UNIVERSITY OF TARTU

Institute of Computer Science

Innovation and Technology Management Curriculum

Maria Häling

Improving operational processes - A process mining case study in e-commerce

Master's Thesis (20 EAP)

Supervisor:
Fredrik Milani, PhD

Tartu 2021

Improving operational processes - A process mining case study in e-commerce

Abstract:

Increasing demand in e-commerce has developed a highly competitive environment for companies, where process efficiency and customer satisfaction are keys to success. To identify opportunities for increased performance, process mining can be used as a data-driven analysis method. Process mining is a tool of business process management that is specifically used to find improvement opportunities in processes based on ta logs. Current research in e-commerce has studied mining in customer weblogs context, but using process mining on business operational processes has yet to be done. Hence, this study explores this aspect from process performance and customer satisfaction angles, providing additional knowledge to existing research. The results showed that process mining could effectively identify performance and customer experience related problems in operational e-commerce processes. Furthermore, combining domain experts' knowledge, redesign heuristics, and analysis results enabled to propose relevant and insightful change opportunities.

Keywords:
Process mining, Business Process Redesign, E-commerce
CERCS: P170 – Computer science, numerical analysis, systems, control

Operatiivsete protsesside parendamine – Juhtumiuuring protsessikaevest e-kaubanduses

Lühikokkuvõte:

Suurenenud nõudlus e-kaubanduses on firmadele tekitanud kõrge konkurentsiga keskkonna, kus protsesside efektiivsus ja klientide rahulolu on edu võtmed. Selleks, et leida protsessi suutlikkuse suurendamise võimalusi, saab kasutada protsessikaevet andmepõhise analüüsi meetodina. Protsessikaeve on äri protsesside juhtimise vahend, mida kasutatakse spetsiaalselt protsesside parendusvõimaluste leidmiseks kasutades andmete logisid. Senistes uurimistöödes on e-kaubanduses kasutatud protsessikaevet vaid kliendipoolsete veebilogide kontekstis, kuid äri operatiivsete protsesside uuringut protsessikaevaga veel läbi viidud ei ole. Seetõttu täiendab antud uurimustöö olemasolevaid teadmisi, kasutades protsessikaevet just protsessi jõudluse ja klientide rahulolu aspektides. Töö tulemused kinnitavad, et protsessikaevega on võimalik leida protsessi jõudluse ja ka kliendikogemusega seotud probleeme e-kaubandusega seotud firma operatiivsetest protsessidest. Kombineerides analüüsi tulemused protsessi ekspertide teadmistega ja protsessi ümber kujundamise heuristikaga, on võimalik mõelda asjakohaseid ja perspektiivikaid muudatuste võimalusi.

Võtmesõnad:
Protsessikaevae, Äri protsesside ümber kujundamine, E-Kaubandus
CERCS: P170 – Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of Contents

1. Introduction

In today's competitive economy, companies need to adapt to market demand and rivalry by increasing efficiency via improvements to business processes [1]. A company can be seen as a collection of processes that work towards the same goal [2]. One way to increase output, reduce costs and time spent per process is to look inwards and streamline the organisation's internal processes. The set of processes is controlled by business process management (BPM). BPM can be defined as the combination of knowledge from IT systems and the managerial aspect to find improvement opportunities in operational business processes [3]. Using BPM has become increasingly popular because of its potential to save costs by boosting productivity [3].

As companies are increasingly utilising a plethora of information systems in their day-to-day operations, the availability of data has never been so reachable [4]. The data reflects all the processes at the core of an organisation by keeping an event log of said operations. These event logs are merely timestamped records of events collected to one data log and are used to perform process mining analysis [5]. Process mining is used as a tool of BPM to increase efficiency and make educated decisions about a company's processes. The process mining manifesto defines it as: "*The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems*" [3]. Process mining produces evidence and real-life based analysis for organisations to make informed decisions regarding BPM to lower costs and increase performance.

From all possible business processes, the study focuses on online sales processes and how to improve them. Online sales processes are data-heavy from consumer reactions to company systems activity logs [6], making these great candidates for data-based analysis via process mining. The problem at hand is that companies do not use available data as efficiently as they could [7], and improvement possibilities are missed out on. Increasing volumes in e-commerce are at an ever-high upwards trend, particularly due to the COVID-19 outbreak, which makes it an important area to research [8].

Currently, there are no relevant studies on the problem because past studies have mainly focused on theorizing on BPM and process mining approaches [3, 4, 9, 10]. Some case studies have researched the industrial section, for example, applying process mining to a construction company [11] and even in the medical field, like this study about emergency room procedures [12]. However, these case studies are irrelevant to the study at hand since the operational processes in these sectors are wildly different from e-commerce. Some case studies have been conducted in e-commerce; however, they focus on the consumer-side of data [6, 7, 13], not on company internal processes, making them unrelatable for the thesis at hand. That being said, these studies can be used to generalise opinions and strategies for an industry or specific process type, but every company is different. To gain insightful results and make personalised suggestions, a separate study is needed to make knowledgeable proposals to the firm at hand.

The current research is about the online sales process of an e-commerce company, which is a Baltics based firm operating since the beginning of the nineties. Currently, the e-store offers all consumer products from electronics to sports equipment, cosmetics, toys and lots more. The website that is up today has been in place since 2014 without any extensive improvements since then. The company uses four central information systems to log orders, invoices, money movement and warehouse movements concerning e-commerce.

Online sales are affected by consumer side processes (making orders) and company operational processes of fulfilling the orders. Since customer-facing data logs are not currently available, the question is if something can be done about the operational processes to improve the company's efficiency and even customer experience through it. In light of that, the thesis seeks to answer the following research question: **What improvement opportunities can be identified via process mining in the operational processes of an e-commerce company?**

The study attempts to uncover how online sales fulfilment is influenced by the organisation's resources (employees), suppliers, delivery partners and other attributes and what opportunities can be found via process mining. The research performs evidence-based discovery and analysis to find data-based answers to the research question. As a result of the analysis, the study makes improvement suggestions to the company's current process.

The research aims to find bottlenecks and inefficiencies for the company's internal process for handling online sales. To find answers to the research question, this thesis uses a case-study approach by making an in-depth analysis of company data. This study applies process mining methods to the firm's internal process logs to find business process improvement opportunities. The findings of the thesis provide evidence-based understandings of the process as currently, the company has a confidence-based perception. In addition, suggestions and redesign ideas to improve the current state will be made to maximise the efficiency of the operational sales process.

The thesis contributes most to the firm's employees, who can improve their daily work tasks to be more efficient, less time-consuming and less stressful. The analysis highlights the shortcomings of the process as well as the reasons for them, so the firm has a list of problems they can delve head-first into solving. The suggestions highlight possibilities for automation, which in turn means minimised workload for the sales personnel. Moreover, the thesis will benefit customers, who can enjoy faster service, more precise timeliness of orders and overall better customer experience.

The remainder of the thesis is organised as follows. Sections two and three will briefly cover the background and related works for the topic at hand. Section four describes the methods used for this study, following which section five covering the analysis results. Section six reflects on the results and includes suggestions for the future with commentary and validity from the company's perspective. Finally, section seven concludes the paper.

2. Background

To give a common understanding of the topics involved in this research, the background chapter explains the main concepts used in the thesis. This part has two fundamental topics – business process management and process mining. Firstly, the thesis explains what the concept of BPM means and how it is used and implemented in organisations. It provides a foundation to move on to process mining, the second topic. The background covers process mining definitions, their relation to BPM and what types of mining there are. In addition to an overview of the theory, specific methods used in the research, later on, are identified.

2.1 Business Process Management

The overarching topic of the research is business process management. As already mentioned above, each company comprises of processes and how they can be improved through business process management.

According to Dumas [5], one example of business processes occurs when companies provide a service or product to a customer. In order to deliver the item, the company goes through many internal processes (like ordering the product). The process itself comprises of different events and activities, where events happen automatically (no time duration), and activities take some time to complete (for example, inspection of arrived product). Usually, there are also decision points in more complex processes, which are points where the course of the process changes based on the decision made. These processes are performed by actors, which could be people or even software acting for an organisation. The process itself involves the movement of physical or intangible items, which can be anything from equipment to electronic documents.

Acknowledgement of processes and process-driven strategic management has its roots in the industrial revolution when technological advances via innovation were first used to increase productivity and efficiency (for example, Henry Ford) [3, 5]. Manufacturers started noticing how process-oriented factories could cut costs and produce larger volumes; hence organisations started to pay more attention to process-first strategies [10]. Further rise of technological advances meant that companies started to innovate their processes by combining new information systems and automatics into the organisation. The emergence of these changes started to radically transform traditional business processes, leading to drastic new ways of conducting business [3]. All of this change brought forward the idea of managing these new, more complex processes to optimise outcomes best, hence business process management [5].

The contemporary definition of BPM would be that it is the science of how tasks are carried out within a firm to leverage possible improvements and establish consistent results [5]. BPM is known for its ability to foremost help reduce costs and increase process efficiency in order to grow an organisation's competitive advantage [3]. To determine which business processes need redesign or improvements, the organisation needs to implement the BPM lifecycle phases

[5]. Different authors have their own interpretations of how this lifecycle is supposed to be, but this research adheres to Dumas' [5] perception of the BPM lifecycle (Figure 1).

The first step is to identify which process needs changes by creating a process architecture, meaning a figure that shows how the company's processes interact with each other. By visualising the processes, the firm is able to specify its scope to what they currently wish to change via BPM. This process architecture help manage the complexity of intertwined information systems and resources and gain insight into problem areas [10]. After relevant processes are found, the cycle moves to the process discovery phase. In this step, the as-is model is made of the processes in question to document how these processes work currently.

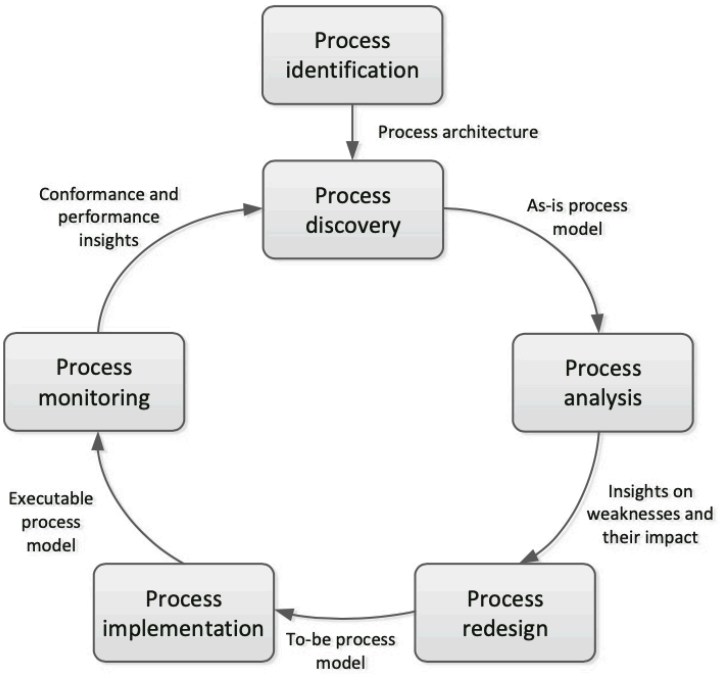


Figure 1. Business Process Management Lifecycle [5]

The third step is to analyse the as-is model by documenting all errors and problems that occur. To achieve the best results, these issues should be quantified where possible by the organisation's performance metrics. Before moving to the redesign, these problems should also be prioritised by impact and effort measures. In the fourth phase, using the previously made analysis, some redesign options are thought of to meet the performance measures. By using the same analysing technique, these ideas are also lined up to find the combination to yield maximised results with minimal effort. At the end of the stage, the to-be model is made.

The next step includes making necessary changes inside the company to implement these new redesign ideas. This step can include developing new information systems, making organisational structural changes or even automating tasks. After the changes are done, the new process model is executed and monitored. Monitoring is essential to analyse if the changes deliver promised efficiency and achieve the previously made goals for the redesign. If new problems arise, these will be dealt with again from the start of the cycle with process discovery.

This research will focus mainly on the discovery, analysis and redesign steps of the BPM lifecycle. The discovery phase will mainly consist of data exploration to understand how the as-is process looks like. After the initial process is identified, process mining will be used in the analysis stage. By running the as-is process through the process mining algorithm, the thesis can identify bottlenecks and shortcomings of the current process. Finding these areas in need of improvement helps the research focus on the redesign aspect of BPM. Further research will be carried out to find topic-specific improvement ideas and automation possibilities as part of redesigning the as-is to the to-be model. Since the case study aims to find evidence-based results, the primary analysis method will be the above-mentioned process mining technique.

Additionally, the thesis will apply redesign principles and heuristics to provide relevant redesigns for the company. According to the BMP fundamentals book [5], when it comes to redesigning, there are two types of approaches – exploitative and explorative redesigns. The exploitative redesign is also called transactional since it does not question the current process structure, instead, emphasis is on resolving problems incrementally. On the other hand, the exploratory redesign also called transformational change, where the fundamentals of the existing process structure are put to the test to realize breakthrough innovation.

The transactional redesign focuses on the heuristics of the process and identifies change based on the four aspects of cost, time, quality, and flexibility. There are nine heuristics, which help find improvement opportunities and propose how to redesign the process [5]. To give an overview of all heuristics and their aims, table 1 was conducted to summarise the approaches.

Table 1. Summary of redesign heuristics, inspired by [5]

H1. Task elimination	- Eliminate non-value-adding steps. - Reduce manual control activities.
H2. Task composition and Decomposition	- Eliminate transportation waste by combining tasks. - Where tasks could be assigned to dedicated resources, they should be split.
H3. Triage	- Specialise an activity, so a general task is split into alternative activities. Like, split insurances to simple and complex cases. - Alternatively, do the opposite and generalise a task that should not be split into many steps.
H4. Re-sequencing	- Activities should be re-orders to minimise over-processing and costs. For example, more expensive tasks should be done last to avoid unnecessary activities and costs.
H5. Parallelism enhancement	- To decrease cycle time, tasks should be done parallelly where possible.

H6. Process specialisation and standardisation	- Specialisation focuses on spitting a process into multiple dedicated ones based on, for example, geography or customer class. Resources are split between the new processes. - Standardisation does the opposite, so that unnecessary specialisation is integrated into one process.
H7. Resource optimisation	- Share workload between idle and overloaded employees. - Specialise people to do what they do best, but avoid not being flexible when needed. - Avoid setups by batching cases into the same task.
H8. Communication optimisation	- Reduce the number of interactions by gathering information. - Optimise the timing of communications.
H9. Automation	- Share data between resources and increase its' availability to avoid duplication - Promote self-service via online forms - Use tracking to keep information about resources - Automate information processing by business rules

The transformational redesign has a total of 5 principles for redesigning the original process structure. These principles are also from BPM fundamentals [5] and are summarised in table 2 for a complete overview.

Table 2. Transformational redesign principles summary, inspired by [5]

P1. Capture information once and at the source	- Data should be captured once and then shared across a system instead of sending it around. - Improve self-service so that the customer performs some tasks themselves.
P2. Integrate information processing into the real work	- Information should be processed in the same place where it is captured in the first place.
P3. Those who use the output should drive the process	- The actor who has the motivation to perform the task should be able to do it. - Decrease centralisation to a certain level to have less bureaucracy and allow people to follow through on tasks they need the output of.
P4. Put decision points where the work is performed	- Empower the employees of the process. - Remove back-and-forth transactions between departments and managers. - Ensure control points to keep the power in check.

P5. Treat geographically split resources as one pool	- Use employees with similar skillsets for the same tasks regardless of their geographical location. - It creates a larger resource pool so that resource utilisation is more even.

2.2 Process Mining

The main idea of BPM is to achieve business process improvements. As mentioned already, these improvements can include increased efficiency, minimised costs, streamlining processes and automation. One tool to find these opportunities is process mining based on event logs of said processes. It is a statistical technique used to assess past behaviour to outline a processes current state [2]. Process mining aims to obtain information about processes from the logs created during the execution of these processes [14].

Event logs come from all possible information systems or resource planning systems an organisation uses. If a user performs an action in the system, it is logged by the system (timestamp, actor) [6]. These collections of data are then collected, and an event log is born. Each of the events in the log can include several attributes like the time stamp, who is performing an activity, what activity is being done, which case is affected or whichever other data the system can capture. Here is where process mining comes to play with algorithms to make sense of these event logs [6]. However, for the majority of process mining techniques, there are three must-have attributes the data log needs to contain: (1) case ID to know in which case the task happened, (2) event or activity, to know which task was performed, and (3) a timestamp to know when the event happened [5].

According to van der Aalst [9], process mining has three primary perspectives: process perspective, organisational perspective and case perspective. The first focuses on how the process is done by looking at the pathways created by the event log and focuses primarily on the process characteristics. Organisational perspective is more about the who, so which resources are at play in the process and how they affect the outcome of the process. The last is about specific cases inside the process. For example, if we talk about e-commerce processes, the first view could be about how a sale gets done on the website. The organisational perspective would be who makes the sale happen / what influences the sale from a resources point of view. From the case view, it could be looking at some instances where people get stuck in the shopping basket, why and how it happens.

In addition to these three perspectives, Dumas [5] refers to four process mining techniques: Automated process discovery, conformance checking, performance mining and variants analysis. The current research uses the first of these mining techniques, so this will be explained further. Automated process discovery uses event logs as input to construct a business process

model based on the events found in the data. It is done by algorithms that detect the behaviour inside the logs and then map it as closely as possible.

One tool of automated process discovery is dependency graphs which aid in visualising event logs [5]. Dependency graphs (Figure 2) consist of two elements – nodes and arcs. Nodes represent the tasks of the process, and arcs signify relationships between different nodes. In these graphs, arcs usually have numbers to signify how many times one task is directly followed by another. Hence, as an example, task H follows task G directly on average in 80 cases out of the total 100. The other 20 cases come from task E.

