This work was submitted to:

Chair of Process and Data Science (PADS - Informatik 9), RWTH Aachen University

A framework to correlate resource behavior with the success of process instances – A case study in a P2P process

Master's Thesis

Author: **Syed Faizan Hassan**

Student ID: **403355**

Supervisor: Alessandro Berti

Examiners: Prof. Dr. Wil Van Der Aalst
Prof. Dr. Ulrik Schroeder

Registration Date: 2022-12-02

Submission Date: 2023-07-14

Abstract

This thesis focuses on improving the efficiency of `manual documents` processing by identifying the `optimal behavior` for resources to optimize the `performance` of the process. Manual processes, which involve tasks requiring human intervention, often suffer from delays, errors, and inefficiencies due to their time-consuming nature. Optimizing these processes is critical for the effective functioning of organizations. Resource working behavior plays a vital role in manual processes, as even small improvements can have a significant impact on the overall performance of the process. The factors influencing resource working behavior in manual processes are examined, including `workload`, `working prioritization pattern`, and `batching`. Finding the optimal workload is essential to ensuring resources are appropriately engaged and challenged. The working pattern refers to how resources organize and prioritize tasks, and adaptability is key for process performance optimization. Batching can improve efficiency but may also lead to monotony or reduced concentration. Various techniques are utilized to identify these factors, such as the `performance spectrum` and the `dot chart method`. However, these techniques primarily analyze single-resource behavior and may lack a comprehensive overview of all process aspects. They may also present challenges in interpretation, especially for non-experts. Additionally, existing techniques often provide visual representations without concrete numerical values or metrics. To address these limitations, this thesis proposes evaluating resource behavior against *key performance indicators (KPIs)* to identify the optimal working behavior for the resources. The analysis aims to provide a `synthetic representation` of results that is simpler to comprehend and interpret than existing graphical outputs. By identifying the optimal working behavior, the overall efficiency of manual document processes can be enhanced, assisting organizations in achieving their goals and improving process performance.

Contents

List of Figures

Chapter 1

Introduction

In today's dynamic business landscape, efficient and effective processes are crucial for organizations to stay competitive and meet customer expectations. However, many organizations still rely on manual processes, which involve tasks requiring human intervention, such as `document review` and `approval`, `customer service`, and other critical activities. These manual processes often suffer from delays, errors, and inefficiencies, hindering the organization's ability to deliver optimal results. The `working behavior` of resources involved in these manual processes plays a vital role in determining process efficiency and performance. Understanding and optimizing resource behavior can provide valuable insights into workload allocation, task execution, and overall process optimization. By identifying areas for improvement and implementing strategies to enhance resource behavior, organizations can streamline manual processes, reduce delays, and improve overall performance.

This thesis focuses on the analysis and optimization of resource behavior in manual processes to enhance process efficiency and performance. The goal is to develop a systematic approach that enables organizations to gain insights into resource behavior, make informed decisions regarding resource allocation and workload assignment, and optimize task execution to maximize process efficiency. The research methodology involves analyzing historical process data, including *key performance indicators (KPIs)*, to evaluate resource behavior and its impact on process performance. Through this analysis, patterns and correlations between resource behavior and process outcomes will be identified, allowing for the development of strategies to optimize resource utilization. To evaluate the proposed approach, a comprehensive evaluation will be conducted using real-world process data. A dedicated tool will be developed, which takes an `event log` and a selected `KPI` from the *list of KPIs* as input. The framework will then be applied to the event log, generating `synthetic tables` that provide valuable insights into resource behavior and its impact on process efficiency. These tables will include information on the `working behavior` of individual resources, `evaluated working behavior metrics`, and `success rates` of the resources. Additionally, there will be a table dedicated to `optimal workload`, offering recommendations for the ideal workload range for each resource. Furthermore, the correlation values between working behavior and process success will be determined and analyzed. These correlations will provide valuable insights into how resource behavior relates to overall process success. By understanding these relationships, organizations can identify areas for improvement and make data-driven decisions to enhance process performance. The results of the evaluation will demonstrate the practical value and applicability of the proposed approach in improving manual process efficiency. Through this research, organizations will gain the necessary tools and knowledge to leverage resource behavior analysis for achieving process excellence and gaining a competitive edge in today's business landscape. The findings and insights derived from this research will empower organizations to optimize their manual processes, reduce errors and delays, and enhance overall operational efficiency. By bridging the gap in understanding and optimizing resource behavior, this thesis aims to contribute to the field of process improvement and optimization.

