping challenge, caused by the fact that many systems do not consider logging as a primary function or only contain finished timestamps (Reinkemeyer, 2020). Which means that logging is often delayed until the system has inactive time or little load and causes sequential events with the same timestamp in the log. The third challenge is longevity challenge, whereas processes with longer cycles are not observed. In other words, some cases are incomplete or still running and for this reason it could not be observed. Solution to this problem could be excluding incomplete cases from an event log. The last challenge is the granularity challenge, which is the scale of level of detail presented in the event log. Generally, the granularity of event log recordings is finer and therefore record each task of process (Dumas Book 2018). Overall, it is important to overcome these challenges and produce and record high level of data quality.

3.4 Conclusion practical contribution

Process management has become more important during the last decades (Dusdat et al., 2005) To increase their competitiveness, many organizations including this accounting firm, have to introduce clearly defined business processes and these processes must be improved continuously. In this chapter, the study has described the importance of a business process and introduced two process model languages. BPMN and Petri net are the most used process mining models.

The chapter also discussed the latest theoretical development in the mining of business process. Process mining can play an important role in addressing the issue of monitoring operational business. It can be used as a tool to discover how people enact processes in the real world (Wen et al., 2009). Dustdar et al. (2005) distinguishes three different perspectives in process mining: process perspective, organizational perspective and case perspective. Van der Aalst et al, (2011) includes the fourth perspective the time perspective. Aalst (2004) distinguished three main types of process mining techniques.

The techniques are: (1) process discovery, (2) conformance, and (3) enhancement.). In order to scope the research in more depth, this research will focus on the process discovery technique.

Within the process discovery method there are three classes namely: (1) interview-based discovery, (2) workshop-based discovery and (3) evidence-based discovery. The interview-based discovery aims at interviewing domain experts to study how the process is executed and the workshop-based discovery method is similar to the interview-based discovery. The evidence-based discovery typically provides the best level of objective and can be distinguished into three sub methods namely document analysis, observation and automated process discovery. The purpose of this research is to understand the business process based on event logs and the automated process discovery is a method that uses event logs to retrieve relevant insights from a business process.

Event logs is a bundle of digital traces that are automatically and chronologically records the actions in the system (Chiu and Jans, 2019) Chiu and Jans (2019) and Reinkemeyer (2020) suggest four attributes that are essential for analyzing the information from event logs in the system. These attributes are (1) activity (event), (2) Process Instance (3) Resource, and (4) Timestamp. Event logs can be very useful to analyze a business process or business productivity, however it is important that all four attributes are included. In doing so, Dumas (2011) identified four major challenges for log data extraction. The challenges are correlation challenges, timestamping challenges, longevity challenge and granularity challenge. Overcoming these challenges can lead to provide high qualitive analyses.

4. Research methodology

The objective of this paper is to provide relevant business insights related to productivity based on event logs. The current problem the management is facing are the complaints from employees that experience high workload. The complaints are based on personal experience and are not supported by facts or evidence. The invoice booking process, is one of the main tasks for the department. Employees work with accounting software system that collects data (event logs). Analyzing event logs can provide the management a deeper insight of what the reality is based on data.

In order to answer the management questions the methodology for this research is separated into two parts. The first part (knowledge collection) is the collection of knowledge from various existing literature on this research subject. It provides the researcher and the management the basic knowledge of this field. The objective here is to describe the importance of business processes, existing business process languages but also discuss the theoretical development within the process mining of business process.

The second part is the preparation and the method of analyzing the event logs from 2017, 2018 and 2019 provided by the accounting firm. Figure 3 visualize the steps it takes to describe the two methodical parts. The section provides practical insights on how data are prepared and analyzed.

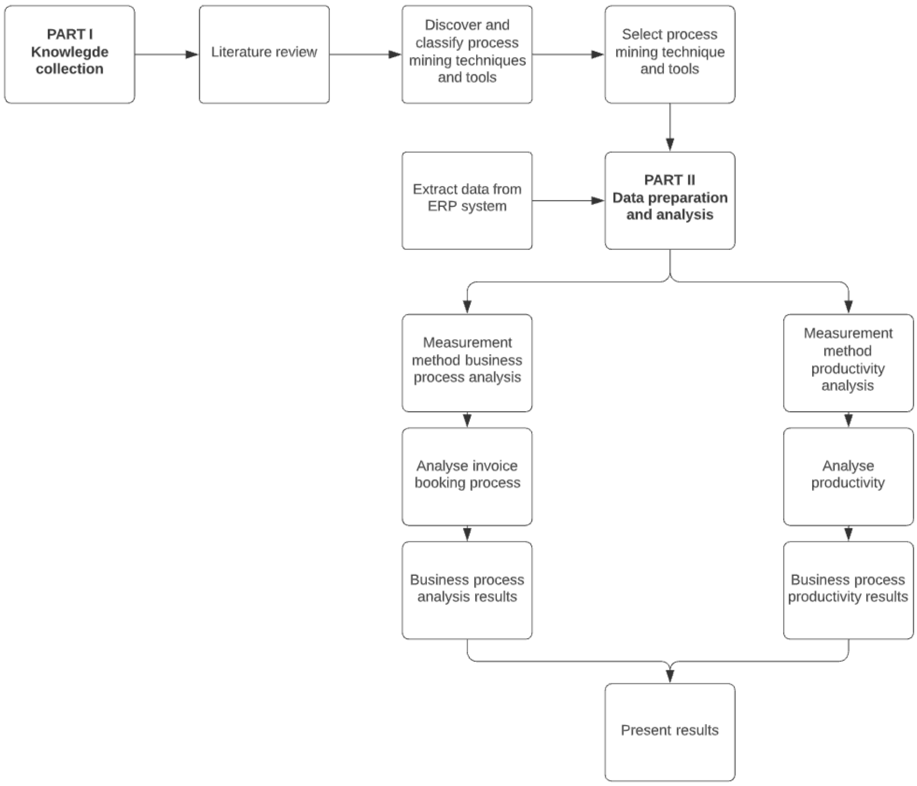


Figure 3

4.1 Knowledge collection

The theoretical framework contributes major support and guide of structuring this research. It is important to select the correct literature review method. here are two methods to write a literature review. The first method comes from the paper Eisenhardt (1998), where the paper describes the process of inducting theory by using case studies. The second method is from Webster and Watson (2002) provide advice, including a framework to prepare writing a literature review in the field of Information System (IS). Since the paper objective is to test the existing theory, Webster and Watson (2002) will contribute more to the research.

The review will examine relevant studies from a broad range of academic journals that explicitly incorporate theories, constructs, and context from Information System research, Business research, Management Technology, and Computer Science. Based on the literature review findings, the research will build a strong foundation of advanced knowledge. Webster and Watson (2002) state that the information systems (IS) field are less published review articles. As a result, our field's progress is impeded; specifically, its complexity of assembling a review of the IS field has been challenging because we often need to draw theories from a variety of fields (Webster & Watson, 2002). The main reason we see so few theoretical articles in the field of Information System relates to the field's youth. To identify relevant literature, we will use the framework of Webster and Watson (2002). The framework consists of a structured table, including papers based on the inclusion and exclusion criteria shown in Appendix 2.