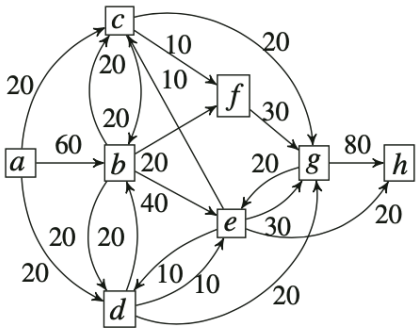


Figure 2. Simple Dependency Graph with Nodes and Arcs [5]

Since process mining builds an initial idea of the as-is process based on event logs, it is crucial to use as detailed data as possible [14]. That said, the quality and scope of the analysis depend on what kind of data logs are available. The dependency graph represents all pathways of the current process, which helps to see where the process slows down and loses efficiency. The completion of the dependency graph sets the base for further process mining analysis. Besides showing dependencies, process mining analysis also helps conduct conformance checking and performance analysis [5]. Conformance checking allows a company to see if its process works as promised – does it meet the set deadline/requirement. Performance analysis, on the other hand, paints a picture of how company resources (time, cost) relate to the process outcomes (quality) [9].

One of the main focuses of this thesis will be performance enhancement by process mining. As mentioned, performance analysis can compile of time aspects, resources used (performers and materials) and quality of work (internal and external) [15]. Milani and Maggi [15] have composed a framework where techniques from different studies have been combined to reach an overarching performance analysis structure. The framework states that the type of performance analysis relies heavily on what type of data is available in the logs. For example, for descriptive performance analysis, the data must have at least the case id, timestamp and activity. This data allows algorithms to compare timestamps and determine the duration of the process, activities and even waiting time if activity end-time is also available.

To conduct resource performance analysis, the framework [15] specifies a need for additional data. Another layer of performance analysis is achieved by adding to the data the resources used in each activity or the person/system who performed it. In addition to gaining insight into how each resource performs, it is also possible to find out low-performers and bottlenecks by comparing logs.

Lastly, the framework covers quality performance analysis. For this, the data needs information about what is the needed outcome and if the case has the desired outcome. Measuring this allows to do conformance analysis based on previously set goals for the process, giving the company an idea of how high-quality their process is.

To make this type of analysis, there are various tools out there that can read event logs and provide process mining outputs. For example, there are tools like Disco, Icris, Rialto and ProM [2]. This research uses a tool called Apromore, a business process analytics tool that provides all process mining functionalities [16]. Apromore has a free community version that allows to upload event logs, then generates the process maps and BPMN models. There are possibilities to do some data filtering as well as flow comparisons and animations. An important feature available is conformance checking, which allows the user to see if processes comply with needed standards. The usage of this tool allows to find bottlenecks in the business process based on the event logs provided.

3. Related Work

This thesis is a case study based on an e-commerce company, so the related work explores what has been studied in this area regarding process mining. Process mining researches in e-commerce have focused primarily on the customer side of data logs, so this section will also review studies related to a variety of organisations' operational process improvements via process mining. Based on available knowledge, research has yet to cover the topic of using process mining for improved customer experience based on operational processes. Thus, the case study wishes to add to the existing literature by increasing customer-centricity of processes from the back-end of a sales process.

The most valuable aspect of e-commerce is data, especially customer data [7]. Current research has uncovered methods for analysing customer behaviour via weblogs [6, 13, 16, 17]. They talk about using customers' past behaviour to predict future behaviour and find improvement opportunities. Poggi [6] researched methods about how to get weblogs in the first place and then used them to predict customer churning points, helping the organisation target these touchpoints to decrease customers dropping from the website. On the other hand, Jiang [13] and Brunk [18] used weblogs for clustering to find the perfect customers to target with offers. Terragni [17] explores a similar topic by producing personalised suggestions based on the customer journey map created from web data logs.

These mentioned authors also point out the lack of process mining research e-commerce, which they believe is mainly due to data sensitivity issues. Thiede [19] provides an overarching literature review of 144 papers about process mining usage. From there, it comes to light that from 2003 to 2016, there were only two researches in the wholesale/retail area. It does not specify online or physical retail. However, e-commerce, in general, is a relatively new and fast-growing sector where proper research has not yet been carried out, especially in BPM and process mining.

Since the thesis aims to investigate the company's backend processes of e-commerce sales, the following paragraph will cover similar researches using internal event logs for process improvements. So far, no studies have investigated these processes based on an online retail company, so other sectors were explored.

One main argument for using process mining analysis is finding resource optimisation opportunities. [20] is a case study where process mining is used for asset management, and similarly [21] explores mining for performance analysis and redesign possibilities. This aim is especially relevant in the manufacturing and industrial sectors, where cost, efficiency and quality assurance are critical [11, 21]. The overarching theme in these studies is how to achieve efficiency by cutting costs. This thesis focuses on performance-related aspects and the customer-centric view of process improvements, but the possibilities for cost reduction will not be overlooked. Van der Aalst [11] also brings out an imperative aspect of teaming up with people who work with the process closely to have successful analysis results. Thus, the study

also focuses on combining process mining with domain experts knowledge and observational data about the process in question.

Another sector with several pieces of research is finance/insurance, which investigate topics from fraud [23] and auditing [24] to also internal efficiency [25]. Weerdt [25] writes how process mining enables to simulate real-life issues so realistically that it is the perfect tool for process improvement. It helps to identify unnecessary loops and data duplications in the system. He discusses how another benefit is increased awareness of company processes among employees, creating a more efficient work atmosphere. These essential aspects (internal efficiency, awareness) will be covered in this thesis by consulting with the company after the analysis results are attained.

To conclude, existing research focuses mainly on either customer data from weblogs or company internal processes in fields not related to e-commerce, thus additional research is needed. In addition to performance analysis, this thesis plans to use operational process analysis for customer experience improvement to bring about a new type of study where a customer-focused approach starts from within the company.

4. Methods

This chapter of the thesis discusses the choices for methodology and design of the research. The setting of the case at hand is described, and the explanation of the theoretical framework is given. As the research is primarily quantitative, the techniques for data preparation and analysis are discussed as well. Lastly, the paragraph covers the qualitative methods used to evaluate and validate the study.

4.1. Case Study Methodology

This research aims to examine how a company can employ data-driven methods for identifying improvement opportunities in their operational processes. Due to this, the thesis follows a case study methodology, which is suitable for exploring real-life issues by implementing theory to practice [26]. A mixed-methods approach is used to gain deeper insight by combining quantitative data analysis with a qualitative workshop. In contrast to other methodologies, case studies provide deeper insight into problems by putting theoretical knowledge to the test. By implementing this research method, corrective changes can be proposed to improve the organisation's way of working.

4.2. Research Question

Every research starts with an area of interest and a problem within that. In order to carry out a case study design, a research question defining how to explore the issue is needed [26]. The area under research in the thesis is e-commerce, a highly digital and competitive industry requiring efficient processes for success. Being digital allows to collect data about various aspects and processes in the business, but the volumes of data are more than what analysts can manually model and analyse. Therefore, data-driven methods like process mining can be used to simplify data analysis of processes. Due to this, the main research question of the thesis is:

RQ: What improvement opportunities can be identified via process mining in the operational processes of an e-commerce company?

The research question is further divided into two parts – improvements for performance and customer experience.

RQ1: What improvement opportunities can be identified with process mining to enhance process performance?

The company in question is interested in reducing its process cycle time to be more efficient and deliver orders to consumers faster. When decreasing time spent per order/case, the company can reduce employees' workload, possibly even reduce costs, and use resources more effectively. Since e-commerce is getting even more competitive in this company's market, gaining an edge in performance can help the company stay ahead of rivals.

RQ2: What improvement opportunities can be identified with process mining to improve customer experience and satisfaction?

Since the firm operates in e-commerce, then consumer satisfaction is vital to have returning customers. In sales, it is always cheaper to keep existing consumers than to acquire new ones [27]. Through transforming sales processes to be more customer-centric and transparent, the company is able to keep old users and maybe even attract new ones. Research has found that personalisation can improve revenue as well as customer retention and satisfaction [28]. Due to this, it is imperative to study how to increase customer experiences within e-commerce.

4.3. Case Study Setting

When choosing a case for this study, four main requirements had to be filled: interest in improving processes, available data to extract event logs, granting access to the data, and allowing access to process domain experts. The company chosen for this case study is highly interested in improving and moving forward, always looking for ways to streamline their business. As e-commerce is highly growing, the need for improving their processes to be scalable has become even more apparent. Specifically, in light of the COVID-19 pandemic, the growth of e-commerce accelerated faster than expected [29], requiring additional workforce in the company.

In the past, the company has chiefly based decisions on confidence-based feelings and some manual analysis. These methods, however, have not identified many opportunities and do not uncover the root causes of issues. Therefore, the firm seeks to use data-driven methods to find new improvement opportunities within its processes.

The company has a total of three information systems (IS) that log data for online orders. Combining all of these IS event logs allowed to extract the operational process for e-commerce orders. Given the interest in improving their processes, the company granted access to these event logs. Also, to gain even deeper insight, contacts to domain experts and their knowledge was allowed at all times. However, due to the competitive nature of e-commerce, the company wished to remain anonymous for this case study.

The firm in question is a retail and wholesale company established nearly 30 years ago, which decided to enter e-commerce around 20 years ago (first web store launched in 2002). Over the past two decades, they have improved their website, automated processes and scaled the business to now offering approximately 500 000 products on their online shop. Their online store has everything ranging from cameras and consumer electronics to toys, cosmetics, and even gardening equipment. Monthly, they get on average 6 610 orders, ranging from 4 500 on slower months and up to 12 660 orders during pandemic restriction months.

Currently, due to COVID-19 and a general rise in online shopping, demand has increased, making it harder to provide the same level of service to customers. It has also increased the

workload for employees, creating more stressful situations and overworking. This study can help them identify weak spots with data-driven analysis instead of the manual methods used so far. For example, using statistics makes it possible to confirm bottlenecks in the system and make data-based decisions around changes in the process, rather than relying on confidence-based feeling.

The process selected for analysis in this research is the operational internal process for handling online orders. The customer side of the process could not be analysed since weblogs were not available, so analysis could not be made. However, the internal process is captured by the IT systems the company uses, enabling the creation of event logs that can be used in process mining. The process covers the company side from order confirmation to package delivery, making it possible to find inefficiencies and improvement opportunities for the firm.

4.4. Case Study Design

To better guide the execution of the research, a case study design was made (Figure 3). As the research questions and process selection have been explained under the case study setting paragraph, this section will continue from the data preparation aspect.

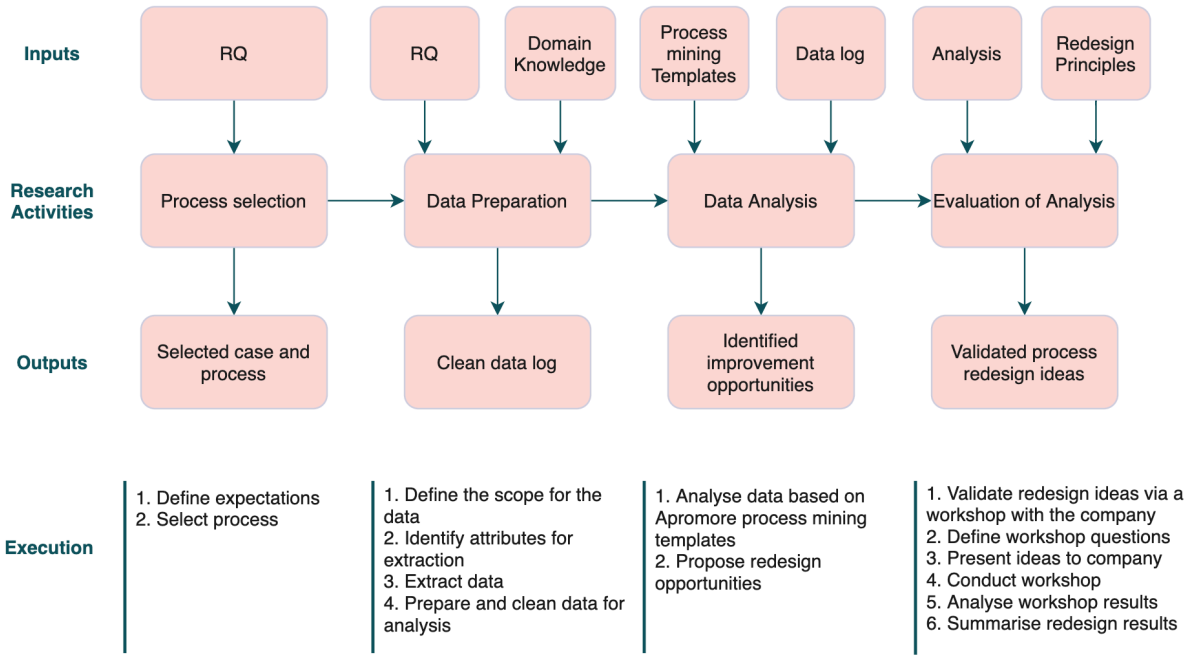


Figure 3. Research Process Design

Data preparation had to be done to organise the data for the analysis step. First, considering the research questions and company opinions, the scope of the data was set so that the most relevant information would be used in the analysis. Then, by consulting with domain experts, all relevant and available attributes needed to be identified so that the data would be ready for extraction. The details on the scope and attributes were forwarded to the person responsible for extraction. After receiving the extracted data, it was overviewed and cleaned so that a relevant

datalog would be ready for analysis. The event log processing and data analysis were illustrated and explained using figures and screenshots from the Apromore portal.

The third step of data analysis took as inputs the previously made data log and process mining theory to guide the analysis. A process mining methodology matching the research aim and the case study design was chosen for the theory. Since the company in this thesis is interested in improving performance, the study used the PM^2 method (Figure 4) for analysis, as it was specifically made to improve processes' performance and find compliance issues [30]. Additionally, Apromore process mining templates were used since PM^2 lacked the clarification in terms of concrete techniques to be used. The exact process of analysis is covered in a later paragraph detailing the analysis steps.

Finally, the evaluation of analysis was based on the data analysis results as well as redesign principles and heuristics to be able to propose improvement opportunities. Based on heuristics and results, the redesign ideas were proposed. To validate these ideas, they were presented to the company representatives, and a discussion workshop was held with the firm. Prior to the workshop, interview-style questions were gathered to guide the workshop. Afterwards, the workshop results were analysed and summarised by applying coding techniques to uncover if the redesigns were relevant and valid in the company's eyes.

To further explain the theoretical background, PM^2 comprises of six steps (Figure 4): Planning, Extraction, Data processing, Mining and Analysis, Evaluation, Improvement and Support [30]. Step 1 reflects case study selection and the process selection stages of the design in this thesis. Stages 2 and 3 of PM^2 are merged for simplicity in the case study design, as they require the same inputs.

The analysis and mining step is equivalent to the data analysis mentioned in the case study design (Figure 3). However, since PM^2 does not specify any process mining techniques, then Apromore templates for process mining were used as additional input. Lastly, stage 5 of PM^2 aligns with the last step of the case study design. This methodology also has step 6 for implementing improvements and supporting change via process mining. However, the implementation of changes is out of the scope of this thesis, thus is excluded from the case study design.

Below is a description of the PM^2 stages in more detail to overview the theoretical knowledge applied in this case study.

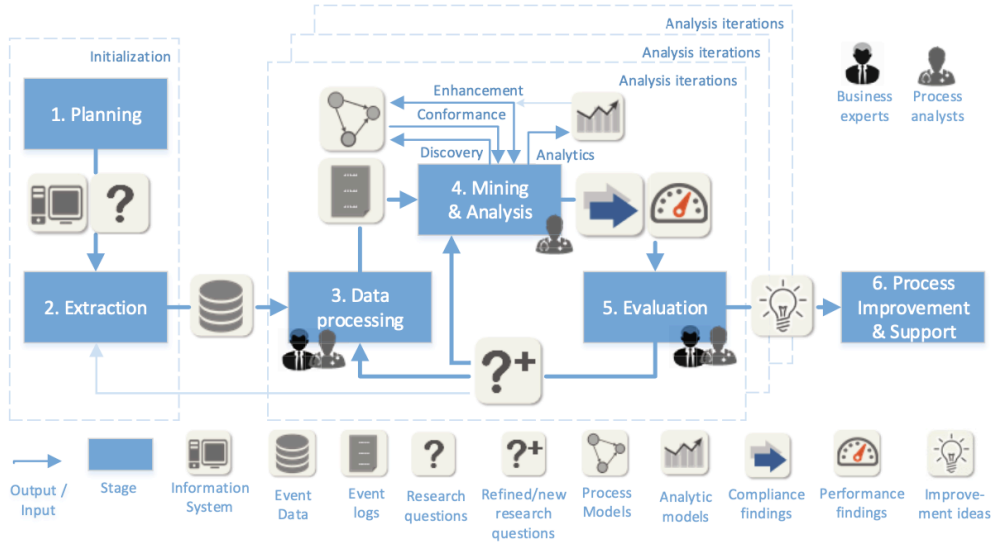


Figure 4. PM2 process description [30]

Step 1 – Planning

The planning stage is meant to set up the expectations and goals for the project. It starts with picking which process needs to be analysed based on data quality and how susceptible to change the process is. Process mining depends on good event logs, so processes without quality data will not achieve great analysis results. After this, project goals and research questions need to be set to understand what outcomes are expected.

Step 2 – Extraction

After process selection comes the data extraction. Before extracting, the scope needs to be set regarding granularity, time period range, attribute selection, and correlations between data. The next step is to extract the data and collect tacit process knowledge from business experts to analyse data more effectively.