1.1 Motivation

In today's business environment,organizations must prioritize the implementation of efficient and effective processes to maintain their competitive edge and meet the ever-growing demands of customers. While manual processes, which rely on human intervention, are a critical component of many organizations, they often suffer from inherent challenges such as delays, errors, and inefficiencies that can significantly impact overall performance. To overcome these obstacles, it becomes imperative to delve into the working behavior of resources involved in these manual processes, unlocking valuable insights and opportunities for improvement. By understanding and optimizing resource behavior, organizations can achieve enhanced process efficiency, cost reduction, and an elevated level of customer satisfaction, all of which contribute to their long-term success.

1.2 Problem Statement

Despite the importance of resource behavior in manual processes, there is a lack of effective techniques and tools to analyze and optimize resource behavior. Existing process analysis tools often focus on automation opportunities or overall process flow, neglecting the specific impact of resource behavior. As a result, organizations struggle to identify areas for improvement, make informed decisions about resource allocation and workload assignment, and maximize process efficiency. There is a clear need for research and solutions that address this gap and provide practical approaches for understanding and optimizing resource behavior in manual processes.

1.3 Research Question

How can the analysis and optimization of resource behavior in manual processes lead to improved process efficiency and performance?

The research question focuses on investigating how the analysis and optimization of resource behavior in manual processes can contribute to enhanced process efficiency and performance. Specifically, the research aims to explore techniques and approaches that can effectively analyze resource behavior, identify areas for improvement, and provide insights on optimizing resource allocation, workload assignment, and task execution. By addressing this research question, the thesis seeks to bridge the gap in understanding resource behavior and provide practical solutions to streamline manual processes, reduce delays, errors, and inefficiencies, and ultimately improve the overall performance of the business processes. The findings of this research have the potential to benefit organizations by providing valuable insights for process improvement initiatives, leading to increased operational efficiency, cost savings, and better customer experiences.

1.4 Research Goals

The primary goal of this thesis is to analyze the working behavior of resources and investigate its impact on process performance. The research aims to achieve the following objectives:

- **Insights into Treatment of Cases:** By analyzing the efficiency of resource behavior based on *key performance indicators (KPIs)* derived from past results, this research aims to provide insights into determining whether cases should be treated as `batch` or as individual entity within the process. Understanding the optimal treatment of cases can significantly enhance process efficiency and resource utilization.
- **Identification of Optimal Workload Range:** By analyzing resource behavior and its correlation with process performance, this research aims to identify the `optimal workload range` for resources. Understanding the range of work assignments that maximizes resource efficiency and effectiveness can help organizations avoid overload or under utilization, leading to improved process outcomes.
- **Identification of Optimal Work Prioritization Pattern:** Through the evaluation of resource behavior efficiency using specific KPIs, this research aims to identify the most effective `work prioritization pattern` for each resource. By understanding and determining the work prioritization pattern that best suits each individual, organizations can significantly enhance task management and scheduling. Empowering resources with knowledge of their optimal work prioritization pattern enables them to efficiently manage their tasks, prioritize activities, and allocate their time effectively. This leads to improved productivity, reduced waiting times, and enhanced overall process performance. The identification of the optimal work prioritization pattern for resources brings about a more streamlined and organized approach to task execution, resulting in optimized resource allocation and improved operational efficiency.
- **Optimal Resource Behavior:** The analysis of resource behavior will aim to discover patterns and behaviors that result in faster process execution. By identifying the `optimal working behavior` for resources, this research aims to provide guidelines for resource management that can streamline process flows, minimize delays, and increase throughput.

Achieving these research goals will contribute to a deeper understanding of the relationship between resource behavior and process performance. The findings and insights derived from this research will provide valuable guidance for organizations seeking to optimize their processes, enhance resource utilization, and improve operational efficiency.