This research implemented three criteria for the literature review. First, the literature review will limit the search to peer-reviewed academic journals and exclude unpublished materials. Second, to ensure the article's quality, we consider high cited papers and exclude papers that are not cited. The numbers of citations counted are based on databases from Sciencedirect, Web of Science, University of Twente library and Google Scholar. However, there are two exceptions made. Since the Information system field is modern, recent papers from 2000 and 2020 will not fall under these criteria. But also, research papers that are not highly cited but are acknowledge in highly cited papers. Third, only articles written in the English language will be examined. The reason for this is to reduce the risk of misinterpretation of other languages.

4.1.1 Theory selection

The theoretical contribution started with conceptualizing the central research question and understanding the current problem. The current problem is the knowledge gab and practical experience in the field of analyzing the business process based on event logs. Therefore, the paper starts first with understanding the basic knowledge of business processes and how to manage processes within an organization.

Current literature and publishers promote process mining as a new tool to improve, control, re(design) and support of the business operating processes (Gartner, 2008). The method process mining aims to discover, monitor and improve real processes by extracting knowledge from event logs retrieved from information systems (van der Aalst, 2011). Professor Wil van der Aalst is one of the founding fathers of process mining. He published his first book on process mining in 2011. A book that fills in the knowledge gab between business process modeling and business intelligence. Process mining is a relatively new but useful method (tool) for many organizations. For this reason, this paper will focus on the implementation methods and the outcome of process mining.

4.1.2 Selection Business model language

A business process model can be expressed in a form of a graph and visualizing an order of process tasks and can be expressed in languages such as BPMN and Petri nets. Choosing an appropriate language requires consideration, because not every modelling language is suitable for all aspects of the process (Leemans, Fahlan & van der Aalst, 2016). Previous chapters have mentioned The BPMN and Petri net model languages see section 2.3.1 and 2.3.2. The BPMN and Petri net are the most common process model language for process mining. The difference between BPMN and Petri net is that the Petri net consists three static elements and is therefore not suitable to discover a process, BPMN is more focused for process discovery (van Der Aalst, 2012). Augusto (2019) argues that Petri nets is the predominant business modeling language. However, in their research they have found that more recently the appearance of methods produce model in BPMN, a language that is widely used standard process modelling (Agusto, 2019). It is also generally easy to comprehend and is highly understandable for business and technical personnel (Leemans, Fahland & van der Aalst, 2016).

4.1.3 Selection Process mining technique

There are three main types of process mining techniques: the process discovery, conformance and enhancement. The research objective is to analyze a business process based on event logs. The conformance and enhancement techniques focus on monitor and improve business processes. The process discovery technique focuses on analyzing the current business process and its activities. This technique is the most suitable for this research because it supports the paper to analyze the invoice booking process based on event logs. The other techniques could be interesting for other studies.

4.1.4. Process mining tool selection

The process mining tool selection is based on three criteria. All three criteria are equally important. The first criterion is the accessibility of the tool. The paper prefers a non-commercial tool, the benefit here is to experiment a tool before purchasing, which saves the firm expenses. The second criterion is the operating system availability. This research is conducted on a MacBook with a IOS operating system, but the firm uses Windows as their operating system. Therefore, the paper suggests a tool that can run on both operating systems. The last criterion is that the tool uses the BPMN model language. Based on the three criteria, Apromore fits the most.

4.2 Data preparation and analysis

In the following paragraphs, focuses on the data preparation and analysis methods. Figure 3 describes the roadmap with its starting point "receive event logs" that eventually leads to "conclusion".

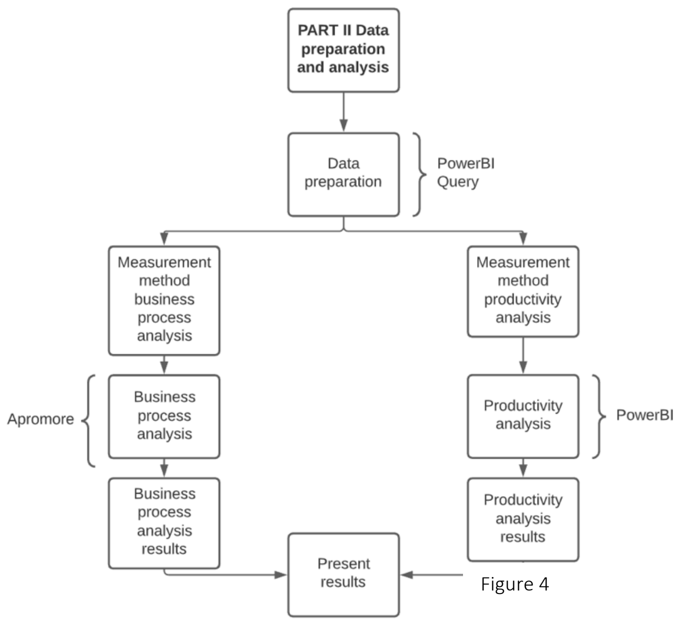


Figure 4

4.2.1. Data tools selection

Frequently, datasets are not delivered in the perfect form and data preparation is needed. Before analyzing the data, we start with selecting a data preparation tool. There are various tools on the market to prepare raw data. The datasets for this research are delivered in .xlsx formats. Since excel was made as a spreadsheet application with the function of keeping all the data sets separate, it has limited capacity per sheet and can become unmanageable as the volume of the data stored grows. In this research, the dataset will be used to analyze a business process and the productivity of the department. The business process analysis will be done with the process mining tool Apromore.

The limitation here is, that current process mining tools solely focuses on presenting a business process. The paper needs to select a tool that could prepare and visualize the datasets.

There are various database management tools on the market. To list a few: R studio, MySQL, Oracle and PowerBI Query. All these database tools require somewhat coding language experience. However, there is a difference between levels of coding language experience. Data management and analysis were performed using Microsoft Power BI. The main reason for this decision is that the financial institution has prior experience in using PowerBI to visualize and PowerBI Query to prepare data. Furthermore, the company provides online courses to all employees within the firm.

4.2.2. Data preparation

The event logs are retrieved from the Enterprise Resource Planning system. Collecting event logs required permission from the ERP system developer. For this research, event logs from the years 2017, 2018 and 2019 will be analyzed. The reason to use multiple years is to indicate if patterns and processes return in sequent pattern(s) each year. Another reason is to detect whether changes in patterns and process occur. In order to conduct the analysis, the datasets will be first prepared. The formats for all three years are unique, which mean that the preparation steps for each event log will be exactly identical.

The first step in this process was to start removing variables with sensitive client data. The dataset contains in total of 46 attributes, but not all attributes will be used to conduct the research. The research excludes variables with sensitive client data, such as client name, company name, email-address and document name. Instead of displaying the name to recognize the resource, the data set will use a unique ID for each case and change employees' number into a random numeric. The reason to exclude this information is related to the European regulations of data protection, including consumer privacy. Therefore, companies should comply with the General Data Protection Regulation, and data can only be used when the person has given their consent. In total 24 attributes will be deleted, which leaves 22 attributes available.

The value in the row contains information from non-department customers. Since the paper conducts a research for one particular division of the accounting firm, the non-department customer will be executed. In doing so, the department has delivered us a list of names from the clients. This dataset includes 300 customer names. We kept in mind that not all customers were and are using the accounting software system and the outcome of rows will be significantly less. The department's list and the three datasets are individually merged. The outcome of total rows is displayed in table 4:

	Objects (rows) before merge	Objects (cases) after merge
Audit file 2017	206.837	63.188
Audit file 2018	262.242	75.196
Audit file 2019	282.548	87.758

Table 4: objects merged for all three years

Based on table 4, the population of interests (also known as cases) are 63.188, 75.196 and 87.758. The cases represent the productivity outcome from the employees working for the accounting firm and the firm's clients. We point to an important difference here between the accounting firm (their employees) and their client. Some clients prefer to perform its own invoice booking process but makes use of the ERP system. The reasoning behind this is to reduce accounting expenses. Other clients outsource the service, whereas the accounting firm will perform the tasks. This research could measure the productivity for the accounting firm but also compare the productivity among their clients. Table 5 shows the difference of total objects between the accounting firm and their clients.