Step 3 – Data Processing

The third stage is meant to prepare the data for analysis and mining techniques. Firstly, the data can be used to create various views for analysis to create specific event logs for each variant. To reduce the complexity of mining, another step of processing is aggregating events. For example, similar tasks of different classes could be grouped as one task but remain as separate events. Thirdly, the data logs should be enriched with additional data points. One option is to create additional attributes based on the available data or by using external data that could be useful in the analysis stage.

The final step of processing is filtering to reduce the complexity of the data log again and create different perspectives. Slice-and-dice filtering is used to remove unwanted events based on their attribute values and statistics. Variance filtering helps to group similar process traces so that analysis can simply distinguish between complex and simple process instances. Lastly,

compliance-based filtering is used to remove events or instances which do not conform to rules or regulations followed in the process.

Step 4 – Mining and Analysis

The analysis stage is where process mining techniques are applied on the event log to answer the research questions set in stage 1. The analysis covers four methods/steps – process discovery, conformance checking, enhancement, and process analytics. The point of process discovery is to get fact-based process models as output which could then be analysed further. Conformance checking compares the process model to the event log data, which simulates real-life behaviour. Comparing the two helps to find inconsistencies and compliance issues within the process.

The enhancement step is similar to the one in processing, where additional information about the process is used to develop insightful analysis. These aspects taken into consideration could be costs, resources, time spent. According to PM^2, process analytics is the final step for the analysis stage. Process analytics is a blanket statement for all other types of analysis methods besides process mining. It suggests using data mining and visual analytics to further enhance the process model.

Step 5 – Evaluation

After the analysis is done, results are used to develop improvement ideas, which are in line with the project goals. The first step of evaluating is diagnosis – interpreting the results, discovering exciting and unusual findings from the usual ones, and identifying further research questions for additional analysis. Secondly, the results need to be verified and validated by comparing the findings to the original data to uncover root causes for issues. During the evaluation, it is important to include business experts involved since analysts are not always the domain experts of the processes in question. Including experts in the process helps achieve results that are valuable to the organisation.

Step 6 – Process Improvement and Support

The final step of PM2 is used to implement the ideas from the evaluation stage, to make changes to the process in question. While the goal of process mining projects might be the achievement of process change, it can become a separate project itself. The results of PM2 analysis can be used as a base for starting a new project for implementing change in the organisation.

Process mining can be used as a source of support for implementing change. By using analysis techniques, process mining can uncover problematic cases or suggest new courses of action. For this stage to be successful, a high-quality IT structure is needed to link real-life results with process mining analysis.

4.5. Data Preparation

The following paragraph covers the data preparation of the research design (Figure 3). The first step of process selection has been already described in the research question and case setting part. To prepare data for analysis, a couple of steps were taken. Firstly, the scope of the data was defined according to PM2 to narrow down the selection of data needed. In terms of time, all orders from the year 2020 were selected to analyse a variety of cases over multiple seasons. Older data was not included since the most recent year is more reflective of the current situation the firm is in.

The main components of the dataset were:

1. The case ID's, which were unique to each webshop order;
2. The activity status (which activity is being performed in the process);
3. The start and end timestamps of each activity.

The original dataset included all possible attributes that the company had gathered data about to gain as much insight as possible. There were two types of attributes: case attributes and event attributes. Case attributes are specific to the order, but event attributes are unique to each activity in the order.

For case attributes, it was possible to collect the following data:

1. Payment method
2. Company (if the order was made under a company name)
3. Shipping method (how it was selected to be delivered)
4. Number of items (how many items the order has)
5. Total value (how much the order cost to the customer)
6. Buum class (defines if the order is a web order or other type of order)
7. Campaign (If the order was made during a campaign)
8. Refund (if the order was refunded)
9. Estimated delivery time
10. Actual Delivery time
11. DT Fulfilled (0 if ETA was not achieved, 1 if promised ETA was fulfilled)

In addition, for events, these attributes were available to add to the event logs:

1. Product ID (unique identification code)
2. Supplier (which supplier the product came from)
3. Supplier country (country where the supplier warehouse is in)
4. Resource (who ordered the products and made the invoice)
5. Category (product category of the item, for example, electronics)
6. Feature group (more specific category for the item, for example, smartphones)
7. Our warehouse (if the item came from the central warehouse or a shop's stock)

After extracting the data, the unprocessed dataset contained 90.7 thousand cases, with 509 variants and a combined 1.7 million events. Total characteristics can be seen in figure 5.

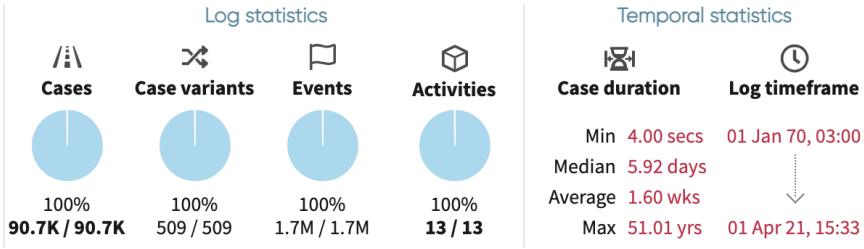


Figure 5. Dataset characteristics before processing

Before processing, the dataset went through scripting to add the attribute DTFulfilled. The original dataset included attributes for ETA and Actual delivery time, but no column stating if the ETA was fulfilled in each case. In order to analyse each case from the ETA fulfilment side as well, a separate column was needed so that Apromore could identify it as a case attribute and provide statistics about this aspect. In order to achieve this, the dataset underwent scripting. The actual delivery time and estimated delivery time were used to derive the corresponding arrival dates based on when the order payment was confirmed. These calculations were made keeping in mind that ETA is expressed in business days, so weekends and holidays were taken into account. After determining the dates for actual and estimated delivery, another column was made to calculate the difference between these two times. The DTFulfilled column was based on the previous column, so if the difference of times was negative, the SLA was broken, and the DTFulfilled equated 0, else the SLA was fulfilled, and DTFulfilled equalled 1.

Next, based on the statistics (Figure 5), it was clear that some processing and cleaning needed to be done to eliminate irrelevant cases and incorrect data to simplify the dataset for analysis. As mentioned, the dataset involved all orders made from the website, but not all of them are relevant to the study. The data log included data from orders that were made by the company's employees as well as a few other variants (like Swedbank bonus program orders). However, only regular webshop orders are relevant in this thesis. Thus, filtering was done so that orders with the label of web store orders were retained. These accounted for 83.4% of all cases, so the data was left with 75.7 thousand cases (Figure 6).

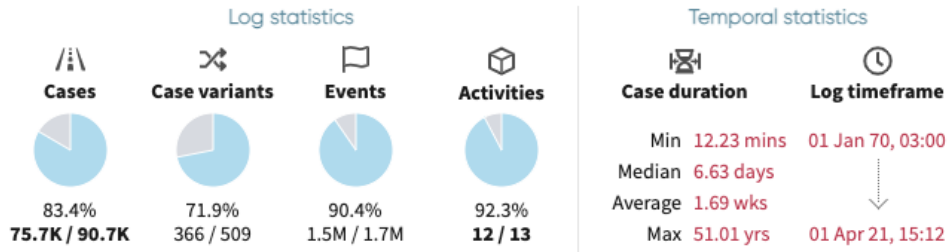


Figure 6. Dataset attributes after initial filtering

When looking at the characteristics, it was evident that there are some issues with timestamps since the maximum case duration is 51.01 years, and the log timeframe starts from the year

1970. To eliminate incorrect data and prepare the log for mining, additional filtering was applied based on PM[2]. The method includes two types of filtering: slice-and-dice, and compliance-based filtering, both of which were applied to the dataset.

Slice and dice filtering

Removing illogical performance: Case performance ranges from 12.23 minutes to over 51 years. In reality, the company has a maximum ETA of 31 business days which is 6.2 weeks, so all cases where performance is longer than 6.2 weeks were removed. For the minimum performance, the minimum delivery is one workday. Sometimes, the salespeople manipulate the system, and the delivery can even be 0.5 hours (direct pick-up from a physical store). Due to this, the minimum performance filter is put to 30 minutes. After all time filtering, 99.2% of the cases are still included in the data (Figure 7).

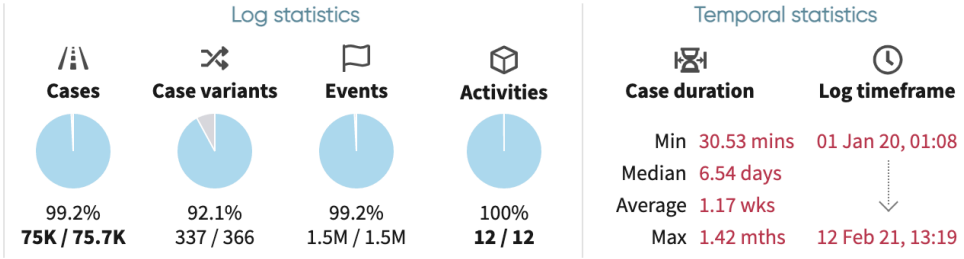


Figure 7. Dataset after performance filtering

Compliance based filtering

The internal order process in this company is set very distinct so that one activity cannot be performed before the other is completed. For example, any of the activities cannot be done before the order is confirmed and the payment is confirmed because salespeople will not receive the order information. Another example would be that the order cannot go to packing unless the invoice has been created because the packing system will not receive order information before an invoice is generated. The process follows a step-by-step logic, and all cases with different pathways have some system error in reporting timestamps and activity statuses.

Removing cases with illogical pathways: Rules were implemented (Figure 8) to ensure that no task breaks the conformance rules of the process and the company system. List of rules:

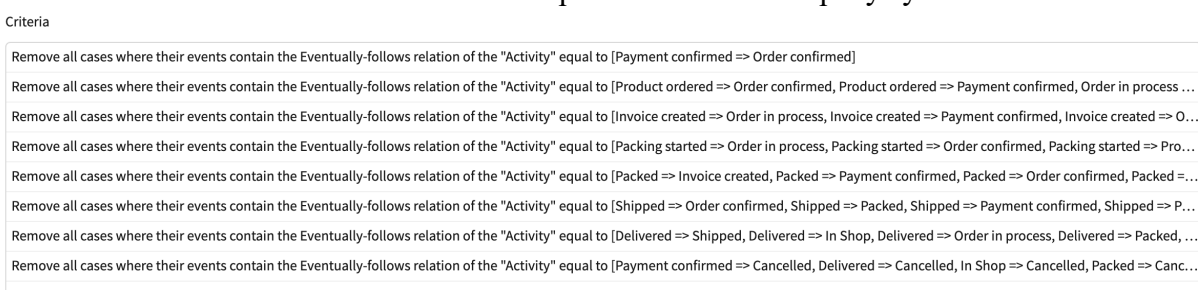


Figure 8. Example list of compliance rules applied to the data

After filtering, most cases were retained, but case variants were reduced, simplifying the mining process. For the analysis and mining, the data was left with 74.9 thousand cases with 267 variants and 1.5M events left (Figure 9), and the final process itself is illustrated in figure 10.

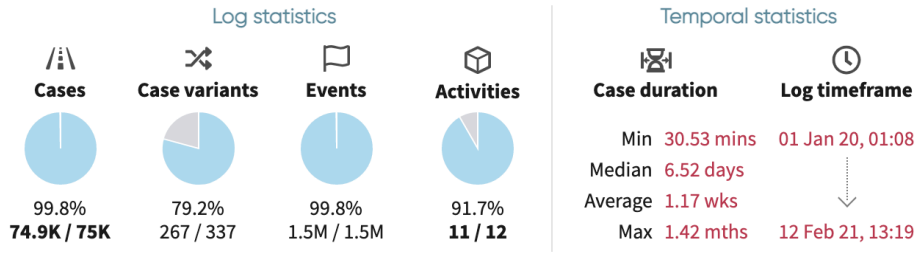


Figure 9. Dataset characteristics before mining and analysis

According to PM2, the data should be enriched with external knowledge as well. Outside factors could have shown certain tendencies (maybe rainy weather means more orders), but since the focus is to streamline the company's internal process, these attributes would stray away from the scope.

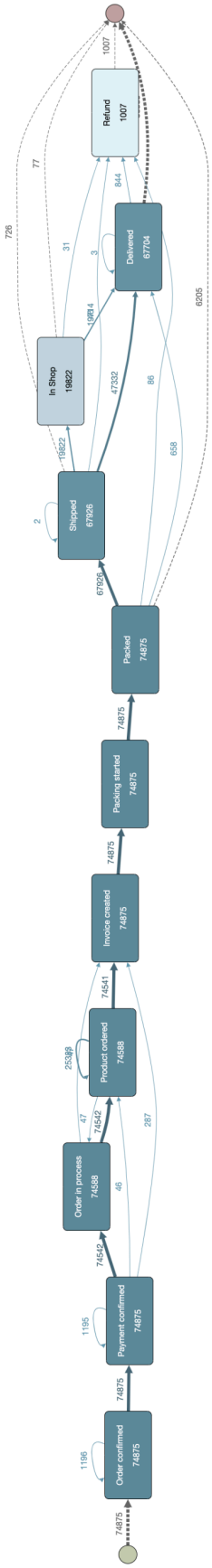


Figure 10. Final process model

4.6. Data Analysis

This chapter includes the description of techniques used in the data analysis to receive the results presented in paragraph 5. According to PM^2, the mining and analysis part included process discovery, Conformance Checking, Enhancement, and Process Analytics. Since the methodology did not have sufficient explanation for these analysis stages, then Apromore templates[1] were used to guide the analysis. Apromore templates are step-by-step guides explaining which techniques to use for what type of analysis. In total, there were four templates covering similar steps to the four stages in PM^2 (steps listed in table 3). These templates were used since they offered detailed instructions on how to detect, for example bottlenecks, compliance and flow issues, and resource utilisation by using process mining to visualize processes.

These Apromore templates are guides to conducting process mining analysis from a variety of aspects. These four templates cover themes from process discovery and compliance checking to performance mining and variant analysis. In table 3, the outline and order of the templates are described.

Table 3. Overview of Apromore templates and analysis components

Automated Process Discovery	1. Flow Analysis 2. Filtered Flow Analysis 3. Frequency Analysis 4. Handoff Analysis
Compliance Checking	1. Flow Compliance Checking 2. Temporal Compliance Checking 3. Resource Compliance Checking 4. Exception analysis
Performance Mining	1. Bottleneck Analysis 2. Workload & Demand Analysis 3. Rework Analysis 4. Over-processing analysis
Variant Analysis	1. Flow Comparison 2. Frequency & Rework Comparison 3. Bottleneck Comparison

Each templated started with a table summarising what needs to be analysed and how it should be done (Figure 11). Then, the templates continued with explaining each of the phases and techniques even further.

[1] Apromore Process Mining Analysis templates are part of a corporate training offered by Apromore. Access to the templates was granted for this research.

Automated Process Discovery: Analysis Template

What?	How?
Flow Analysis • Analyze the process structure & main case variants • Identify parallelism, branching points & rework loops • Analyze case entry and exit points, and check for incomplete cases	• Visualize the most frequent case variants using Case variant filtering and the Case inspector • Use abstraction sliders to focus on the most frequent activities and dependencies. • Switch to BPMN view; inspect the behavior around the gateways • Check the dotted lines in the process map (arcs emanating from the start event or leading to the end event). Unexpected dotted lines indicate some cases are incomplete
Filtered Flow Analysis Analyze different slices of the event log (groups of cases) or fragments (groups of events) separately	• Use Case filtering to retain/remove specific cases, e.g. cases that have a specific path of activities • Use Event filtering to retain/remove subsets of activities, e.g. separate automated or non-core activities, or separate activities belonging to different systems.
Frequency analysis Analyze the most frequent activities and transitions	• Use the Activity inspector • Use the color-coding in the process map to find most frequent transitions • Consider switching between multiple frequency metrics (max, avg, …) • Use the Activities tab (frequency view) in the Dashboard plugin
Handoff analysis Analyze handoffs between workers, teams, groups, org units	• Switch between perspectives in the Process Discovery plugin and assess handoffs between resources • Use the Resource tab (frequency view) in the Dashboard plugin

Figure 11. Example of Process discovery template outline

All steps included images to showcase how the analysis needs to be done in the Apromore portal (Figure 12) and a detailed step-by-step guide down to which filters to use and buttons to push. For example, for conducting frequency analysis, the template (Figure 12) clearly states which button to toggle to achieve the frequency view in Apromore.

Frequency Analysis

Apromore can visualize the process map by using color – coding techniques to find the most frequent arcs.

1 Analyze the most frequent activities and relations

By selecting the "Max" ("Min") option from the "Frequency" drop-down menu, we can see the maximum (minimum) number of cases activities and transitions have been observed in.

For example, there is at least one case where activity "Repair (Complex)" was performed four times.

Figure 12. Example of instructions on how to conduct frequency analysis

The analysis itself followed the outline pictured in figure 13. Most of the steps listed in table 3 were followed within the analysis, but not all were relevant, so some were skipped.

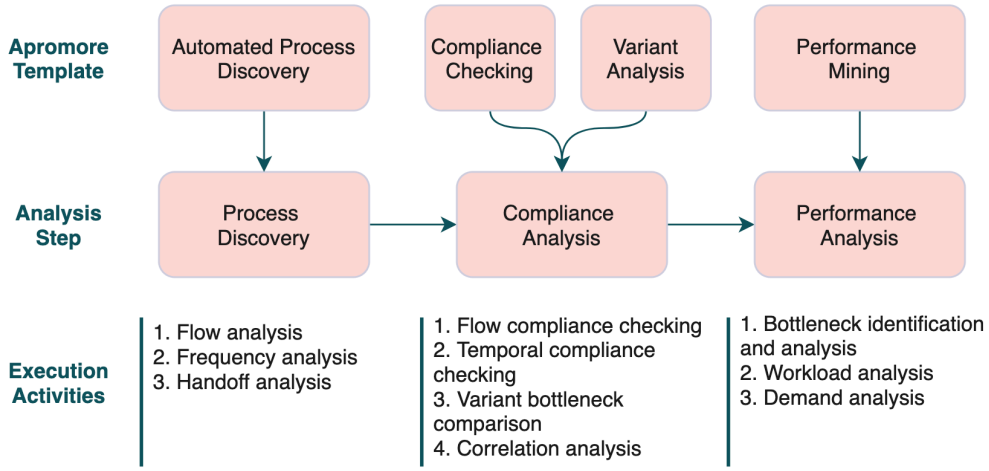


Figure 13. Outline of the analysis process followed in the research

Automated Process Discovery

The first step was about process discovery, where visualisation and filtering were used to discover differences in how the process was conducted. Process discovery included flow analysis, frequency analysis and handoff analysis. Filtered flow analysis was not done per se since different types of filtering and case slicing was done throughout the analysis, so a dedicated analysis chapter was unnecessary.