	Firm's objects (cases)	Clients objects (cases)	Total objects (cases)
Audit file 2017	45.574	17.614	63.188
Audit file 2018	54.803	20.393	75.196
Audit file 2019	59.631	28.127	87.758

Table 5: objects divided in sources

Once the population is extracted, the next step is to ensure that each process procedure consists of all essential attributes (Reinkemeyer., 2020) (Dumas et al.,2018). The four attributes are:

(1) Activity (event),
(2) Process Instance (Case ID),
(3) Resource, and
(4) Timestamp

The event log from the firm presents a complete invoice booking process in one row. This means, that one complete row consists of all activities, resources and timestamps. The goal here is to classify and group the activity, resource and timestamp. Each process activity should be listed under each other based on the time order. Table 6 explains a small part of the current event log format:

Activity	**Source**	**Timestamp**	**Activity**	**Source**	**Timestamp**	**Activity**	**Source**	**Timestamp**
Receive invoice	Firm 1	01/01/2017 06:01	Tag invoice	Firm 1	01/01/2017 08:01	Book invoice	Firm 1	01/01/2017 08:01
Receive invoice	Client ABC1	02/01/2017 07:11	Tag invoice	Client ABC1	01/01/2017 08:53	Book invoice	Client ABC1	01/01/2017 09:02

Table 6

Since the tables does not consist of unique Case ID's, the following step is to create Case ID's for each objects of interest. The Case ID's number is equal to the total object of interest. Table 7 shows an example of one complete activity.

ID	**Activity**	**Source**	**Timestamp**	**Activity**	**Source**	**Timestamp**	**Activity**	**Source**	**Timestamp**
1	Receive invoice	Firm 1	01/01/2017 06:01	Tag invoice	TKH01	01/01/2017 08:01	Book invoice	Firm 1	01/01/2017 08:30
2	Receive invoice	Client ABC1	02/01/2017 07:11	Tag invoice	Client ABC1	01/01/2017 08:53	Book invoice	Client ABC1	01/01/2017 09:02

Table 7

The next step is to enlist each activity including the four attributes under each other. Table 9 demonstrates the ideal format. The ideal format is one of the requirements for the process mining tool Apromore. Furthermore, this format also provides a clear overview of each activity that has taken place and also detects missing values in the event log.

ID	Activity	Source	Timestamp
1	Receive invoice	Firm 1	01/01/2017 06:01
1	Tag invoice	Firm 1	01/01/2017 08:01
1	Book invoice	Firm 1	01/01/2017 08:30
2	Receive invoice	Client ABC1	02/01/2017 07:11
2	Tag invoice	Client ABC1	01/01/2017 08:53
2	Book invoice	Client ABC1	01/01/2017 09:02

Table 9

Data preparation is an essential involvement for this research. It creates consumer protection through deleting all sensitive consumer information; solely focuses on a particular population; ensures the correct attributes are included and finally structure the right format. The next step is to describe the measurement methods for the business process analysis and the business productivity

4.3 Process mining tool - measurement method

The process mining tool supports the analysis of the invoice booking process based on logs. The tool includes various statistical measure options including the mean, median, min and max. It also presents the business process with main variant and non-variant options.

The pm tool separates the statical measure into two overlays. The first overlay focuses on the case frequency, the second overlay measures the duration between each activity. The measurement for the frequency is calculated based on the sum of each individual activity for the reason that it creates and evident overview on where most frequent activities were performed.

The case duration measures the time it took between activities. It could measure the case productivity, but also detect bottlenecks. The measurement for the case duration will be measured based on the median. The median measures the middle number by separating the higher half from the lower half of a population. Other measurements are easily influenced by outliers, whereas the median exclude potential outliers. Considering that the dataset consists of the accounting firm and their client source, the research will compare the productivity amongst these two sources.

The tool can provide a very detailed business process. It can include or exclude non-variants by hiding some of the most infrequent arcs and creates a more readable map. This research has decided to include the entire business process tables in the Appendix and highlight the most important findings with tables in chapter 4 results.

4.4 Productivity - measurement method

In order to measure the departments invoice booking process productivity, the first step in this process was to start with evaluating available attributes (variables) in the event log. The attributes are: Case ID, Activity, Source and Timestamp. The research focuses first on the time perspective. The results here could lead to indicate which timeframes are most productive during the hour of the day, day, week, month, quarter and amongst each year.

The hour of the day is broken down in usual work hours which are from 08:00 till 17:00. Anything not between those hours are identified as before and after work hours. A week consist of seven days in which are five working days, Monday till Friday. There are in total 12 months in a year and can indicate which month of the year employees are most productive. For many accounting firms, it is important to file the VAT declaration each quarter before the deadline. All registered businesses in the Netherlands have their financial income and expenses registered in the general ledger and are ready to file the Dutch VAT-form. The VAT declaration months are in January, April, July and October. To compare the results, the following formula will be used:

Total sum activities for day* */ Sum total activity for the year.*

*Day, week and month

The paper then measures the productivity from the department based on employee's performance. During the three years, the department had over 70 different employees participated within the invoice booking process. The research measures the productivity based on the total cases of carried out by each employee divided by the total cases of that year. The outcome is emphasized in percentage, so the results can be compared with other years. Since each year many employees participate in the invoice booking process, the paper has decided to focus on presenting the results till 80% of the total cases. The entire analysis including the results of productivity will be presented in the Appendix. The formula to measure the productivity is:

Sum cases for each employee / Sum total cases for the year

5. Results

The results can be classified into two sections which are the business process analysis and the business productivity.

The business process analysis starts with comparing the log statistics. Here, the paper has used process, case and time perspective three of four process mining perspectives. The perspectives are used as tool to find good descriptions in possible paths, process models, characteristics of a case and timing of events. Once all BPMN models are demonstrated the results were compared amongst each other.

The business productivity analysis focuses on the case and timeframe perspective. Here we discover the productivity within the department amongst employees, but also distinguish the productivity level spread throughout the year.

5.1 Business process analysis – log statistics

Table 10 compares some of the main characteristics of the log statistics for the years 2017, 2018 and 2019. The table indicates the total cases, events, activities and case variants based on the productivity produced by the accounting firm and the clients from the firm.

	2017	2018	2019
CASES	63.188	75.196	87.758
EVENTS	195.946	231.812	269.005
ACTIVITIES	6	5	5
CASE VARIANTS	27	13	18

Table 10

What stands out in the table is the increase of total cases in each year. This means that between 2017 and 2018 an increase of 19% of cases occurred and between 2018 and 2019 16.71%. The events in the table are directly connected to the total events. Each individual case composes a number of events. The average case in 2017 consists of 3.10 events, 2018 consists of 3.083 events and 2019 consists of 3.07 events. What is interesting about the data is that the case variants reduce by more than half between 2017 and 2018. But then increase with 5 case variants between 2018 and 2019.

Table 11 in previous chapter compares the total objects of the population between accounting firm and their clients and summarizes the most interesting aspect of that table. As Table 11 shows, there is a significant difference between 2017 and 2018 to 2019.

CASES	ACCOUNTING OBJECT	CLIENT OBJECTS	TOTAL OBJECTS
AUDIT FILE 2017	72.1 %	27.9 %	100 %
AUDIT FILE 2018	72.9 %	27.1 %	100 %
AUDIT FILE 2019	67.9 %	32.1 %	100 %

Table 11

5.1.1. Business process – BPMN model 2017

Appendix 1 presents a process map of the entire invoice booking process in 2017. The results are classified into standard and non-standard variants. Figure 5 presents the BPMN model with all six activities including the total cases displayed. The model presents the main variant with a thick blue arrow. The standard variant is as follows: Start -> Document received -> Tagged -> Booked -> End.

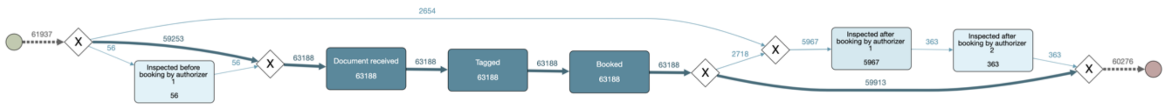


Figure 4

The non-standard variant including all activities is as follows: Start -> Accepted by author 1 before booking -> Document received -> Tagged -> Booked -> Inspected after booking by authorizer 1 -> Inspected after booking by authorizer 2-> End.

Figure 5 presents the median time between activities without a breakdown between the sources. The median time between activities for the standard variant between document received and tagged is 52 minutes and between tagged and booked 4.97 hour.

Figure 5

The non-standard variant notices an apparent disparity in the activities "Inspected before and after booking by authorizer one or two". The model shows that the median time between "Inspected before booking by authorizer 1" is 4.87 months and the activity "Booked" and "Inspected after booking by authorizer 1" took 3.57 months.

The median time for the source clients between activities for the standard variant of document received and tagged is 19 minutes and between tagged and booked 1.57 hour. The non-standard variant includes the activity "Inspected after booking by Authorizer 1" adds another 3.85 months in the timeline.

The median time for the firm's standard variant is 1.48 hours between "Document received" and "Tagged" and 7.78 hours between the activities "Tagged and Booked". The non-standard variant includes three more activities, which increases the time of the business process by 8.41 months.

4.1.2. Business process – BPMN model 2018

Appendix 2 presents the process map for the invoice booking process in 2018. Figure 6 shows the BPMN model including all five activities. Equal to the model of 2017, the main variant is outlined with a thick blue arrow with no difference in activity sequence compared to the year before.

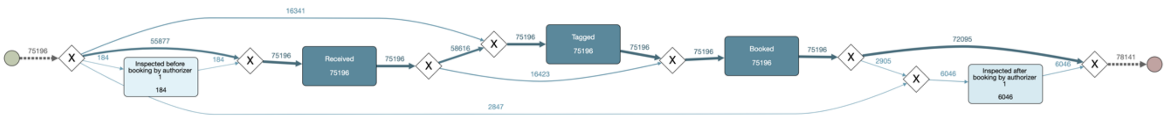


Figure 6

However, there is a difference in activities for the non-standard variant. In 2017, there was one more activity. The sequence for the invoice booking process in 2018 is: Start -> Accepted by author 1 before booking -> Document received -> Tagged -> Booked -> Inspected after booking by authorizer 1 -> End. The difference here is that the "Inspected after booking by authorizer 2 is not included.

Figure 7 presents the median time between activities without a breakdown between the sources. The median time between activities for the standard variant is between document received and tagged is 13.38 hours and between tagged and booked 3.38 days.

Figure 7

The non-standard variant notices an apparent disparity in the activities "Inspected before and after booking by authorizer one". The model shows that the median time between "Inspected before booking by authorizer 1" is 5.72 months and the activity "Booked" and "Inspected after booking by authorizer 1" took 3.62 months.

The median time for the source clients between activities for the standard variant of document received and tagged is 1.98 hours and between tagged and booked 1.42 hour. The non-standard variant includes the activity "Inspected after booking by Authorizer 1" adds another 3.91 months in the timeline.

The median time for the source for the standard variant is 18.13 hours between "Document received" and "Tagged" and 5.17 hours between the activities "Tagged and Booked". The non-standard variant includes two more activities, which increases the business process's time by 9.34 months.

4.1.3. Business process – BPMN model 2019

Appendix 3 presents the process map for the invoice booking process in 2019. Figure 8 shows the BPMN model, including all five activities. Equal to the models 2017 and 2018, the main variant is outlined with a thicker blue arrow with no difference in activity sequence than the other years. The difference is in the non-variant process. The "Accepted before booking by author 1" is excluded.

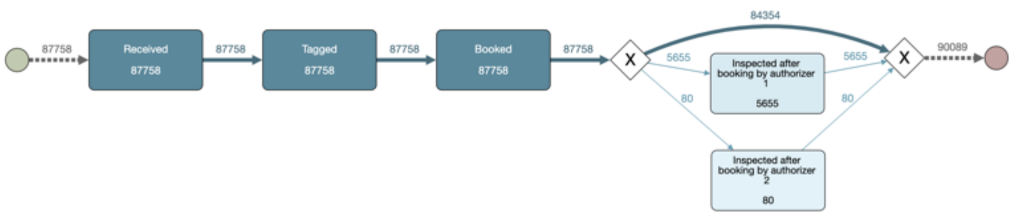


Figure 8

Figure 9 presents the breakdown of time between activities. The median time for activities in the standard variant is between document received and tagged is 1.98 hours, and between tagged and booked, 2.25 days.

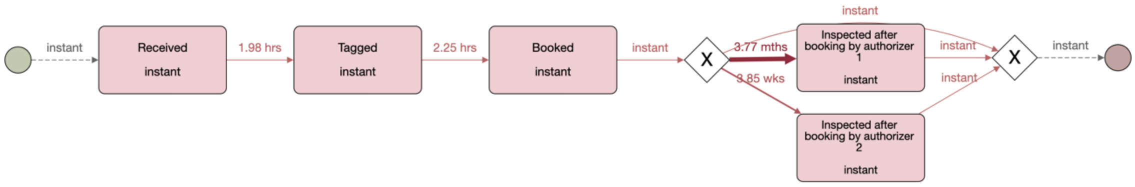


Figure 9

The non-standard variant notices an apparent disparity in the activities "Inspected after booking by authorizer 1 and 2". The model shows that the median time between "Inspected after booking by authorizer 1" is 3.77 months and the activity "Booked" and "Inspected after booking by authorizer 2" took 3.85 weeks.

The median time for the source clients between activities for the standard variant of document received and tagged is 1.33 hours and between tagged and booked 57 minutes. The non-standard variant includes the activity "Inspected after booking by Authorizer 1" adds another 4.39 months.

The median time for the firm in the standard variant is 2.32 hours between "Document received" and "Tagged" and 3.45 hours between the activities "Tagged and Booked". The non-standard variant includes two more activities, which increases the time of the business process by 20.29 weeks

5.1.4 Comparison results between BPMN models

Table 12 compares the difference in total time spent for the main variant between all three years.

MAIN-VARIANT	CLIENT	FIRM'S	ALL CASES
2017	1.89 hours	9.26 hours	5.9 hours
2018	3.4 hours	23.3 hours	16.76 hours
2019	2.28 hours	5.77 hours	4.23 hours

Table 12

Table 13 compares the difference in total spent for the non-variant that includes all activities between all three years.