Flow analysis was based on the process map and the BPMN model to identify parallel tasks, rework, and incomplete cases. It was done mainly by using the visualisation tools Apromore provides. Then frequency analysis was done by using the dashboard functionality and colour-coding options to find irregularities in frequencies. Lastly, the template could not be used for handoff analysis since the data log did not include resource handoffs between different resources. However, task handoffs were discussed and analysed for parts of the process which the log fails to capture. The analysis statistics from Apromore were used to visualise certain parts of these tasks not captured by data logs.

Compliance Analysis

Secondly, compliance analysis was done to identify if any violations of rules occur in the process (for example, mandatory steps are skipped). This step included flow compliance checking and temporal compliance checking. Additionally, from the variant analysis template, variant bottleneck analysis was applied, and correlation checking was done outside of Apromore template instructions.

The first step of flow compliance was to understand if any mandatory steps have been skipped, thus enabling to identify any errors in the process. Then, temporal compliance analysis helped to understand if the company is keeping their service level agreement (SLA) - delivering items on time - to the customers. To further understand the issues of temporal compliance, variant bottleneck analysis was used on both variants (where SLA was fulfilled and where it was not)

to gain insight into the root cause. Lastly, a small correlation table was made to find compliance correlation causes.

Resource compliance checking was not done for a similar reason why handoff analysis could not be executed. Since the process does not include resource interactions and restrictions on cases, then there was no compliance rule to be analysed. Additionally, exception analysis and further variant analysis was not implemented since it focuses mainly on different and infrequent case variants. However, this data log had 267 case variants, where mostly the difference was due to one activity, and it did not break any process rules. Thus, variant analysis would not have provided any valuable insight into the process.

Performance analysis

Lastly, performance mining was applied to the event log to understand which attributes cause delays in the performance of the process. It comprised of bottleneck analysis, workload analysis and demand analysis.

To find bottlenecks, process maps overlays were used to identify the longest taking waiting points in the process. Activity bottlenecks could not necessarily be found since the timestamps of the event log only allowed to see the time it took to get from one activity to the next. Workload and demand analysis was made to understand if any resources were overutilized and see how demand affected the work of the process. Apromore dashboards gave an overview of resource utilisation as well as demand frequencies. Additional log slicing and filtering were done to compare different variants of the log in the dashboard.

Variant analysis

As mentioned already, variant analysis was not deemed to be necessary for this process log due to the nature of the case variants in the data. Additionally, differences in attributes in the cases and their effects on the process were discussed in the above analysis parts. Thus, this analysis step would not provide additional valuable insight into answering the research questions since the variance is mainly related to one activity.

4.7. Evaluation of Analysis

The last stage of the research evaluated the analysis results by providing redesign ideas and validating the proposed improvements based on the company's opinion. Firstly, redesign ideas were proposed based on analysis results and redesign principles and heuristics to find answers to the research questions.

In order to evaluate the need and relevancy of the proposed changes, a workshop was carried out with the company representatives. The redesign ideas were presented to the company one by one, and then questions about relevancy, strategic fit and implementation difficulty were asked. Instead of conducting individual interviews, the session was held in a workshop format

to facilitate conversations from different perspectives and get immediate feedback on proposed redesigns. It also allowed for different positions in the company to better understand the needs and restrictions of others.

In total, there were three people selected to participate – a sales specialist, the head developer, and the CEO. The sales specialist was included to gain insight into how the changes help with the everyday work with the process. The developer was included in the discussion to understand the limitations of the changes in terms of IT systems. Finally, the CEO was interviewed to find out how these changes play into the overall strategy and direction the company is taking.

The workshop lasted 1 hour, during which the redesign opportunities were proposed, and after each redesign, questions were asked to facilitate discussion around validity. Firstly, a description of the data was presented to give an overview of which data was included in the analysis. It contained the data set description, selection of attributes, and the most noticeable data trends. This was followed by an explanation of the research questions and the issues found regarding each question. The significant analysis results were presented one-by-one where each was illustrated with the essential numerical impacts and figures. The presentation was structured so that each issue had analysis highlights for justification, followed by possible redesign opportunities and corresponding impacts if these changes would be applied. To address the severity of specific results, statistics and averages were presented to illustrate the situation. Additionally, bar charts and colour-coded tables were shown to draw attention to the most crucial bottlenecks. The analysis and impacts focused mainly on the change effect on resources, cycle time, and customer experience.

The workshop was semi-structured to guide the discussion but leave room for additional questions and comments. For each redesign, four questions were presented:

- Do you consider this redesign as an improvement opportunity?
- Were you already aware of this issue/opportunity?
- Do you think this redesign is relevant and has the wanted impact? If not, then why?
- How likely are you to implement something similar to this?

Besides these questions during the workshop, the participants were also asked to rate each idea from their perspective to evaluate the relevance of the research to the company quantitatively. Each participant received a survey after the workshop to avoid group-think in the results. In the feedback form, they had to indicate on a scale of 1 to 5 their opinion of the redesign opportunity. The CEO and sales specialist filled it from a relevancy perspective to grade how important is the change to the daily work and company, and the developer assessed the ease of implementation of the redesign. For relevancy, the scale was interpreted so that one meant not relevant and will not implement, and five stood for highly relevant and need to implement. For ease of implementation, the scale indicated five for hard to implement redesigns and one if the idea was straightforward to implement.

To analyse the survey, the CEO and sales specialist results were averaged to get a score of relevancy for each redesign. As the developer was the only person who could assess the ease of implementation from a technical standpoint, the results had to rely on one set of numbers. In order to visualise the attractiveness and importance of each result, an action priority graph was made. The action priority matrix is a graph where ideas are plotted based on importance and efforts to understand which projects should be implemented next [31]. The matrix includes four quadrants: Quick Wins (high impact, low effort), Major Projects (high impact, high effort), Fill-ins (low impact, low effort), and Thankless Tasks (low effort, high impact) [31]. For this study, the impact aspect was replaced with relevancy as a factor to make a guide for the company on which redesigns they should implement first. To make the matrix, an open-source platform called Miro was used.

Due to the confidentiality of the company's strategic aim, the session was not recorded and thus not transcribed. However, notes were allowed to be taken during the interview regarding the participants' reactions, answers and ideas. These notes were coded and organised regarding the research questions to identify relevancy for both questions. Additionally, the survey scores were added to each of the redesign opportunity to identify the most minor and most viable ideas.

5. Results

The following chapter contains the analysis results based on the aforementioned research questions. The main goal of the results is to find answers to what improvement opportunities can be identified with process mining techniques. Results are based on the Apromore templates, which were used as a reference point to guide the analysis.

5.1. Process Discovery

Flow analysis

Firstly, flow analysis was used to identify parallelism, rework and incomplete cases. No parallelism was apparent from the process map, but rework and incomplete cases were identified (Figure 10). There was some **repetition by self-loops** in the process (Figure 10). In 1.6% of the cases, the "Order confirmed" activity was repeated. When analysing the system behaviour further, it was apparent that this activity was repeated when the customer switched their payment method (Figure 14). However, it is unclear why the system takes payment method selection as a new confirmation of the order (instead of the confirm order button in the shopping cart).

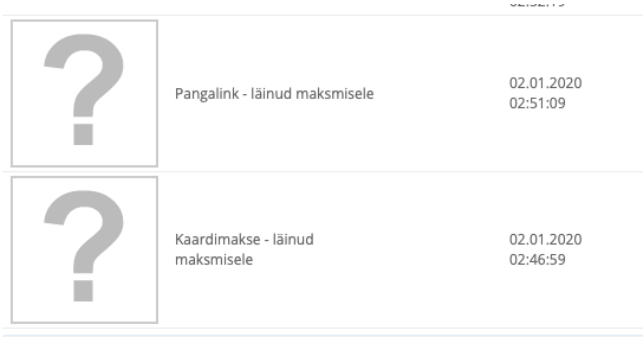


Figure 14. Customer changed payment method from credit card to bank link

In 1.6% of cases, "Payment confirmed" was also repeated. In most cases, it happened because the system thought that "Hire purchase accepted" is the same as "Payment confirmed", thus recording the activity twice instead of logging them as different activities (Figure 15). In some cases, the double confirmation reason could not be identified since the system just showed double payment confirmation, both with the same timestamp.

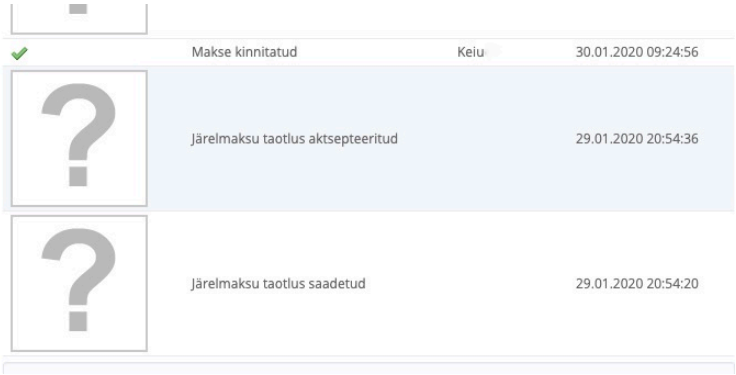


Figure 15. Payment confirmation of a hire purchase order

From figure 10, it can also be seen that “Product ordered” was also repeated a lot, but this repetition was logical since a case with eight products has the “Product ordered” step done eight times. Other repetitions in the process were minimal and mainly represented human error, for example, someone had marked the order as delivered twice by accident.

Lastly, incomplete **cases** were identified to find the root cause behind orders not completing. There were three clear trends of cases that did not get to “Delivered” status and ended not completed. There were 7008 cases that did not go on from “Packed” status, 803 cases that did not go forward from “Shipped” status and 77 cases that ended with “In Shop” status (Figure 16).

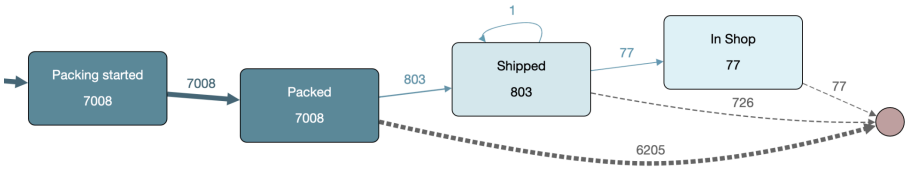


Figure 16. Incomplete cases with no correct ending

The “In shop” orders were cases where the customer had not come to pick up the package yet, or the shop employees forgot to mark that an order was picked up. For orders that get stuck in the “Shipped” phase, there are two possible reasons:

1. Shop employees have not updated that the package has arrived to the shop, meaning that the customers also have not received a notification that the order is ready.

2. The system had known incompatibility issues with DPD tracking numbers, so status updates did not reach the company system after dispatching. This is illustrated by figure 17, where it is seen that most incomplete cases are related to DPD (Kuller and DPD pakipood both operated by DPD).

Value (0 / 7)	Cases	Frequency
Kuller	3422	48.83%
DPD pakipood	2421	34.55%
Omniva pakiautomaati	695	9.92%
Smartpost	246	3.51%
PH poodi	153	2.18%
Smartpost Finland	63	0.90%
Matkahuolto	8	0.11%

Figure 17. Shipping methods for cases with status "Shipped"

Value (0 / 6)	Cases	Frequency
Kuller	3355	54.07%
DPD pakipood	1965	31.67%
Omniva pakiautomaati	634	10.22%
Smartpost	245	3.95%
Matkahuolto	5	0.08%
PH poodi	1	0.02%

Figure 18. Shipping methods for cases with the status "Packed"

With “Packed” end status, the orders account for 8,3% of all orders in the study, thus the root cause should be analysed further. This happens when the system has no tracking number associated with the order, so the order status updates do not reach the company’s system. However, this issue also means that the customer does not receive any updates about the status

anymore. This status has the same problem with DPD orders as the previous one (Figure 18 shows most cases related to DPD again), but there are a few other reasons it could happen.

1. For example, the system can only record one tracking number, but orders with many large items are shipped separately and have many tracking numbers. However, most orders only contain 1 item (Figure 19), suggesting that this issue is not the main reason for this incompleteness.
2. Another cause could be that customers have contacted the company to change their shipping method, thus the system does not record the new method, and the statuses get stuck in “Packed”. Any of the systems do not record the changes in shipping method, thus this option could not be analysed further.

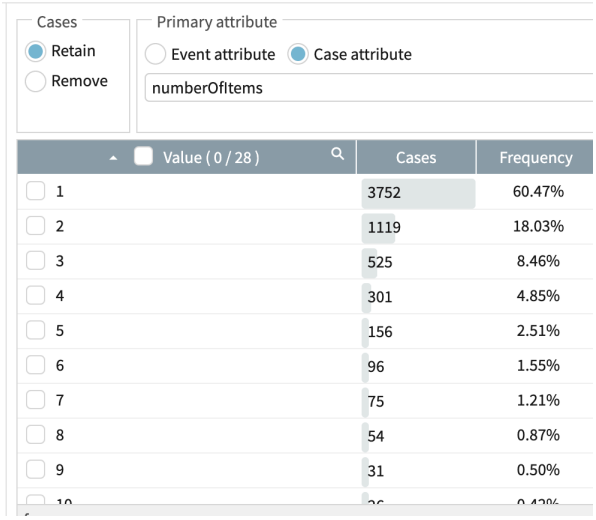


Figure 19. Most frequent number of items for cases ending in "Packed" status

Frequency analysis

After flow analysis came another part of process discovery called frequency analysis, which helps analyse most frequent activities in the process and find causes in different frequencies. From analysing the process dashboard, it was seen that all activities are done equally, except “In shop” and “Refund” activities (Figure 20). The differences for “Shipped” and “Delivered” have the same reasoning as brought out in the flow analysis.

The number of refunds is low compared to the volume of delivered cases based on the event log. Conversely, when consulting with domain experts, they estimated that around 25% of refunds are done immediately after payment confirmation. These refunds, however, were not captured by the data log due to a data error. Nevertheless, when investigating further why these refunds happen immediately after payment, the reason turned out to be the ETA. The current customer UI only shows items’ ETA on the individual product page, and then the overall ETA is displayed in the order confirmation letter. According to customer support and domain experts, consumers overlook the ETA on the products’ page and assume that since the products

show status “in stock”, they will be delivered in a few days, thus creating many refunds after the customer gets the order confirmation letter.

Moving to the “In shop” status, its frequency indicates that most customers prefer alternative shipping methods to picking up from the store.

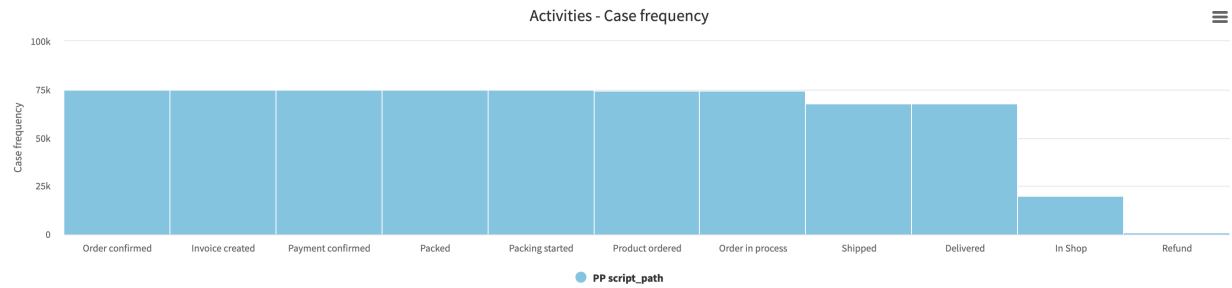


Figure 20. Activities frequency dashboard

When looking at the distribution of shipping methods in cases (Figure 21), Omniva is the most popular and only 26.69% of cases are ordered with pick-up in the stores. The variance in shipping methods is influenced by several methods from price to customer, delivery speed, to purely consumer preferences over shipping companies. The results are also influenced by COVID-19 since, during 2020, shops were closed at certain weeks, so customers had no choice but to get delivery via parcel terminals/couriers. To understand the divide between shipping companies, additional consumer analysis should be done to determine the root cause of these preferences.

Value (0 / 7)	Cases	Frequency
Omniva pakiautomaati	23130	30.89%
PH poodi	19987	26.69%
Smartpost	12508	16.71%
Kuller	11317	15.11%
DPD pakipood	7757	10.36%
Matkahuolto	112	0.15%
Smartpost Finland	64	0.09%

Figure 21. Distribution of shipping methods in the dataset

The data log did, however, enable to identify the delivery speeds for each of the methods (Figure 22). Orders were ready for pick up in the stores in 21.58 hours, making it the fastest option for the customers. Thus, the second fastest, thus more consumer-friendly, was Smartpost, with delivery in 28.18 hours, and the longest was DPD parcel terminal with 51.36 hours. These results give opportunities to explore promoting certain shipping methods over others, but costs to the company should also be considered (which method is cheapest to operate with). The costs of delivery partners were not included in the data, so they could not be analysed.