NON-VARIANT	CLIENT	FIRM'S	ALL CASES
2017	16.3 weeks	36.56 weeks	36.67 weeks
2018	17.01 weeks	40.76 weeks	40.73 weeks
2019	19.07 weeks	20.33 weeks	20.26 weeks

Table 13

The results in this paragraph indicates that there is a difference of productivity based on various factors. The differences are mainly visible in the audit year, between sources but also between main variant and non-variant. The next chapter, therefore, moves on to discuss potential causes on the difference and also explains which impact(s) the results have.

5.2 Productivity results – Power BI

The results for this paragraph describe the productivity level spread throughout the year and amongst employees. Figure 10 represents the productivity level during workdays and weekend. For all three years, the most productive hours are between 8 am and 11 am. There is 35.84% for 2017, 35.1% for 2018 and 34.4% for 2019 invoice booked between these hours. What is remarkable is that after 5 pm on average, 4.12 % of invoices were booked, which means that workers are still productive after office hours. Activities after working hours had slightly increased from 3.99% in 2019 to 5.36% in 2018 to 5.35% in 2019.

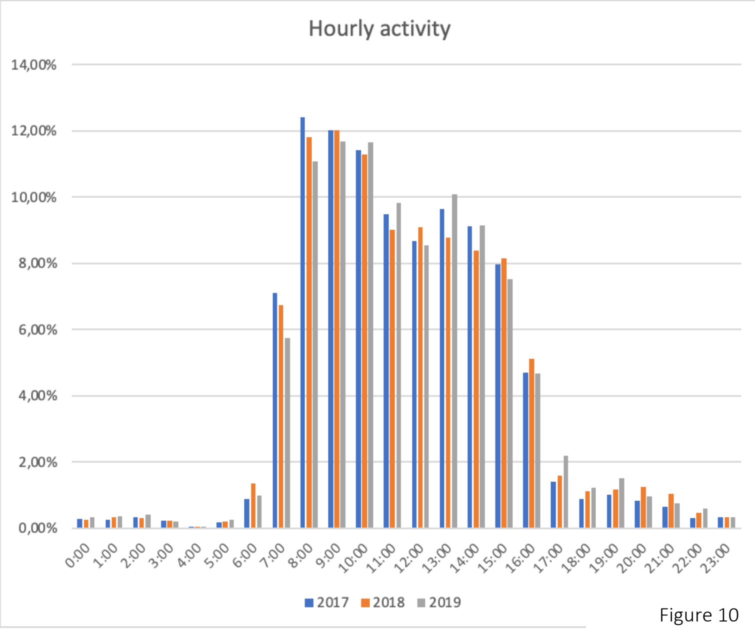
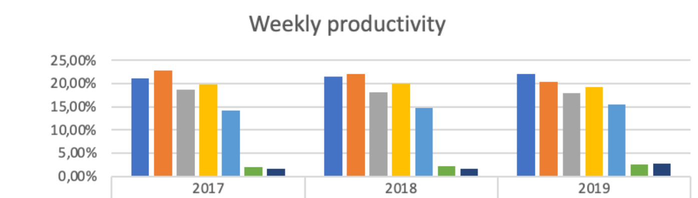


Figure 10

On the contrary, peak hours have been stable. The result has also found that 9.25% for 2017, 9.44% in 2018 and 8.30% in 2019 productivity was measured before starting a working day. The slight decrease of 1% from the number in 2019 compared to 2018 and 2017 is mainly related to fewer people working earlier.

5.2.1 Days of the week analysis

The productivity level in the week varies between days. Figure 11 shows that in 2017 and 2018, the week's most productive day is Tuesday but changed to Monday in 2019. The top three productive work weekday are Monday, Tuesday and Thursday. The least productive day of the workweek is on Friday. The table also shows an increase of 1.5 times working at the weekend from 3.41% in 2017 to 5.21% in 2019.



	2017	2018	2019
Monday	21,15%	21,46%	21,95%
Tuesday	22,86%	22,07%	20,28%
Wednesday	18,72%	18,05%	17,87%
Thursday	19,72%	19,98%	19,30%
Friday	14,14%	14,71%	15,39%
Saturday	1,88%	2,07%	2,58%
Sunday	1,53%	1,65%	2,

Figure 11

5.2.2 Month Analysis

Based on the analysis from figure 12, October is the most productive month, and August is the least productive month following February and May. Results also indicate that the VAT-Months belongs to the top productive months and the least productive months are often the following month after the VAT-months.

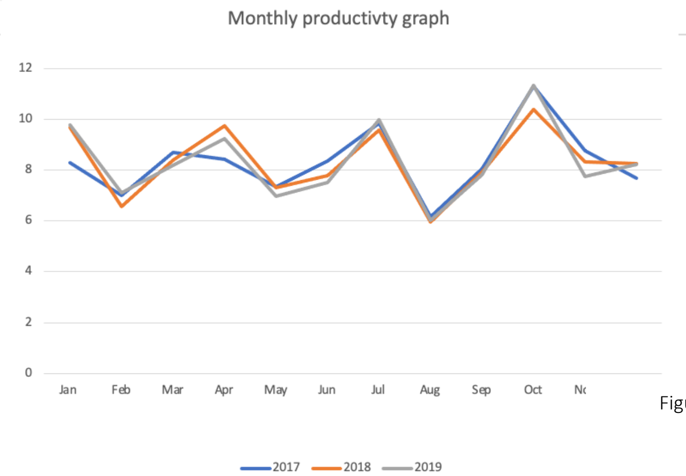


Figure 12

5.2.3 Employee productivity analysis

As earlier described in section 4.1 and table 10, total cases and events starting from 2017 has increased each year subsequently from 19% between 2017 and 2018, 14.4% between 2018 and 2019. Table 14 shows the average cases each participant has on average for each individual year. More cases were carried out by each participant compared to the first year.

	2017	2018	2019
CASES	63.188	75.196	87.758
PARTICIPANTS	43	34	42
AVERAGE CASES EACH PARTICPANTS	1470	2212	2090

Table 13

Figure 13 represents the results for the productivity in 2017. In total 15 employees counted for 80.46% of total cases. Employee 70 (16.18%) has produced more cases compared to the last eight peers combined.

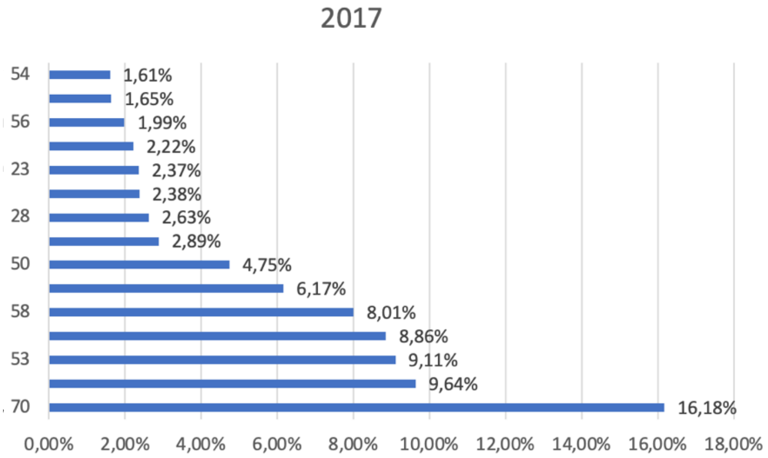


Figure 13

Figure 14 counts in total 9 employees that are good for 81.71% of the total cases. The disparity between the most productive employee and the second most productive is less large compared in 2017. A comparison of the two results reveals that less employees in 2018 were involved but more cases had to be carried out. What stand out in the table is the top three employees (participants 1, 20 and 17) were active in 2017 as well but were not marked as the most productive members in previous year.

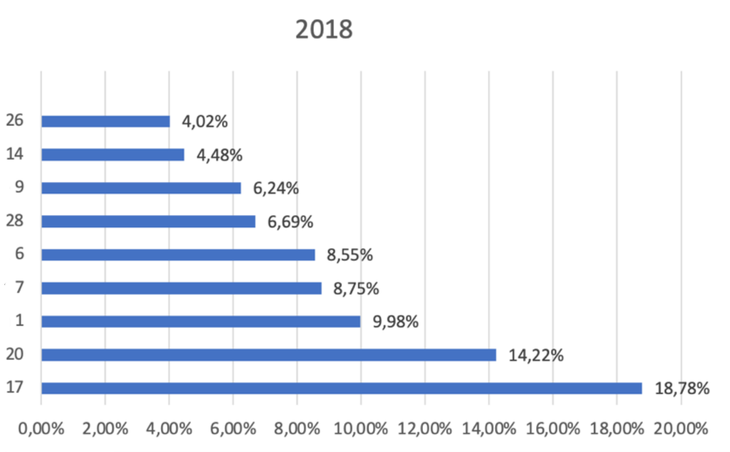


Figure 14

Figure 15 presents the analysis of 2019. What is striking about the figure in this table is that the productivity is more equally divided amongst each other. In total 10 peers represent 80.22% of the total productivity. Employee 7 was the most productive in that year, whereas a year before it was ranked in the fourth place.

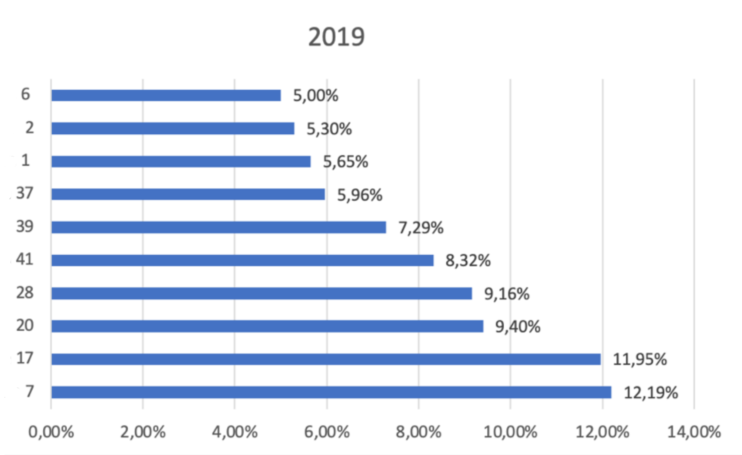


Figure 15

In summary, these results show that the year 2018 was the most productive year based on the average case for each employee. It also shows that, employees 1, 20 and 17 are active in all three years, but employee 17 started as an average performance in 2017 becoming one the most productive employees the years after. Furthermore, the table expresses that employee's productivity is very diverse amongst each other. In the next chapter, therefore, moves on to discuss the conclusion based on the results.

6. Analysis

Based on the previous chapter there are several matters from the analysis that are interesting for the management. This section will focus on relevant outcomes for the firm. It will answer the following management questions: *"What is the current invoice booking business process?", "Are there are bottleneck process mining technique can detect in the invoice booking process?", "Is there a discrepancy in productivity divided among employees?" and "Is there a discrepancy in productivity level between certain time period?"*

The first insight detected from the analysis is the increase of invoice process through the ERP system. This means during the years more invoices are received and booked. There are two main causes that has led to the increase. The first reason is the interchange from paperwork to digitization. The second reason could be an increase of new customer in the database.

The next insight is related to the business process. The current invoice booking process is divided into two variant paths. They are called the main variant path and non-main. The main variant path is a business process that are often followed and could be seen as a general procedure. The non-variant path is a process that deviate from the main variant path. Also seen as a process that is not often followed. Beside the difference in paths, another difference is the time that it took to finish the entire business process. The main variant path takes of serval hours average whilst the non-main variant path takes up to 30 weeks. The time difference comes from one extra process in the non-main variant path, also known as authorizing. Authorizations plays an important role to keep the quality of the activity by detecting potential error or fraud. This extra process focuses on controlling the data entered in the ERP-system. If the information is correct the authorizer will approve, and data will be registered in the financial ledger. The extra check and approval take time, and the entire process is depending on the approval from the authorizer. If the authorizer does not check nor approve, the invoice could circulate for weeks to month in the system, which causes the bottleneck. Question such as: Is the extra check needed, and which risks are there if the firm excludes authorizers are considerable.

The third insight is related to the productivity among employees. The analysis show that there is a substantial difference among peers.

7. Conclusion

This study set out with the aim of assessing possible opportunities from process discovery for the accounting firm. For this reason, this research has raised the following research question: "*What insights can be derived from analyzing the invoice booking process based on event logs from a financial institution's enterprise resource planning system?"* The firm's benefit from analyzing the event log is that it creates in-depth knowledge and insights of their current business process. It provides or confirms based on facts business process-related questions, including insights that they have never been aware of. The business learns new insights that include identifying a switch to digitalization, creating awareness of potential risks or productivity losses, discovering current occurring bottleneck(s), and calculating the productivity level. The following paragraphs describe the new insight more in-depth.

The first finding is a significant increase of case ID's each year. In other words, more invoices are booked via the ERP system (accounting software system). The reason for an increase in cases is led by favouring working digitally. But also, the business has been encouraging their clients to deliver invoices via the system.

The second finding is an increase in the insourcing invoice booking process from clients of the firm. It seems that clients are favouring booking the invoices themselves without support from the accounting firm. One reason why client insources is saving accounting expenses. The other reason is to have control over this particular process. Some companies prefer to have their financial overview updated at all time. However, the accounting firm may not always meet this requirement. The two reasons are potentially important driving factors of potential losing accounting hours, which directly reflects on creating less revenue for the financial institution.

This particular analysis has also investigated the difference between the accounting firm and their clients average time of booking an invoice. The outcome is that clients invested five times on average less time than when the accounting firm carried out the same task. There are two possible underlying factors. The first reason is that the accounting firm works for different clients, and each client has additional requirements and expectations. So, when they work on the administration, it takes time to adapt. The second reason is that when accounting firm receives invoices from the client, it takes time before someone can work on it. In other words, invoices are in the system ready to book but are pending for the best possible date to work on.

The third finding has found the bottleneck within the business process. The analysis has distinguished the invoice booking process into the standard variant and non-standard variant path. The main-variant path is a process map is the process instances that have paths that are aligned with the firm's standard business processes. In contrast, the non-standard variant has paths that deviate from the standard business process. Whilst comparing the processing time between these two paths, the most apparent finding for this analysis is that the authorization causes the bottleneck. Whilst, on average, it takes less than nine hours to complete the process, the non-variant takes on average 32 weeks.