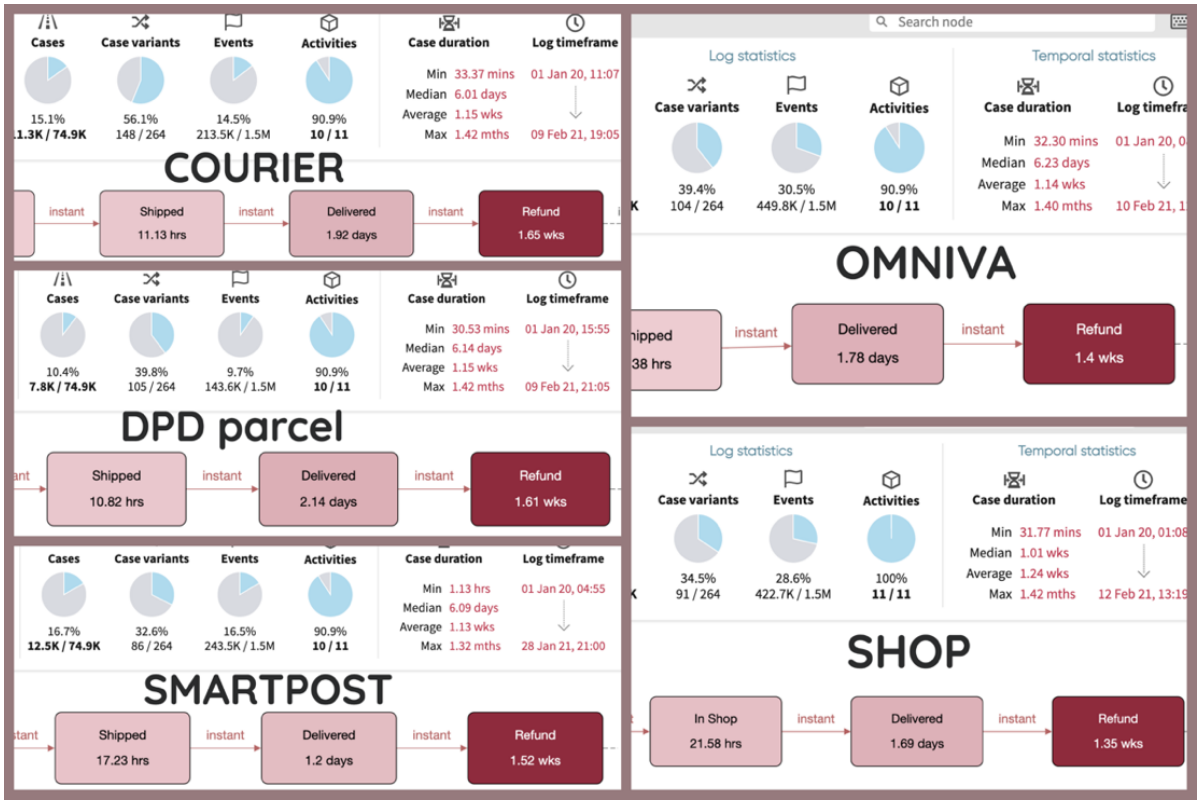


Figure 22. Shipping methods duration differences

Handoff Analysis

The final part of process discovery was handoff analysis, which is used to find any resource handoff related issues. According to the event log, the process does not include handoffs since each resource interacts with the system and not with other resources directly. On the other hand, according to domain experts, the parts of the process not recorded by data have some handoff issues.

Firstly, when customers pay via bank transfers, the accounting team has to manually check the incoming payments. Then, this information is forwarded to the sales specialists, who manually confirm the payment in the system and adds payment information to the order. Almost 4% of orders are made via bank transfers, and the average waiting time to get the payment confirmation is 1.28 days, whereas, in other methods, it is almost instant (Figure 23). This processing way includes a lot of information transport and waste and manual operations that should be removed or reduced.

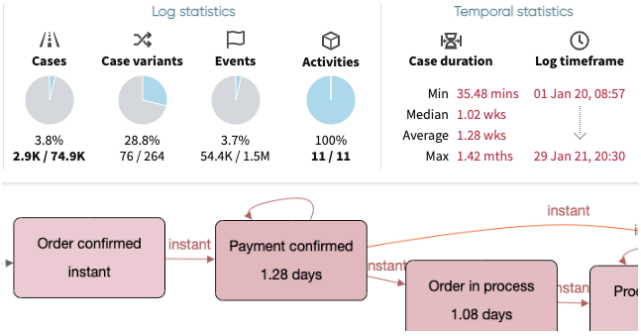


Figure 23. Statistics for cases with bank transfer payment method

According to domain experts, there are also possible issues with product ordering. Currently, all suppliers are divided among sales specialists, and each person is responsible for ordering products from their supplier. Then, each incoming order is divided among sales specialists manually according to whose suppliers are most apparent because this person will then be responsible for making the invoice and gets the sales commission from the order. However, if the order contains other specialists' suppliers as well, then the responsible specialist depends on other employees and if they have ordered their products on time. If products are not ordered on time, then the order will not get to the customer in the promised ETA.

When looking at the average SLA fulfilment rates for all suppliers, the most problematic ones can be identified from figure 24[2]. Supplier 79 has the worst SLA fulfilment, where only 10.53% of cases get delivered on time. Based on this, most late suppliers can be addressed to ensure a higher SLA fulfilment rate. According to domain experts, the lateness of the suppliers mostly depends on the sales specialist's ability to order items on time. It is rarely the suppliers' fault that products get stuck in delivery and are late. Thus, items ordering by sales specialists should be addressed. The complete list of supplier analysis can be found in the appendix.

Supplier	SLA = 0	SLA = 1
17	32,65%	57,09%
76	27,31%	58,95%
7	43,84%	46,16%
57	29,16%	57,88%
52	29,70%	44,73%
8	33,68%	57,03%
99	58,79%	30,31%
101	35,85%	50,81%
13	27,23%	55,38%
34	35,44%	54,95%
20	32,12%	35,97%
95	21,02%	46,50%
36	32,63%	51,58%
91	20,65%	47,83%
73	36,21%	44,83%
21	41,07%	44,64%
75	19,44%	36,11%
72	36,36%	51,52%
103	32,25%	40,63%
80	28,57%	57,14%
112	50,00%	26,92%
30	41,67%	41,67%
79	63,16%	10,53%
116	61,54%	30,77%
96	38,46%	23,08%
48	70,00%	20,00%
16	50,00%	50,00%

Figure 24. Most late suppliers according to SLA fulfilment

5.2. Compliance Analysis

Firstly, by looking at the flow of the process, **skipped activities** were identified (Figure 25). Apromore allows visualising processes by minimum frequency, which highlights possible violations.

[2] The percentages of the two columns do not add up, since the rest of the cases had SLA = 2, which means delivery time was not available to calculate fulfilment.

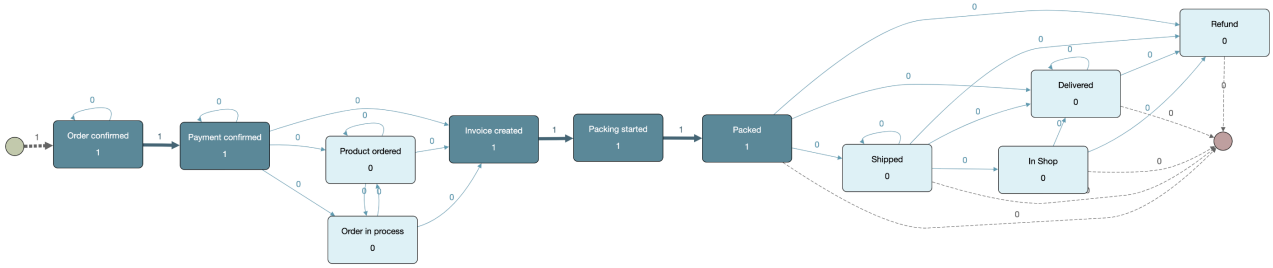


Figure 25. Minimum frequency of performed activities

The missing activities of "Shipped", "In Shop", "Delivered" are explained by process discovery analysis since some orders are not completed, thus do not have these activities. The "Order in process" and "Product ordered" should, in theory, happen in every case, but there are 287 cases where this has been skipped. When analysing this more in-depth, it seems that the orders are related to products available straight from the warehouse. There were two possibilities here:

1. The invoice was made so quickly that the system did not have time to update the order status.
2. The item was out of stock, and upon arrival, it was not ordered through the system, so the order did not go through these two steps. Skipping these two steps decreases the transparency for customers and possibly also customer satisfaction.

Secondly, **temporal compliance checking** was done to understand which cases did not arrive within the promised ETA, thus violating their service level agreement (SLA). To measure this, the scripted column of DTFulfilled was used. A value of 0 meant that the ETA was not fulfilled. The value of 1 indicates that the package arrived in time, and 2 signified cases where the actual delivery time was missing from the data, so fulfilment could not be calculated. Overall, only 64.13% of cases arrive to customers in time (Figure 26), which means that over 19.5 thousand cases were unsatisfied since they did not arrive when promised.

Value (0 / 3)	Cases	Frequency
0	19688	26.29%
1	48016	64.13%
2	7171	9.58%

Figure 26. Delivery time fulfilment in all cases

When analysing the cases where **SLA was broken**, it is seen that the biggest bottleneck is waiting for products to arrive at the warehouse so that an invoice could be made. On average, the company waited 5.21 days for merchandise before invoices could be made (Figure 27). Another bottleneck was around delivery time. Based on domain experts' knowledge, each item should be delivered by the next day, but the average delivery time is 2.86 days (Figure 27). These bottlenecks indicate issues with suppliers as well as delivery partners. This analysis further confirms the above differences between suppliers SLA and different shipping methods delivery times as well.

Figure 27. Average duration of activities in cases where SLA is not fulfilled

Compared to cases where SLA is fulfilled, there are significant changes in the process times. On average, invoices can be made one day earlier, and items are delivered 1.4 days faster as well (Figure 28), meaning that something has to be done differently in these cases.

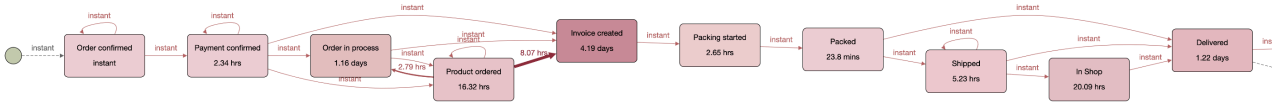


Figure 28. Average duration of activities in cases where SLA is fulfilled

When consulting domain experts, attention was brought to the fact that customers cannot see the "product ordered" status, so for them, the whole time, the product is in "order in process" status. This means that the customers are not aware of what is happening with their order, and on average, they wait 4-5 days before getting a new status update. Customer support specialists also mentioned that consumers are often confused by the term "order in process", and they get calls and emails daily asking about the order status to get an update.

It should also be considered that the SLA fulfilment can be influenced by how fast the customers went to pick up the order. The "Delivered" timestamp comes when the customer receives the package, but sometimes consumers do not always go to pick it up on the same day that the package arrives. Thus, it is possible that the SLA was kept in more than 64% of cases (package arrival to parcel terminal/shop), but the customer did not come to pick it up.

In order to further analyse which attributes are correlated with SLA fulfilment, a correlation table was made. The analysis shows that no single attribute is heavily correlated with fulfilling the promised ETA to customers (Table 4). The most significant correlation is with the shipping method, indicating that certain delivery partners provide better services than others. There is also a slight negative correlation between delivery time and campaign, which indicates that orders made during a campaign are a bit more likely to be delayed.

Table 4. Attribute correlation with delivery time fulfilment

	DTFulfilled
DTFulfilled	1
supplier	0,08199429
resource	0,03476669
numberOfItems	0,05619631
shippingMethod	**0,22281767**
campaign	**-0,0837708**
totalValue	0,07938527
paymentMethod	0,08817146
supplierCountry	0,06162156

5.3. Performance Analysis

The third stage of analysis was performance mining, which consisted of bottleneck analysis, workload, and demand analysis. In terms of **bottlenecks**, the largest was waiting for the invoice to be made (figure 29). The invoice cannot be made until all products reach the main warehouse, so the process waits behind suppliers to deliver their merchandise. The bottlenecks also became apparent in the first two steps of the analysis, further confirming the issues' importance.

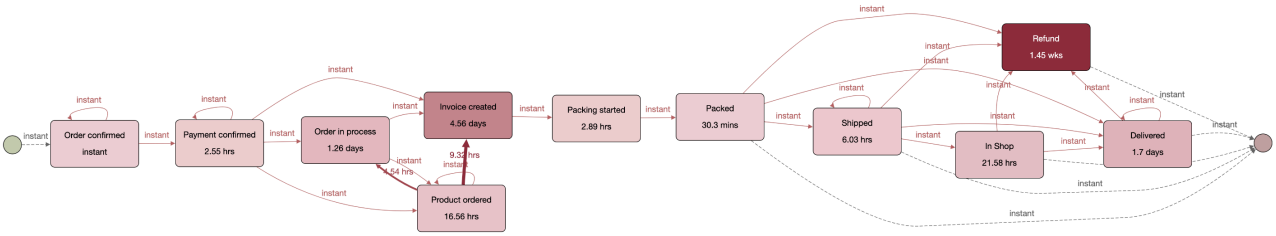


Figure 29. Average duration of activities and waiting times

For **workload division**, the Apromore dashboard gave an overview of how cases are distributed between resources (Figure 30). The case frequency calculates all cases where the resource interacts with a case, meaning invoice making and product ordering. Thus, the figure shows the overall workload and involvement of a person in cases. One thing to keep in mind is that Selena joined the team in April and took cases from Keiu to lessen her workload. Selena was then replaced by Maria in June, meaning that Keiu's caseload would have been the biggest without additional workforce. Considering that all employees work with the same contract, the workload division should be more equal to avoid burning out.

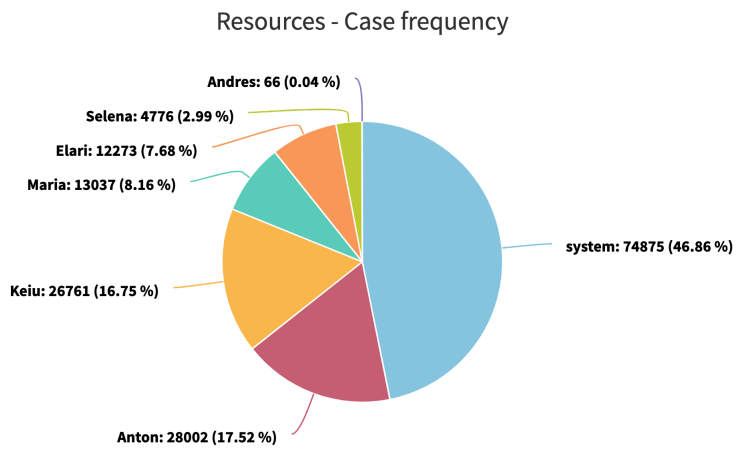


Figure 30. Division of cases between specialists

In addition to the workload division of the sales specialists, the analysis also took into account another resource related to the process - the physical stores. When an order contains products not in stock in the main warehouse, it is ordered from one of the physical locations. Currently, the system is built to order the products from shops based on a predetermined sequence, which is reflected by the distribution of cases. This, however, creates an uneven workload for store employees since some locations have more than three times the items to handle than others (Figure 31). According to domain experts, the system also clashes with sending products to the stores to be sold. It frequently happens that a product is ordered from a store where it just arrived, so the product itself does not even make it to the store's shelf. Whereas in another store, an identical product has been sitting on the shelf for a long time. This task is currently done semi-manually by sales specialists, so when items are ordered too late, they do not make it to the following transportation cycle, and orders are possibly delayed.

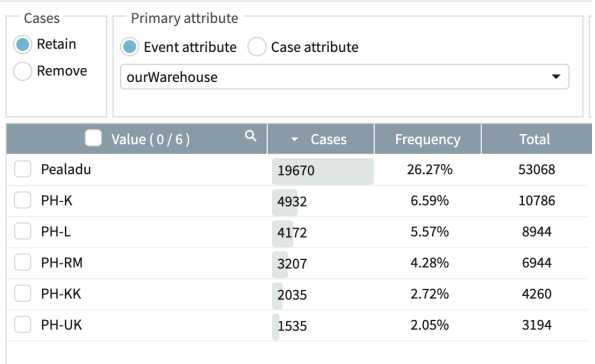


Figure 31. Case distribution for physical stores

Finally, **analysing the demand**, there are definite periods in the year where demand was more significant (Figure 32). For example, there are more active cases during the April-May lock-down when shops were closed. Another peak can be seen in November when the company has its annual biggest sale and the Black Friday sale. This high peak continues in the whole of December, due to holiday shopping.

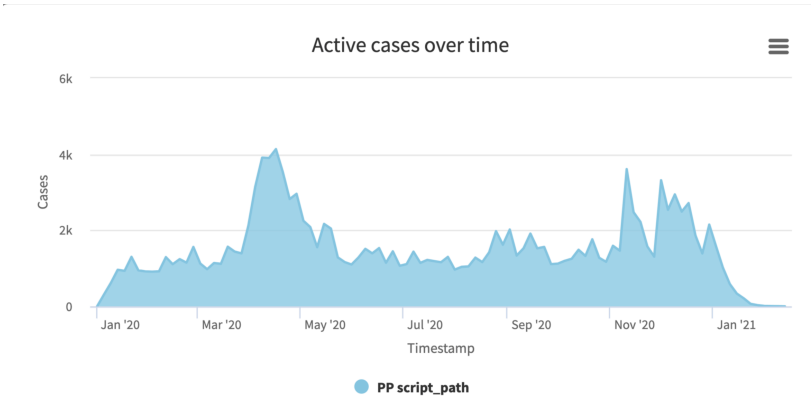


Figure 32. Active cases over the year 2020

Another graph was made comparing SLA fulfilment to understand what impact this varying demand has on process performance. When plotting cases where SLA was fulfilled to cases where it was not, it can be seen that during these peak times, late deliveries are more frequent (Figure 33). This could be due to an increase in workload and inability to handle it from the company's side, but it can also be influenced by COVID-19 implications on deliveries and transportation worldwide. To avoid such radical differences in SLA fulfilment, the process should be redesigned to handle sudden increases in cases.

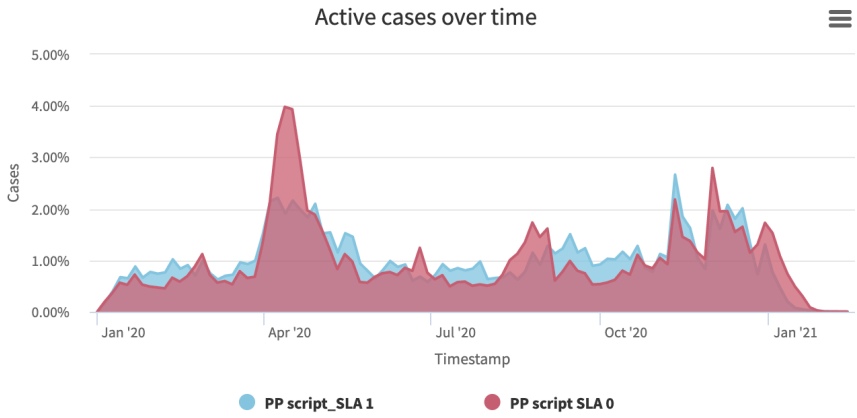


Figure 33. Relative active cases for SLA variants

5.4. Manual Tasks Analysis

The operational process involves two additional manual tasks, but the event log could not capture the time spent on these activities. Thus, analysis of these tasks was based on domain experts knowledge and statistics around case attributes related to these tasks.

Firstly, the process includes a step where all incoming orders are manually divided between sales specialists based on specific business rules, mostly revolving around suppliers connected to the order. On average, there are around 1440 cases each week that need to be divided, thus taking even up to an hour each day. However, when looking at the case attributes, it is seen that 67% of cases only have 1 item in the order, meaning that an automatic system could be easily put into place.

The second manual process is making invoices. During busier periods, the sales specialist can make altogether more than 900 invoices per day, which again occupies a large portion of their workday. To make matters worse, the invoicing is unevenly distributed between resources, similarly to the resource analysis results (Figure 35).

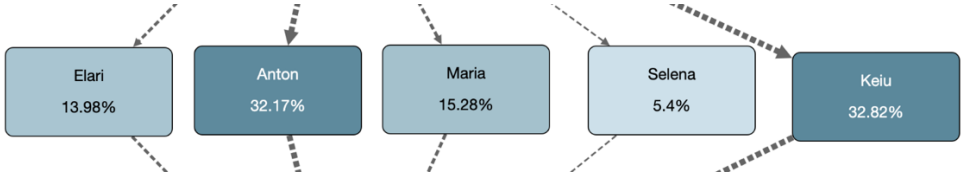


Figure 34. Invoice making case distribution between resources

5.5. Summary of Results

To conclude, the analysis identified issues within the process in aspects where performance improvements could be made and problem areas where customer experience could be enhanced further. The main findings are presented in table 5.

From the process discovery step, multiple issues arose. Flow analysis allowed for issues regarding incomplete cases and incompatible systems to be detected. Frequency analysis and variant comparison of the process enabled to identify possible cycle-time inefficiencies that should be addressed. Investigating and visualising handoffs allowed to detect unnecessary manual tasks as well as explanations regarding SLA fulfilment. Conversations with domain experts about the process flow additionally revealed a lack of transparency in the process from the customers' point of view. Additionally, consumers were not aware of their ETA's while ordering, thus being disappointed when it turned out longer than expected.

The second method was compliance analysis, where SLA fulfilment was the main issue that surfaced. By using variant analysis again, the root causes of not delivering orders on time were found. In addition, to case attribute contributed issues, the process also showed problems around scalability in high-demand seasons.

Finally, process performance evaluation further confirmed the bottlenecks around suppliers, invoicing, and delivery. The results were coupled with resource utilisation, where the uneven workload was the main issue. After consulting with domain experts, more manual tasks unearthed, and their effects revealed yet again unequal workloads for process employees.

Table 5. Identified issues summary

Performance-related issues		**Customer experience related issues**	
1.	Delivery partner inefficiencies	1.	Not picked up packages in the shop
2.	Manual tasks that could be automated	2.	Low ETA/SLA fulfilment
3.	Uneven workload between resources	3.	Not enough transparency in process flow
		4.	Confusing order ETAs

6. Discussion

The thesis set out to find improvement opportunities for increased performance and better customer experience for the selected company. In order to identify issues, process mining techniques were used on the data log. This section of the paper takes the analysis results as inputs and applies redesign heuristics and principles to provide ideas on how to improve the as-is process. For customer experience related changes, the assumptions are that consumers retention, satisfaction and possible revenue is higher when processes are more customer-centric and transparent [28].

The discussion is divided into two sections based on the research questions, where possible solutions to analysis issues are presented. Later on, the presentation of workshop results and company feedback is presented to validate the redesign ideas, and limitations of the thesis are discussed.

6.1. Redesigns to improve performance

The first research question wanted to find "What improvement opportunities can be identified with process mining to enhance process performance?". It focused on finding issues and opportunities in the process for performance-related redesigns. There were a total of 3 issues detected from the analysis for which a variety of redesigns were found. The primary performance-related issues were: Delivery partner inefficiencies, Manual tasks in need of automation, and Uneven workload between sales process employees and for physical store's employees.

The first issue under discussion is the delivery partner inefficiencies. Currently, the order of shipping methods presented to the consumers is random, but this should be reorganised based on which delivery partner is the fastest in delivering. So, in the shopping cart, the company could implement a change in the UI on their website so that customers can see the fastest options first. By implementing this small change in their user interface, the company can decrease cycle time and, as a bonus, also increase customer satisfaction. The impacts on cycle time could reach even 30 hours per case if consumers opted for in-shop delivery instead of DPD parcel terminal pick-up. Even if the shop delivery is left out of the equation, the delivery times of outside delivery partners have a drastic difference in cycle times. The fastest (Smartpost) is able to deliver packages 14 hours quicker than the next best company (Omniva). Thus, average cycle time and also delivery speed to customers could have a significant impact.

In this problem, the redesign was based on the re-sequencing heuristic (H4). Re-sequencing generally applies to the rearrangement of activities within a process to increase performance. In this case, the re-sequencing of UI changes the customer's process, and cycle time is decreased due to it.

The second problem area is focused on manual tasks within the whole process. There were three major tasks that are today done manually but could be at least semi-automatic. These tasks are bank transfer payment checking, order dividing, and invoice making. Unfortunately, the timestamps in the data log did not capture the duration of these activities, so the precise time saving could not be analysed. Instead, the domain experts knowledge was consulted to learn more about these tasks and find suitable solutions.

In the as-is process, the accounting team checks the bank transfer payments, and then the information is sent via email to the sales team, who completes the status change and payment confirmation for an order. Currently, around 55 orders per week need to be confirmed this way, and in total, almost 4% of orders are paid through transfers. Due to the information transportation waste and time spent on manual confirming, the task should be redesigned. If possible, the payment confirmation should be automatic through connecting to bank API. If automation is not possible, the task completion should be transferred to the accounting team. It would avoid information transportation and time spent on sending individual emails. Also, since it is the accounting team who at the end needs this information, it should be in their interest to complete this task. The redesign could be achieved by granting the accountants access to the ordering system so that they are able to change order statuses.

For this tasks redesign, multiple heuristics and principles were relevant. H1 (task elimination) and H2 (task composition) both support the cut down of information transportation and reduction of manual checks. These heuristics are also in line with P1 (capture info once), which emphasizes sharing information via platforms, not sent individually and P3, according to which the processes should be performed by those who need the output.

Secondly, another manual task in the current system is dividing orders to sales specialists. One specialist is explicitly in charge of dividing all incoming orders among others, based on whose suppliers are present in the order. On average, around 1440 orders per week need to be manually allocated. Due to the volume of orders, this task can take more than an hour a day. The task could be replaced by automatic dividing based on business rules. For example, orders containing one persons' suppliers will be assigned to that person. As 67% of cases only contain one item, then even getting automatics for these types of orders would save a lot of time for the employees. The decrease in workload has a waterfall effect on their workday as well. Specialists can focus on other tasks since their time is freed from this activity. It means that invoicing can be done faster since orders are automatically divided, so the specialist can start making invoices first thing in the morning. Additionally, since invoicing speed will increase, the order cycle time will decrease due to expedited time to packing and shipping.

Order dividing automation is supported by H1, as the dividing of tasks does not provide any immediate value to the customer nor the business, it should be removed.

The final manual task is the making of invoices. Currently, invoices can be made when all items in the order arrive in the warehouse. Thus, based on stock levels, the system could be

taught to make the invoice once the inventory level of all items is 1. On faster days, the sales team makes over 800-900 invoices per day, which take a good chunk of time from their whole workday, so automation would save time. It will also shorten the cycle time since invoicing would be done as soon as the items arrive to the warehouse, not when the specialist gets to making the invoice. Additionally, domain experts expressed that due to the workload of invoicing, sometimes the warehouse is able to pack faster than specialists can make invoices. When implementing this redesign, the warehouse would have a steady flow of work that would not depend on the salespeople's actions.

All three redesigns for manual tasks are additionally based on H9 (Automation) by creating automatic systems based on business rules. In the automation of bank transfers, checking and confirming payments would be replaced by an API connection, allowing to save 1 hour of work per day for accounting and sales teams. It would also mean faster cycle times since currently, payment confirmation for bank transfers takes around 1.28 days. Secondly, by automating order dividing, the sales team would have less work every day, ranging from 15minutes on slow days and a few hours on busier days (based on domain expert knowledge). Lastly, automating invoice making by incorporating business rules into the invoicing system could save up to 5 hours of work per day combined (one invoice takes around 20 seconds to make) and increase delivery speed to customers by getting packages to packing faster.

To tackle the third issue of uneven case frequency and workload among sales specialists who handle the web orders, redistribution should be made. Currently, the workload is connected to which suppliers are whose, and the division is not based on any data or statistics, more on a legacy. The analysis showed that some employees have more than twice the cases than others. To change it, the company should redistribute the suppliers considering the supplier popularity in web orders. Since the sales specialists' salary also depends on the order size, each supplier's average order value should be considered as well.

Moreover, the analysis uncovered uneven workload on physical stores, to the extent that some stores hare three times the load than others. Also, according to domain experts, the process is done manually, which is another reason for the change. In the as-is process, ordering from stores was manually done, but the choice of the store was predetermined according to a list set many years ago. To eliminate the manual aspect of the process, this ordering should be done automatically as current IS in the company are capable of this change. When implementing the automatics, it should be considered that products need to be ordered first from the closest store and then from wherever the product has been sitting the longest.

Implementing these changes improves the workload distribution between employees, which helps to avoid overworking and stress. Using underutilised resources could also improve the cycle times since cases from over-worked resources can be processed faster. Automatic ordering from stores provides faster cycle times since automated product ordering ensures that items get on the next transportation round, thus arriving to the customer more quickly.

Additionally, when ordering "older" products from stores, the company can keep the average shelf-life of products smaller and ensure a more even distribution of goods in the stores.

The redesigns for issue 2 are based on H7 (Resource optimisation), where the focus is on sharing workload among idle and overworked employees and H9 (automation). By letting the system order items from the stores based on business rules, sales specialists will have one less task to perform daily or even hourly.

6.2. Redesigns to improve customer experience

The second research question was, "What improvement opportunities can be identified with process mining to improve customer experience and satisfaction?" It aimed to find opportunities in the process which could be redesigned to result in better customer experience. As success in e-commerce relies heavily on consumers, their retention and thus satisfaction [28], it is crucial to streamline the business processes to be more customer-centric. From the analysis, there were four issues that could have improved customer satisfaction: Not picked up packages, Low SLA fulfilment, Lack of transparency in the process, and Confusing order ETAs.

The first improvement opportunity is regarding the orders that do not get picked up from the stores. Based on the analysis, people usually pick up their orders from the store within 1.65 days after it has arrived, but there were also cases where clients forgot to come at all. An email notification is sent out when the package arrives, but the as-is process does not send any follow-up emails. The results showed at least 77 packages, where customers did not come, and which would eventually be refunded. According to domain experts, the shops are unable to keep track of the orders, thus no overview of cases is available. To alleviate the inventory issue as well as get customers to pick up their orders, an automated notification system should be built. For example, if the order has not been picked up in 2 weeks, the system should send another notification to the customer. This would help customers receive their orders which they have forgotten, thus increasing satisfaction. By lessening the amount of packages waiting in the shop, the employees can also more easily locate order which people come to pick up, providing a better customer experience.

It is supported by H8 (communication optimisation) and H9, where communication with the customer is optimised and automated via email notifications based on the order status and other available information.

The second issue that tarnishes customer experience is orders arriving past their promised ETA. In 2020, over 26% of orders were delivered late, meaning that over 19.5 thousand packages were delivered to disappointed consumers. When looking at the bottlenecks in the results, it showed that waiting to make the invoice, since suppliers' merchandise had not arrived, had the most considerable delay in the process. When consulting domain experts, they gave insight that the suppliers themselves are not at fault, and it is usually the sales specialists who order items

too late, making the clients' orders late too. To tackle the issue, the company should change their system for ordering goods from suppliers. If possible, ordering should be automized via API solutions since results indicated that automatic supplier orders have a 10% higher chance of arriving within the SLA. Where not possible, the notification system for sales specialists needs to improve, so that products are ordered on time. The results provide a comprehensive list of which suppliers need special attention in regards to the issue so that the company can prioritise which partners should be addressed. When implementing the change, customer experience will be improved since orders are more likely to arrive on time.

The redesign is based on H9 since automation implementation is recommended to improve the as-is process. Supplier ordering should be made automatic via XML or API solutions where possible. This has been implemented for several suppliers already, but the majority are still done manually right now. By implementing this change, customer orders would be 10% more likely to arrive on time.

Thirdly, the issue regarding the lack of transparency in the order statuses that customers can see needs to be redesigned. On average, the customer waits for 4.5 days in the "Order in process" status, but domain experts gave light that consumers often cannot understand its meaning. As a result, customer support has to answer questions regarding order statuses daily. To improve customer understanding and increase transparency, the company should provide more insightful status updates. For example, instead of saying "order in process", they could specify that products are ordered and are on their way to the warehouse. From the analysis, it also became known that orders shipped in multiple batches also do not provide customers relevant information from order statuses. Thus, clients contact the firm when they receive one package and are still missing some items. This could be improved by conveying to the customer that the package will arrive in multiple instances via their order tracking page and status updates, and also by providing all tracking numbers available so that the clients have all relevant information about their orders.

Implementing these changes will increase customer experience through transparency. By providing more insightful information to clients, the process becomes more customer-centric. Additionally to increases in customer satisfaction, the changes could lessen the impact on customer support.

The last issue is regarding ETA confusion in customers. Currently, the ETA is shown to the customers on each products website and in the purchase confirmation letter. According to sales specialists' knowledge, clients often do not notice the ETA before purchasing and then cancel their order once they see it in the letter. They estimate that around 25% of refunds happen directly after payment confirmation and receiving the letter with an ETA. It also happens that customers purchase items with different delivery times but expect the order to arrive according to the shortest time. One redesign idea, in this case, would be to show the ETA in the shopping cart before order confirmation. This way, the customer can see the total delivery time for the whole order before it is confirmed and money is transferred.

The redesign would decrease cancellations, thus also workload on employees who have to handle the refunds. An estimate from sales specialists was that around 350 refunds per year happen due to this error, and one refund takes 5 minutes to handle, so that total time savings could be almost 30 hours per year. Additionally, customers are more satisfied since they are not disappointed after order confirmation when they find out the ETA is longer than they thought.

6.3. Workshop Results and Feedback on Redesign

During the workshop, analysis results were presented, and feedback was gathered from the company regarding relevance and implementation. For each redesign, four questions were presented to understand the company's understanding of these issues and their relevance to the redesign.

The company agreed that all presented issues are significant and need to be tackled, however, they did not think that all redesigns were worth implementing. All participants agreed that manual tasks need to be reduced, and automation is vital to be able to scale the business and the process. Below are the detailed feedbacks for all proposed redesigns and a summary table of the average numerical feedback to the changes (Table 6). The scale for relevance is 1 for not relevant and 5 for extremely relevant redesigns. For ease of implementation, the scale is similar, 1 means easy to implement, and 5 is hard to implement.

Table 6. Redesign relevance and implementation scores from the company (1 – not relevant/easy to implement, 5 – highly relevant/hard to implement)

Issue	Redesign	Relevance	Implementation ease
Delivery partner inefficiency	Reorganise shipping UI in shopping cart	5	1
Manual tasks	Automatic bank transfer checks	4	3
	Automatic order dividing	5	3
	Automatic invoicing	5	4
Uneven workload	Redistributing supplier responsibility	4	3
	Automatic ordering from shops	5	3
Not picked up packages	Automated follow-up notification	2	3
Low SLA fulfilment	New/Automated ordering system for supplier orders	5	2
Lack of transparency	Additional and more precise status updates	2	3
	Updates regarding divided parcel delivery	3	3
Confusing ETA	ETA in shopping cart	1	2

Reorganising shipping UI – The company knew that there were some issues with transportation companies, but they were not aware that the delivery times had such drastic differences. All participants were confident that the delivery methods were already ordered in the most beneficial way, but as the analysis showed, this turned out to be false. They agreed that the change on the website needs to be implemented as soon as possible, which is also illustrated by the relevancy score of five. Additionally, rearrange the UI does not require a lot of time and effort, so ease of implementation was also scored a five. Thus, this would be a quick and necessary improvement for the company.