In general, organizations use authorization as an approach to detect errors or fraud. It could lead to the prevention of order or price errors and fake invoices with high sum payments to an unknown bank address. However, based on the analysis waiting for approval from authorizers takes up a lot of time. To reduce the time, one of the solutions could be limiting authorization by focusing solely on invoice with specific criteria's including a high sum of payments and unknown debtors and creditors. Besides focusing on reducing time, *the department can ask the question: is an authorization necessary, or does the extra inspection decreases errors?*

The last finding is related to the productivity within the department. The analysis focused on productivity based on daily working hours, days, weekends, months, quarters, years and finally among peers. The most productive hours are between 8 and 11 o'clock. The graph shows a short productivity dip in the afternoon between 11:00 and 13:00, caused by a lunch break. It also shows that during the years more productivity and work are delivered after working hours. A possible cause is the increase of cases for each employee. In 2017, on average, 1470 invoices were booked by each participant, it has increased by 150% of productivity between 2017 and 2018 and 142% between 2017 and 2019. In other words, more invoices on average were booked during the last years among employees.

The analysis shows that the productivity level in the week varies between days. The top three productive work, weekdays for all three years, were Monday, Tuesday and Thursday, and the least productive day of the workweek is on a Friday. The results also indicate an increase of 1.80% productivity in the weekend. In other words, employees tend to work more often on Saturday and Sunday. Another pattern the paper has found is that October is the most productive month following January, April and July. The least productive month is August since many people are on holidays. The explanation behind their pattern has to do with filing for VAT. These VAT declaration months are in January, April, July and October. The following months after the declaration are less occupied.

The final productivity analysis is a comparison of productivity amongst peers. In 2017, 43 participants worked on 63.188 total cases and an average of 1470 each. Interestingly, it took solely five participants to count for over 80% of confirmed cases. The discrepancy between the most productive participant and the second most productive were 4133 invoices (more than 6.5%). In 2018, nine employees from the financial institution produced in total 81.71% productivity. The discrepancy between the most productive employee and the second most productive is less prominent compared to 2017. A comparison of the two results reveals fewer employees in 2018 were involved, but more cases had to be carried out. Finally, in 2019, ten members from the financial institution represents 80.22% of the total productivity. What is striking about the analysis is that the productivity is more equally divided amongst each other.

To conclude, results show that 2018 was the most productive year based on the average case for each employee. It also showed that employees 1, 20 and 17 were active in all three years. Then, employee 17 started as an average performer in 2017, becoming one the most productive employee in 2018 and 2019. The analysis also shows a high employee turnover during these three years. Employees that were active in 2017 were no longer engaged in other years. Instead, a flow of new hires was retained.

The opportunities process discovery creates for the accounting firm are understanding one particular business process and could also be implemented for various business processes within different departments. This study indicates that process discovery with event logs from ERP systems creates and in-depth knowledge of a business process, including detecting bottleneck and confirming or rejecting management speculations and/or assumptions based on facts.

8 Discussion

This chapter discusses limitations, theoretical and practical implications and topics for future research.

8.2 Theoretical and practical implications

The outcomes of this study are practical and theoretical relevant, especially for an organization in the financial industry. Literature review in this paper has played an essential role in building primary ground for conducting the process discovery analysis. Current literature in process mining tends to focus more on describing techniques rather than explaining the practical implications. The gap between process mining theory and implementing the techniques in practice are, for many businesses, a big step to take. Organizations have none to little experience in managing their information embedded in their operating system(s). Process mining is a technique that exists over more than two decades and still not a method used by organizations. The method requires technical as well as theoretical knowledge that is often unknown or new for managers. Besides that, an event log is a dataset that stores data from the entire population. The datasets are often large and included with many attributes (variables) and require considerable time to invest.

In contrast to being unfamiliar with the process mining technique, this paper's main practical implication is the steps on how to prepare and analyse event logs. The study has shed light on possible opportunities for business productivity analysis, such as calculating the invoice process booking time, comparing productivity among peers, and analysing productivity during the year. The knowledge shared with the management could lead to more similar analyses for other departments within the firm, such as analyzing event logs from the audit department. One of the audit department's functions is to assess the quality of internal controls where the department inspects documents and reports or trace transactions through the client's financial report system. The use of event logs could be beneficial since it can detect outliers that are potential errors or fraud.

There are three practical pieces of advice the paper would like to share with the management. The first piece of advice is to claim more authority and control from data collected by the ERP system provider. The management will have event logs available and have the control including and excluding attributes. The second advice is to speak to employees who are experiencing a high workload. Oral explanation of employees' feelings and experiences can either confirm or reject the analysis in this paper. An example here is that employee 17 was a top performer in 2018 and 2019. Do they particularly experience a high workload, and what can the management do to support them? The last practical advice is to create a forecast model that predicts the productivity for the upcoming years. The results and tables show that in certain months and years, more productivity is measured. The management can use these data to prepare the planning for the future and make management decisions such as holiday planning, prediction to recruit extra personnel and plan training for less busy seasons.

8.3 Future research

This research has several questions that remained unanswered at present. One of the questions is related to 2020's event log. In March 2020, companies worldwide, including the Netherlands, abruptly shuttered their offices and instructed employees to work from home indefinitely due to the pandemic. The event log of 2020 could become a great source of data since it can analyze the impact of productivity whilst working remotely.

The second question that remained unanswered is related to control of the data collection. The limitation in this research was that the supplier of the accounting software system had the authority of the data. The accounting firm had no control of including or excluding attributes such as the end-timestamp and source of the person who entered the invoice in the accounting system. If each data entry and activity within the ERP were logged, the paper could conduct a more precise analysis, including detecting outliers between activities. An example of the end-timestamp is the moment when the invoice is delivered to the ERP system. Most of the information has been already entered by the system. An administrator has to check if the data is correct or re-enter the information into the system. However, in some cases, the ERP system does not recognize (some) information and can't automate and backfill the data into the system. In this case, it takes more time in correcting incorrect data and re-enter new or additional information. Logging these data could be beneficial in calculating the real-time of each activity. It also could support the management to understand the time difference in tasks between when the system fails to recognize the data from the invoice.

8.4 Limitations

Specific to every research, this study comes across some limitation which needs to be taken into account when making interpretations. First, a standard limitation is that the theoretical grounding on which this research is based belongs to the structured literature review realized by the author of this paper. It could be possible that some additional relevant literature has been overlooked.

Perhaps the most impactful limitation of this study is the control and ownership of the event logs. The accounting software system supplier decides which attributes from the data are allowed to be shared with the company. It also determines which attributes and values are excluded and included. The limitation here is that a third party is able to alter or delete existing data. In addition to that, the research had no authority to obtain additional data that might be relevant for this study. An example of a specific limitation for this particular study is that the end timestamp date was missing. Including the end-stamp date could provide the analysis with a more precise time between two activities.

The third limitation is that this study has solely focused on the invoice booking process. The productivity analysis is exclusively based on how many invoices were booked, but other important and relevant daily business tasks were excluded. In other words, the representation of productivity is based on one business process activity and does not represent the department entire productivity.

The fourth limitation is related to calculating the Return on Investment (ROI) for this research. The ROI calculates the net income of an investment. In this essence, it isn't easy to calculate what the return could be since we cannot calculate the impact of this analysis.

The last limitation is to find errors in the techniques the paper had used. The techniques used in conducting the analysis are based on literature review and logical thinking, personal knowledge and experience. Furthermore, the research was dependent on various support systems such as Apromore, PowerBI and Excel Query. Run time and lag with a large dataset had occurred often. The run time of Excel Query and PowerBI had taken up several hours for each analysis. This chapter discusses limitations, theoretical and practical implications and topics for future research.

9. References

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *Proceedings Of The 1998 ACM SIGMOD International Conference On Management Of Data - SIGMOD '98*. https://doi.org/10.1145/276304.276314

Aguilar-Savén, R. (2004). Business process modelling: Review and framework. *International Journal Of Production Economics*, *90*(2), 129-149. doi: 10.1016/s0925-5273(03)00102-6

Burattin, A. (2015). *Process Mining Techniques in Business Environments* (pp. 1,33). Springer International Publishing.

Bussler, C., Jablonski, S., & Schuster, H. (1996). A new generation of workflow-management-systems: beyond Taylorism with MOBILE. *ACM SIGOIS Bulletin*, *17*(1), 17-20. https://doi.org/10.1145/236410.236429

Chiu, T., & Jans, M. (2019). Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness. *Accounting Horizons*, *33*(3), 141-156. https://doi.org/10.2308/acch-52458

Cook, J., & Wolf, A. (1995). Automating process discovery through event-data analysis. *Proceedings Of The 17Th International Conference On Software Engineering - ICSE '95*, *4*, 73-82. https://doi.org/10.1145/225014.225021

De Weerdt, J., De Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, *37*(7), 654-676. doi: 10.1016/j.is.2012.02.004

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. (2018) *Fundamentals of business process management* (2nd ed., pp. 421-426, 428-435). The Springer.

Dustdar, S., Hoffmann, T., & van der Aalst, W. (2005). Mining of ad-hoc business processes with TeamLog. *Data & Knowledge Engineering*, *55*(2), 129-158. https://doi.org/10.1016/j.datak.2005.02.002

Eisenhardt, K. (1989). Building Theories from Case Study Research. *Academy Of Management Review*, *14*(4), 532-550. https://doi.org/10.5465/amr.1989.4308385

Champy, J., & Hammer, M. (1993). *Reengineering the corporation* (p. 11). HarperCollins e-Books.

Jans, M., Alles, M., & Vasarhelyi, M. (2010). Process Mining of Event Logs in Auditing: Opportunities and f Challenges. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2488737

Jans, M., van der Werf, J., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems With Applications*, *38*(10), 13351-13359. https://doi.org/10.1016/j.eswa.2011.04.159

Jans, M., Alles, M., & Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal Of Accounting Information Systems*, *14*(1), 1-20. https://doi.org/10.1016/j.accinf.2012.06.015

Key features - Apromore. (2021). Retrieved 17 December 2020, from https://apromore.org/platform/key-features/

Leemans, S., Fahland, D., & van der Aalst, W. (2016). Scalable process discovery and conformance checking. *Software & Systems Modeling*, *17*(2), 599-631. doi: 10.1007/s10270-016-0545-x

Leemans, S., Fahland, D., & van der Aalst, W. (2016). Scalable process discovery and conformance checking. *Software & Systems Modeling*, *17*(2), 599-631. doi: 10.1007/s10270-016-0545-x

Mabert, V., Soni, A., & Venkataramanan, M. (2003). Enterprise resource planning: Managing the implementation process. *European Journal Of Operational Research*, *146*(2), 302-314. doi: 10.1016/s0377-2217(02)00551-9

Mans, R. S., Van der Aalst, W. M. P., & Vanwersch, R. J. B. (2015). *Process mining in healthcare: evaluating and exploiting operational healthcare processes (p24)*. Springer.

Motahari-Nezhad, H., Saint-Paul, R., Casati, F., & Benatallah, B. (2010). Event correlation for process discovery from web service interaction logs. *The VLDB Journal*, *20*(3), 417-444. https://doi.org/10.1007/s00778-010-0203-9

Ould, M. (2007). *Business process management*. British Computer Society.

Pajić, A., & Bečejski-Vujaklija, D. (2016). Metamodel of the Artifact-Centric Approach to Event Log Extraction from ERP Systems. *International Journal Of Decision Support System Technology*, *8*(2), 18-28. https://doi.org/10.4018/ijdsst.2016040102

Petri, C. (1966). Communication with Automata. *Reconnaissance - Intelligence Data Handling Branch*, *1*. Retrieved from https://edoc.sub.uni hamburg.de/informatik/volltexte/2010/155/pdf/diss_petri_engl.pdf

Reinkemeyer, L. (2020). *Process mining in action* (1st ed., pp. 3-6). Springer, Cham.

Robert Jacobs, F., & Weston, F. (2006). Enterprise resource planning (ERP)-A brief history. *Journal Of Operations Management*, *25*(2), 357-363. doi: 10.1016/j.jom.2006.11.005

Soibelman, L., & Kim, H. (2002). Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases. *Journal Of Computing In Civil Engineering*, *16*(1), 39-48. doi: 10.1061/(asce)0887-3801(2002)16:1(39)

T. Davenport, Putting the enterprise in the enterprise system, Harvard Business Review 76 (4) (1998) 121–131.

Turner, C., Tiwari, A., Olaiya, R., & Xu, Y. (2012). Process mining: from theory to practice. *Business Process Management Journal*, *18*(3), 493-512. doi: 10.1108/14637151211232669

Umble, E., Haft, R., & Umble, M. (2003). Enterprise resource planning: Implementation procedures and critical success factors. *European Journal Of Operational Research*, *146*(2), 241-257. https://doi.org/10.1016/s0377-2217(02)00547-7

van der Aalst, W., & Basten, T. (2002). Inheritance of workflows: an approach to tackling problems related to change. *Theoretical Computer Science*, *270*(1-2), 125-203. https://doi.org/10.1016/s0304-3975(00)00321-2

van der Aalst, W., ter Hofstede, A., & Weske, M. (2003). Business Process Management: A Survey. *Lecture Notes In Computer Science*, 1-12. https://doi.org/10.1007/3-540-44895-0_1

van der Aalst, W., & Weijters, A. (2004). Process mining: a research agenda. *Computers In Industry*, *53*(3), 231-244. https://doi.org/10.1016/j.compind.2003.10.001

van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: discovering process models from event logs. *IEEE Transactions On Knowledge And Data Engineering*, *16*(9), 1128-1142. https://doi.org/10.1109/tkde.2004.47

van der Aalst, W. (2012). Process Mining. *ACM Transactions On Management Information Systems*, *3*(2), 1-17. https://doi.org/10.1145/2229156.2229157 van der Aalst, W. (2016). *Process mining* (pp. 8, 25, 29-50, 55-58). Springer.

Webster, J & Watson, R., (2002). Analysing the past to prepare for the future: Writing a literature review a roadmap for release 2.0. *Journal Of Decision Systems*, *26*(2), 129-147. https://doi.org/10.1080/12460125.2020.1798591

Wen, L., van der Aalst, W., Wang, J., & Sun, J. (2007). Mining process models with non-free-choice constructs. *Data Mining And Knowledge Discovery*, *15*(2), 145-180. https://doi.org/10.1007/s10618-007-0065-y

Wen, L., Wang, J., van der Aalst, W., Huang, B., & Sun, J. (2008). A novel approach for process mining based on event types. *Journal Of Intelligent Information Systems*, *32*(2), 163-190. https://doi.org/10.1007/s10844-007-0052-1

Weske, M. (2007). *Business process management* (pp. 4, 73-96). Berlin: Springer.

Appendix 1,2 and 3