Automated bank transfer matching - The CEO was not aware of how manual the task was; thus, the problem had never risen. The company was aware that bank transfer payments are problematic, but the severity of wastes in the process was not apparent. The developer had brought up automation before, but bank APIs are currently not compatible with the firm's IS. Due to lack of resources, this redesign could be implemented, but the focus would be on making it semi-automatic and shifting the workload on the accounting team as proposed. Due to the incompatible API's, the implementation of this task was scored a three since manual business rules will have to be made. The issue's relevance was graded as a four since it is crucial to eliminate manual tasks and workload, but the number of orders related to this issue was small (under 4%).

Automated order dividing – Again, the CEO was unaware that this activity was so manual and time-consuming because sales specialists never brought it up. The redesign was highly relevant, in their opinion, as they implemented a similar semi-automatic version just a few weeks before the workshop. The sales specialist noted that even the current semi-automatic version already saves hours of work each week. The presentation of results gave even more ideas to the developer on how to improve the system to make it fully automatic. This process was highly relevant and got scored a five since the redesign saves time and influences every webshop order. As orders come in different forms and types, this redesign would need several business rules put in place, thus ease of implementation was three.

Automated invoicing – Regarding invoicing, the CEO mentioned that the company has been talking about this change for a couple of years now since the developers of the system in use actually provide this functionality. So far, the redesign has not been implemented due to the peculiar effects it would have on accounting processes. However, after seeing the analysis results and brainstorming about possible implementations during the workshop, the company will try harder to find a way to implement this redesign soon. This improvement got a five for relevance because, again, the redesign influences each case, so the effects on workload and process performance would be huge. Due to the nature of the legacy IS in place at the company and specific accounting procedures that cannot be changed, the implementation of automatic invoicing was scored a two.

Redistribution of workload/supplier responsibility – The uneven workload distribution has also been known for quite some time for everyone, but the fast pace of work due to the

pandemic has not given a chance to carry out changes. The CEO also mentioned that this redesign idea needs to be put on hold until certain parts of the process are made automatic, as this could influence the workload on sales specialists already. For example, automating sales specialists daily tasks like product ordering and invoicing would alleviate some of the uneven workloads as the system would perform these tasks instead of a person. Thus, the company is interested to see how much the workload difference is after automating these steps and then continue to redistribution of who is responsible for which supplier. Since automation is more of a priority, and the redesign issues could be affected by other improvements, the relevance score for this opportunity was four. The implementation was a three since coming up with a 100% fair system would take a lot of analysis and time. Additionally, sales specialists would have to cross-pollinate their knowledge on suppliers to make the re-distribution happen, so time on training would have to be spent.

Automated ordering from shops – The issue was known prior to the research, and the company has already made plans to implement a similar solution to the proposed redesign. The awareness of the issue comes to light every day since employees are able to see the amount of items ordered from each shop every day after placing orders. What is more, the company needs to present reports on average item age (how old is the inventory), so it is in their interest to have systems that support decreasing average inventory age. As the manual aspect of the task should be eliminated, and the company wishes to decrease the average self-life of products, then the relevancy was rated high, a five. The redesign would need the interaction and communication of several IS, so the ease of implementation was rated a three.

Automated follow-up notifications – The interviewees see this change as an opportunity to improve, but they believe a fully automatic solution would not be the best at the moment. Since shop works still make mistakes marking the orders as “delivered”, there needs to be a checking point before the automated notifications. The firm believes this redesign is something to think about, but it is not a priority at the moment due to the scale of the impact. The low priority is illustrated by a relevance score of two, although the firm agreed that the issue itself was a relevant problem. Again, implementation is of medium ease (3) since different IS need to be integrated.

New ordering system for supplier orders – All participants were shocked to hear that the number of orders delivered on time was only around 64%, so the severity of the issue was not evident previously. From the discussion, the sales specialist and CEO also agreed that most of the delays come from the sales teams work, not due to supplier shipping issues. The CEO mentioned that the number of on-time deliveries needs to increase and measures should definitely be taken. The sales specialist also hinted that some new automated ordering systems were coming for certain suppliers. The company plans to take into consideration the supplier analysis to prioritise changes in the system. Since the firm wishes to please customers, the relevancy of increasing SLA fulfilment was scored a five. The ease of implementation from the company side is four since the system already possesses automation possibilities.

More transparent statuses – The company was aware of the issue, and they have tried to solve it in the past, but no better option has come to light so far. This redesign's underlying issue is definitely a problem, but the improvement idea will not be implemented. The CEO feels that changing the statuses to be more transparent would harm the company and reveal to the customer parts of the process that they want to keep secret. They would love to implement some change in the future when they can come up with a solution that is both customer-centric and company-friendly. Since the solution could be detrimental for the company, then the relevancy was a two. However, when the company can develop a firm-friendly option, then implementing will not be too hard (score of three) since prior statuses are already linked between systems in a similar way.

Updates and tracking for several packages – The issues were known for some time since the company has had issues with DPD systems for a while. Although this redesign does not influence that many orders, it would be a nice-to-have feature on the website. Since the company migrated to a different shipping-related IS, the change would be easier to implement than before, so it is somewhat likely that a similar change will be carried out in the near future, according to the developer. This is considered a relevant idea, but the number of impacted cases is low, so the score from the company was three. The ease of implementation was three, as a new delivery-related system has already been put into place, so integration with the customer portal would be doable and not too hard.

ETA in shopping cart – The CEO agrees that they could always be more customer-centric and accurate, but at the moment, the company's interest weigh over it. He also believes that, in large part, this is an issue due to the consumers' lack of paying attention when shopping. The particular redesign of providing the ETA in the shopping cart will most probably never be done. The CEO continued saying that currently, almost no e-shop has this functionality and that it is for a reason. Providing the ETA in the shopping cart would likely decrease sales, especially the people who would otherwise make a purchase and then cope with the long delivery time. In the company's opinion, the change would solve one problem for the customer but create a new one for the firm. Similarly to the status updates, this redesign is believed to harm the company, so the relevance was scored a one. That being said, this is something that could easily be implemented if the need would arise, since ease of implementation was four.

Considering the feedback and the scores, it can be said that the research proved relevant and impactful to the company. Process mining techniques were able to find significant improvement opportunities, which is especially demonstrated by the amount redesigns the company wishes to implement as soon as possible. In addition, the research proved helpful, particularly due to the data-driven nature of the analysis since the firm did not have any visualisation of process analytics prior to this thesis and relied on confidence-based decision-making.

In order to illustrate the relevance of each redesign and provide a guide for implementing the changes, an action priority matrix was made (Figure 35). The matrix plotted each redesign

opportunity based on the survey scores to identify which changes should be tackled first. This chart also showcases once more how most redesigns and identified issues were highly relevant to the company. According to the matrix, the company should first tackle the highest relevance issues, so redesigns 1, 8, 6, 3, and 4. Based on the availability of resources in the company, further prioritisation regarding effort should be made. After implementing these, opportunities 5 and 2 should be tackled next, followed by redesign 10, as these are the following highest relevance areas. Redesigns 9 and 7 should not be implemented since they are not relevant and would require more than average efforts. Opportunity 11 could be tackled at some point, but the low relevancy makes this redesign insignificant as well.

Figure 35. Action Priority Matrix of Redesign Ideas

6.4. Feedback on Apromore

The research results could not have been achieved without Apromore templates and the Apromore portal for process mining. As mentioned in the methods already, the templates were extraordinarily straightforward and easy to use. Each template had an excellent summary followed by a step-by-step guide to the actual analysis. One critique about them would be that the four templates repeated each other a little. For example, if the aim is to find the root cause of issues, then some type of variant analysis is already done during the other three templates.

Secondly, the Apromore portal for process mining was also a great tool to use for this analysis. The system offered a large variety of filtering and visualisation opportunities, complete with even a dashboard analysis functionality. It was effortless to manoeuvre inside the program and create a variety of views and log variants. On the other hand, when it came to exporting the

data, the portal could be improved. After applying some filtering, there was a need to export the data back to a CSV file again, but for some reason, Apromore changed the data format. The biggest issue was that when uploading the data, it needs to have start and end timestamps in different columns, but the exported format had batched them into the same column. Other than the minor issues mentioned above, Apromore was an excellent tool for process mining analysis.

6.5. Limitations

The aim of the thesis was to analyse what improvement opportunities can be found with process mining techniques in order to redesign a company's operational processes. As described in the methods, the research used a mixed-methods single case study methodology, where quantitative process mining analysis was combined with a qualitative workshop to recommend redesigns to the firm at hand. Nonetheless, the research possesses several limitations, which will be discussed from three aspects: internal validity, construct validity, and external validity.

Internal validity is concerned with how realistic are the causal relationships found in the research case [26]. Since the thesis is a descriptive case study based on a single company, it is acknowledged that the theoretical offerings of the research are limited, and no theory building was attempted. As the results reflect the processes and opinions of one company, then the threat remains.

Secondly, there are construct limitations to the case study methodology. Construct validity is used to understand if the described case represents reality [32]. This threat to validity can be decreased by employing triangulation – using several sources for data collection [32]. The data for this study was collected from various sources, including data logs, domain experts, company websites, a workshop, and a small survey. The information was also gathered from several people with different backgrounds in the company in order to represent a variety of opinions and alleviate the threat even further. The workshop was used to validate the realisability of the ideas for the company. However, this data collection method increased the risk of misunderstandings, but the semi-structured nature and discussion of specific details allowed to lessen the threat. In order to reduce the probability of invalidity even further, a simple survey was used to confirm the understandings of the workshop. This was illustrated by the survey answers matching with the discussion of the workshop. All in all, the study results were highly relevant to the real-life situation the firm was in, as some of the results were already implemented after the presentation of the research.

Lastly, external validity is concerned with how generalisable the results are [26]. Since the thesis focuses on a specific e-commerce related process, the study is not expandable outside this industry. While the research could be applied to other commerce cases, it is not proposed that the study is relevant for all e-commerce related processes, as each firm has company-specific processes in place. That being said, there might be parts of the experience that could be transferrable, but given the context of the case, further research should be conducted to confirm generalisability.

7. Conclusion

The competitive nature of business, and e-commerce especially, has pushed companies to improve their processes to increase efficiency. Due to COVID-19, the rise in demand has increased the need for change even more. To realise these changes, business process management tools like process mining can be applied for a data-driven and insightful approach.

Current research in e-commerce process mining has focused on customer-side processes, not business operational ones, thus the thesis contributes a new perspective to the industry's research. This study looked at the operational process of sales handling in e-commerce. Particularly, the research wanted to see what process performance improvements could be identified with process mining (RQ1) and which redesigns to improve customer satisfaction could be found (RQ2). To answer the research questions, a data log was extracted from the company's information systems and analysis was conducted via PM^2 process mining methodology and Apromore process mining guides. The results were combined with redesign heuristics to propose improvement ideas for transforming the as-is process. During the process, company representatives with domain knowledge were consulted to ensure highly insightful and relevant results.

The analysis was able to identify seven issues in the current process with process mining tools, and in total, eleven redesign opportunities were presented to alleviate the problems. The analysis was able to capture process bottlenecks, manual tasks and their impact on process performance, thus redesigns of automation and reorganisation were proposed to decrease process cycle time. The analysis also showcased the lack of transparency to consumers and the company's disability to fulfil their promised delivery time, thus several redesigns on making the process more customer-centric were proposed to improve order delivery and consumer satisfaction. In order to further validate the results and redesign concepts, company representatives were consulted in a workshop format, where feedback was given in terms of relevance, prior knowledge of the issue and plans on implementing suggested changes. Based on the company's feedback during the workshop, the identified issues proved highly relevant. Although they acknowledge that the problems were relevant, some of the proposed redesigns were not in sync with the firm's strategic goals and operating principles, thus not all proposed changes will be implemented in the future.

In addition to providing the researched firm with a guide on process redesigns and improvements to implement, the research also provided evidence of using process mining for identifying improvement opportunities in an e-commerce company's operational process. However, the thesis was based on only one company's event log and is specific to the industry, so relevancy for other companies should be proved with further similar studies. To expand research into e-commerce even more, process mining of operational processes should be combined with the mining of customer weblogs, allowing to analyse the interactions the company has with customers and its effects on both consumers and the firm's internal processes.

References

[1] J. Heikkilä, 'From supply to demand chain management: efficiency and customer satisfaction', *J. Oper. Manag.*, vol. 20, no. 6, pp. 747–767, 2002, doi: 10.1016/S0272-6963(02)00038-4.

[2] D. Dakic, D. Stefanovic, I. Cosic, T. Lolic, and M. Medojevic, 'Business Process Mining Application: A Literature Review', in *DAAAM Proceedings*, 1st ed., vol. 1, B. Katalinic, Ed. DAAAM International Vienna, 2018, pp. 0866–0875. doi: 10.2507/29th.daaam.proceedings.125.

[3] W. M. P. van der Aalst, 'Business Process Management: A Comprehensive Survey', *ISRN Softw. Eng.*, vol. 2013, pp. 1–37, 2013, doi: 10.1155/2013/507984.

[4] S. Smirnov, H. A. Reijers, M. Weske, and T. Nugteren, 'Business process model abstraction: a definition, catalog, and survey', *Distrib. Parallel Databases*, vol. 30, no. 1, pp. 63–99, Feb. 2012, doi: 10.1007/s10619-011-7088-5.

[5] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*, 2nd ed. 2018. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer, 2018. doi: 10.1007/978-3-662-56509-4.

[6] N. Poggi, V. Muthusamy, D. Carrera, and R. Khalaf, 'Business Process Mining from E-Commerce Web Logs', in *Business Process Management*, vol. 8094, F. Daniel, J. Wang, and B. Weber, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 65–80. doi: 10.1007/978-3-642-40176-3_7.

[7] R. Kohavi, 'Mining E-Commerce Data: The Good, the Bad, and the Ugly', p. 6.

[8] J. Koetsier, 'Covid-19 Accelerated E-commerce growth "4 to 6 years"', *Forbes*, Jun. 12, 2020. Accessed: Nov. 20, 2020. [Online]. Available: https://www.forbes.com/sites/johnkoetsier/2020/06/12/covid-19-accelerated-e-commerce-growth-4-to-6-years/?sh=ee29b3e600fa

[9] W. van der Aalst *et al.*, 'Process Mining Manifesto', in *Business Process Management Workshops*, vol. 99, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 169–194. doi: 10.1007/978-3-642-28108-2_19.

[10] M. Weske, *Business process management: concepts, languages, architectures*. Berlin ; New York: Springer, 2007.

[11] W. M. P. van der Aalst *et al.*, 'Business process mining: An industrial application', *Inf. Syst.*, vol. 32, no. 5, pp. 713–732, Jul. 2007, doi: 10.1016/j.is.2006.05.003.

[12] Á. Rebuge and D. R. Ferreira, 'Business process analysis in healthcare environments: A methodology based on process mining', *Inf. Syst.*, vol. 37, no. 2, pp. 99–116, Apr. 2012, doi: 10.1016/j.is.2011.01.003.

[13] Y. Jiang and S. Yu, ‘Mining E-Commerce Data to Analyze the Target Customer Behavior’, in *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)*, Adelaide, Australia, Jan. 2008, pp. 406–409. doi: 10.1109/WKDD.2008.90.

[14] W. M. P. van der Aalst, ‘Business Process Management Demystified: A Tutorial on Models, Systems and Standards for Workflow Management’, in *Lectures on Concurrency and Petri Nets*, vol. 3098, J. Desel, W. Reisig, and G. Rozenberg, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–65. doi: 10.1007/978-3-540-27755-2_1.

[15] F. Milani and F. M. Maggi, ‘A Comparative Evaluation of Log-Based Process Performance Analysis Techniques’, *ArXiv180403965 Cs*, Apr. 2018, Accessed: Jan. 07, 2021. [Online]. Available: http://arxiv.org/abs/1804.03965

[16] Apromore Pty Ltd, ‘Apromore’, 2020. https://apromore.org (accessed Nov. 19, 2020).

[17] A. Terragni and M. Hassani, ‘Analyzing Customer Journey with Process Mining: From Discovery to Recommendations’, in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, Barcelona, Spain, Aug. 2018, pp. 224–229. doi: 10.1109/FiCloud.2018.00040.

[18] J. Brunk, C. Methner, A.-K. Cordes, S. Thiermann, and T. Goerke, ‘Trace Clustering for an Online Freight Exchange at TIMOCOM’, p. 64.

[19] M. Thiede, D. Fuerstenau, and A. P. Bezerra Barquet, ‘How is process mining technology used by organizations? A systematic literature review of empirical studies’, *Bus. Process Manag. J.*, vol. 24, no. 4, pp. 900–922, Jul. 2018, doi: 10.1108/BPMJ-06-2017-0148.

[20] B. T. Greyling and W. Jooste, ‘THE APPLICATION OF BUSINESS PROCESS MINING TO IMPROVING A PHYSICAL ASSET MANAGEMENT PROCESS: A CASE STUDY’, *South Afr. J. Ind. Eng.*, vol. 28, no. 2, Sep. 2017, doi: 10.7166/28-2-1691.

[21] N. R. T. P. van Beest and L. Maruster, ‘A Process Mining Approach to Redesign Business Processes - A Case Study in Gas Industry’, in *Ninth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2007)*, Timisoara, Romania, Sep. 2007, pp. 541–548. doi: 10.1109/SYNASC.2007.50.

[22] M. S. Uysal *et al.*, ‘Process Mining for Production Processes in the Automotive Industry’, p. 8.

[23] M. Jans, J. M. van der Werf, N. Lybaert, and K. Vanhoof, ‘A business process mining application for internal transaction fraud mitigation’, *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13351–13359, Sep. 2011, doi: 10.1016/j.eswa.2011.04.159.

[24] M. Jans, M. Alles, and M. Vasarhelyi, ‘The case for process mining in auditing: Sources of value added and areas of application’, *Int. J. Account. Inf. Syst.*, vol. 14, no. 1, pp. 1–20, Mar. 2013, doi: 10.1016/j.accinf.2012.06.015.

[25] J. De Weerdt, A. Schupp, A. Vanderloock, and B. Baesens, 'Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization', *Comput. Ind.*, vol. 64, no. 1, pp. 57–67, Jan. 2013, doi: 10.1016/j.compind.2012.09.010.

[26] M. N. K. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students*, Seventh edition. New York: Pearson Education, 2015.

[27] Campbell, Patrcik, 'Customer Acquisition VS. Retention: Where are Your Dollars best Spent?', *ProfitWell*, May 21, 2020. https://www.profitwell.com/recur/all/customer-acquisition-vs-retention (accessed May 01, 2021).

[28] T. W. Jackson, 'Personalisation and CRM', *J. Database Mark. Cust. Strategy Manag.*, vol. 15, no. 1, pp. 24–36, Oct. 2007, doi: 10.1057/palgrave.dbm.3250065.

[29] 'The impact of Covid-19 on e-commerce', *Ecommerce News Europe*, Jan. 15, 2021. https://ecommercenews.eu/the-impact-of-covid-19-on-ecommerce/ (accessed May 01, 2021).

[30] M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van der Aalst, 'PM^{2}: A Process Mining Project Methodology', in *Advanced Information Systems Engineering*, vol. 9097, J. Zdravkovic, M. Kirikova, and P. Johannesson, Eds. Cham: Springer International Publishing, 2015, pp. 297–313. doi: 10.1007/978-3-319-19069-3_19.

[31] 'Action Priority Matrix', *ProductPlan*. https://www.productplan.com/glossary/action-priority-matrix/ (accessed May 10, 2021).

[32] R. Yin K., *Case study research: design and methods, 4th edn*. SAGE Publications, Thousand Oaks, CA, USA, 2009.

Appendix

I. Additional Data Analysis

Figure 1. Supplier analysis comparison in terms of SLA fulfilment

Supplier	SLA = 0 %	SLA = 1%	Supplier	SLA = 0 %	SLA = 1%	Supplier	SLA = 0 %	SLA = 1%
10	33,30%	57,24%	34	35,44%	54,95%	109	14,81%	66,67%
14	24,66%	64,24%	115	15,65%	82,31%	112	50,00%	26,92%
17	32,65%	57,09%	81	25,27%	65,84%	30	41,67%	41,67%
26	17,72%	70,44%	50	11,61%	78,17%	92	0,00%	75,00%
55	18,93%	66,63%	39	17,41%	64,78%	79	63,16%	10,53%
64	9,89%	83,61%	118	33,47%	58,78%	62	44,44%	22,22%
76	27,31%	58,95%	70	13,00%	71,00%	58	7,14%	57,14%
23	13,35%	80,82%	24	33,50%	57,36%	96	38,46%	23,08%
85	10,83%	78,10%	111	24,86%	69,73%	116	61,54%	30,77%
89	18,08%	75,21%	77	24,59%	67,21%	48	70,00%	20,00%
86	12,16%	82,01%	45	22,35%	60,34%	59	20,00%	60,00%
88	20,20%	68,89%	20	32,12%	35,97%	66	33,33%	66,67%
5	23,44%	62,06%	95	21,02%	46,50%	84	20,00%	60,00%
90	13,97%	73,15%	105	25,66%	63,16%	16	50,00%	50,00%
7	43,84%	46,16%	110	27,08%	62,50%	33	75,00%	0,00%
57	29,16%	57,88%	107	4,29%	84,29%	83	33,33%	66,67%
65	27,47%	66,11%	11	18,27%	64,42%	97	66,67%	0,00%
6	24,33%	62,06%	87	30,77%	54,81%	47	50,00%	50,00%
69	18,79%	71,29%	98	15,69%	70,59%	114	100,00%	0,00%
52	29,70%	44,73%	29	15,15%	60,61%			
44	25,12%	59,38%	67	20,62%	76,29%			
8	33,68%	57,03%	36	32,63%	51,58%			
99	58,79%	30,31%	35	21,74%	54,35%			
18	14,50%	75,84%	91	20,65%	47,83%			
104	12,94%	79,16%	113	27,06%	65,88%			
28	17,51%	71,35%	32	31,88%	57,97%			
22	25,22%	65,78%	27	25,76%	54,55%			
101	35,85%	50,81%	73	36,21%	44,83%			
71	23,49%	62,97%	93	5,17%	74,14%			
94	21,49%	68,28%	117	22,41%	67,24%			
56	16,64%	73,75%	21	41,07%	44,64%			
38	12,80%	78,11%	82	23,91%	60,87%			
15	16,38%	75,81%	106	34,78%	58,70%			
102	16,17%	76,65%	40	8,33%	66,67%			
78	18,58%	67,22%	75	19,44%	36,11%			
53	4,53%	71,12%	72	36,36%	51,52%			
13	27,23%	55,38%	103	32,25%	40,63%			
37	32,02%	60,56%	108	25,81%	67,74%			
63	21,00%	70,17%	80	28,57%	57,14%			

Figure 2. Delivery method comparison in terms of delivery speed

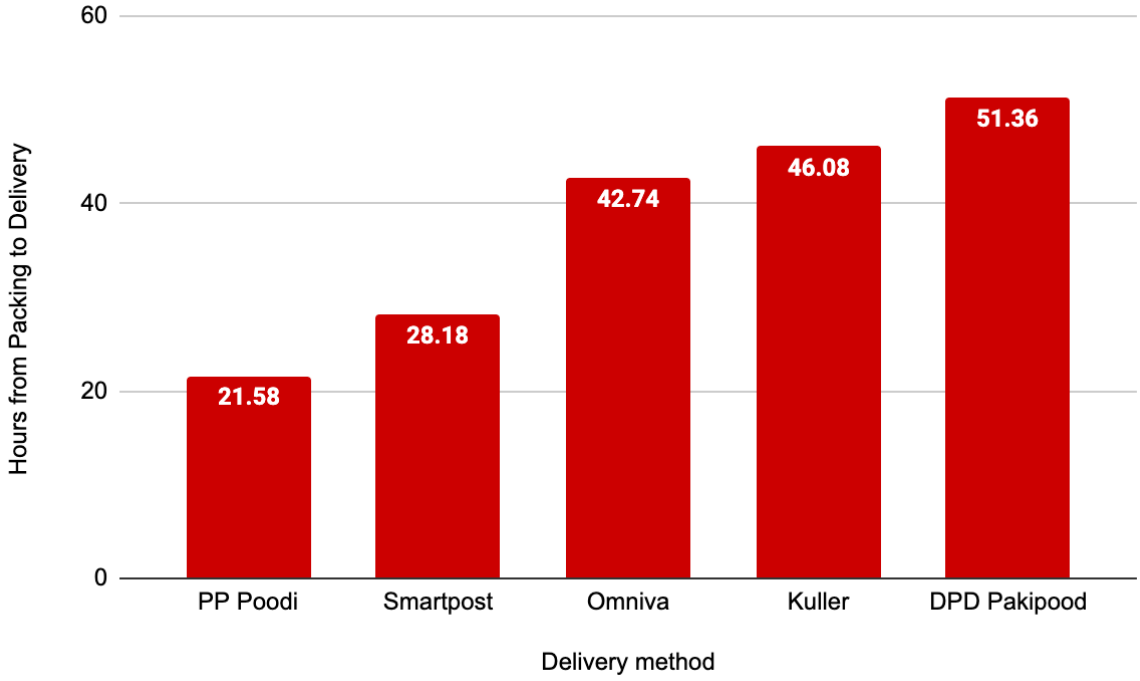


Figure 3. Correlation analysis between dataset variables

	DTFulfilled	supplier	resource	numberOfItems	shippingMethod	campaign	totalValue	paymentMethod	supplierCountry
DTFulfilled	1								
supplier	0,082	1							
resource	0,035	0,279	1						
numberOfItems	0,056	0,027	0,116	1					
shippingMethod	0,223	-0,024	0,035	0,094	1				
campaign	-0,084	0,015	0,011	0,119	-0,055	1			
totalValue	0,079	-0,104	-0,006	0,093	0,103	0,002	1		
paymentMethod	0,088	-0,044	0,033	0,026	0,089	-0,048	0,261	1	
supplierCountry	0,062	0,443	0,247	0,043	-0,058	-0,007	-0,094	-0,047	1

Figure 4. Relative cases over time, to depict workload division and case frequency

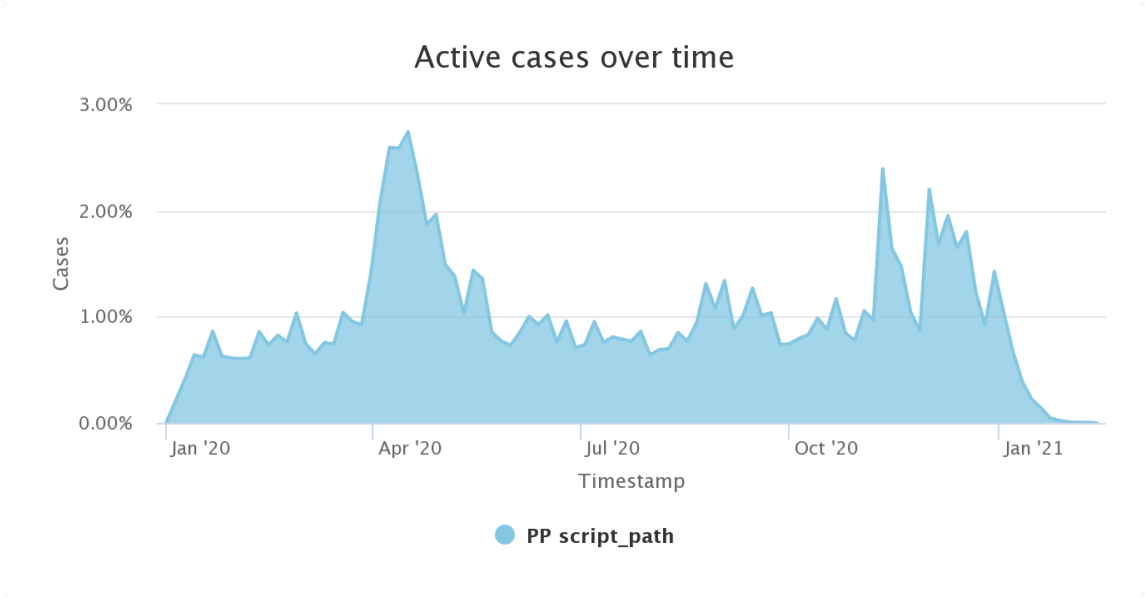


Figure 5. Case duration distribution, where green is up to median case duration of 6.52 days

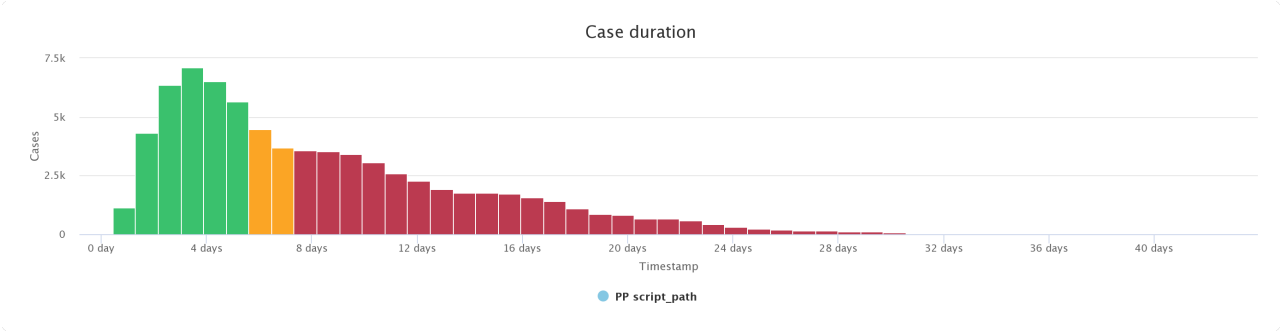


Figure 6. Number of cases bought as private person and as a company

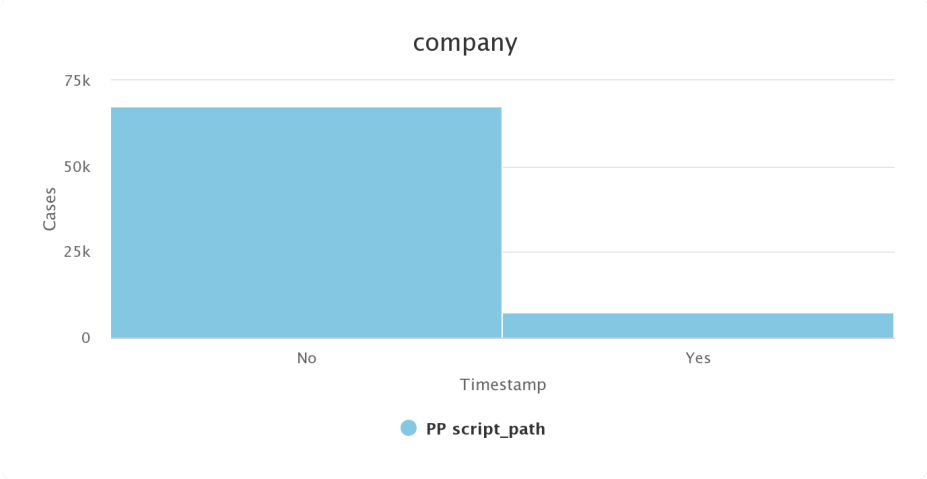


Figure 7. Number of orders made during campaigns and non-campaign times

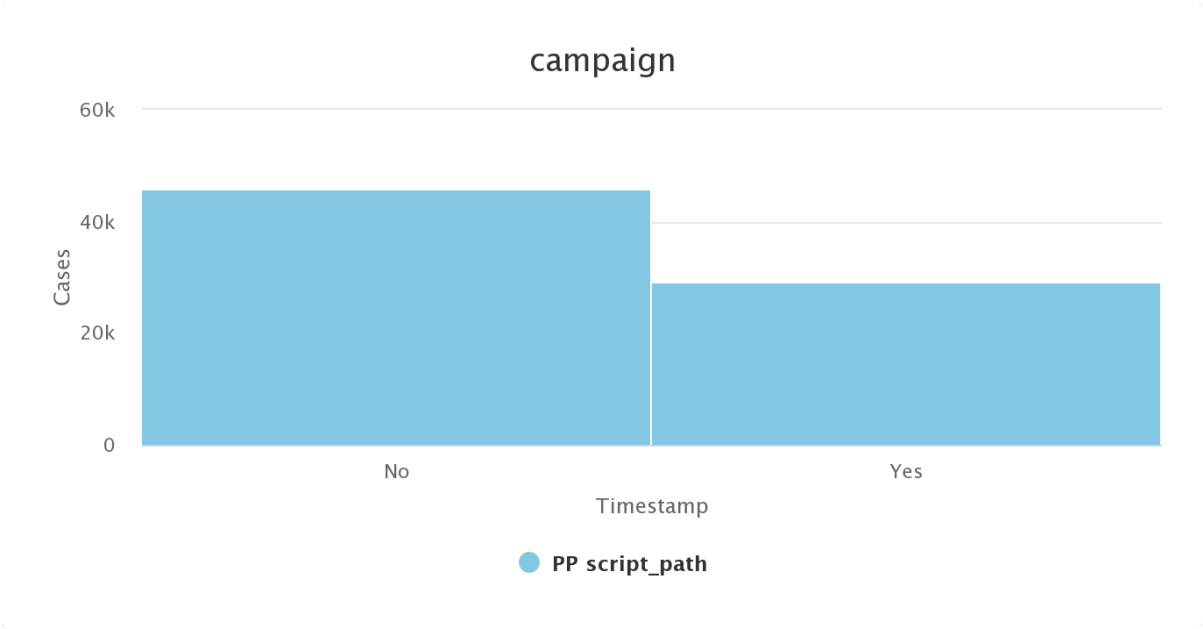
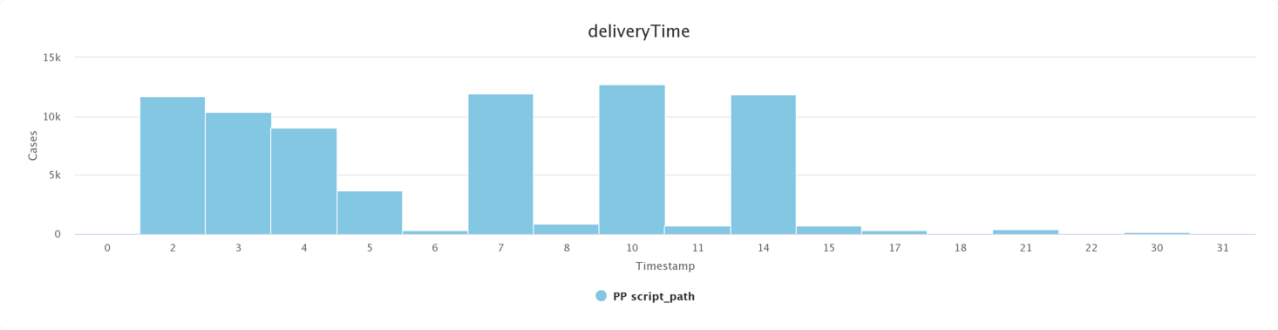


Figure 8. Distribution and frequency of cases for each promised ETA.



II. License

Non-exclusive licence to reproduce thesis and make thesis public.

I, Maria Häling,
(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Improving operational processes - A process mining case study in e-commerce
   (title of thesis)

   supervised by Fredrik Milani, PhD.
   (supervisor's name)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Maria Häling

14/05/2021