1.5 Contributions

The primary focus of this thesis is to analyze resource behavior and its correlation with process performance in order to identify optimal working behavior. By leveraging process mining techniques, the thesis proposes a methodology that allows for the comprehensive analysis of resource behavior and its impact on process efficiency. Through this analysis, the thesis aims to contribute to the field of `process optimization` by providing organizations with valuable insights into resource behavior patterns and their influence on overall process performance. To facilitate the analysis, a dedicated tool will be developed as part of this research. The tool will take an event log and a selected KPI as input, allowing organizations to easily apply the proposed methodology to their own process data. The tool will generate insightful synthetic tables that provide a holistic view of resource behavior, including working behavior metrics, success rates, and optimal workload recommendations for each resource. These tables will enable organizations to assess the effectiveness of resource behavior and identify areas for improvement. Furthermore, a comprehensive evaluation will be conducted using real-world process data to validate the effectiveness and applicability of the proposed approach. The evaluation will assess the correlation between resource behavior and process success, providing organizations with quantitative evidence of the impact of resource behavior on overall process performance. This evaluation will demonstrate the practical value and reliability of the proposed methodology in improving process efficiency and achieving operational excellence. The contributions of this thesis extend beyond theoretical concepts, providing organizations with a practical tool and a validated methodology for optimizing resource behavior and enhancing process performance. By leveraging these insights, organizations can streamline their operations, reduce inefficiencies, and achieve improved operational outcomes. This research equips organizations with the means to make data-driven decisions, optimize resource behavior, and ultimately gain a competitive edge in today's business environment

1.6 Thesis Structure

The remainder of this thesis is structured as follows. In chapter 2, we discuss related work. In chapter 3, we present the basic preliminaries used throughout the thesis. chapter 4 focuses on the methods proposed to determine the working behavior of resources and their success rates. In chapter 5, we provide detailed explanations of the concrete implementation details involved. In chapter 6, we evaluate the results obtained using our proposed framework. chapter 7 involves the analysis of the extracted results and the challenges associated with the proposed methods. Finally, in chapter 8, we conclude the work, discuss future research opportunities, and present our prospects for further development.

Chapter 2

Related Work

The following section will provide an in-depth exploration of related work, exploring literature in the areas of `SAP and Process Mining`, `Goal-Oriented Process Mining`, `Organizational Mining`, with a focus on `batch activities`, `work prioritization` and `workload`. This literature review serves as a comprehensive exploration of existing research and provides a solid foundation for understanding the current state of knowledge in these domains. By examining previous studies, methodologies, and findings, this chapter aims to identify key insights, frameworks, and gaps in the literature. It will shed light on the advancements made in each area, highlight the existing challenges, and uncover potential opportunities for further research.

2.1 SAP and Process Mining

`SAP` and `process mining` are two important technologies in the field of business process management. SAP is a leading *enterprise resource planning (ERP)* software that helps companies manage various business processes, including human resources, finance and supply chain management. Process mining, on the other hand, is a technique that uses data from IT systems to analyze and visualize business processes.

Process mining is important for improving operations in SAP for several reasons:

- **Increased visibility:** Process mining provides a comprehensive view of the entire process, making it easier to identify areas for improvement and track progress over time.
- **Data-driven decision making:** By analyzing process data, process mining enables organizations to make informed decisions based on real-world evidence, rather than assumptions or guesses.
- **Improved efficiency:** Process mining helps to identify inefficiencies and bottlenecks in the process, allowing organizations to streamline operations and reduce costs.
- **Enhanced customer experience:** By improving the process, organizations can deliver a better customer experience, which can lead to increased customer satisfaction and loyalty.

- **Compliance:** Process mining helps organizations to ensure compliance with regulations and company policies, reducing the risk of penalties and legal action.
- **Improved agility:** Process mining provides real-time insights into process performance, enabling organizations to quickly respond to changing business needs and market conditions.

There have been several research works and case studies that have investigated the application of process mining in SAP systems. These studies have demonstrated the potential benefits of using process mining to improve the efficiency, transparency, and compliance of SAP-based business processes. Some of the areas where process mining has been applied in SAP systems include:

- **Supply chain management:** Process mining has been used to analyze and optimize SAP-based supply chain processes, such as procurement-to-pay and order-to-delivery.
- **Workflow optimization:** Process mining has been used to identify bottlenecks and inefficiencies in SAP workflows, such as customer order processing and purchase order processing.
- **Compliance monitoring:** Process mining has been used to monitor the compliance of SAP-based business processes, such as financial reporting and tax compliance.

The case study *"Global Chemical Manufacturer Saves $1.2M by Optimizing Order to Cash (OTC)"* describes how a chemical manufacturer used `process mining` to identify inefficiencies in their `order to cash` process, which encompasses all the steps from receiving an order to collecting payment. OTC processes that are slow or difficult to use can cost businesses money, lose clients, harm their reputations, and even cause regulatory problems. By analyzing their event data, the manufacturer was able to identify Unusual cases, impact of changes on the case duration, pinpoint bottlenecks and delays and make targeted improvements to reduce lead times and increase productivity. This resulted in a savings of *$1.2 million* annually, as well as improved customer satisfaction and reduced risk of order cancellations. The case study highlights the potential of process mining to drive business process improvements and generate significant cost savings.

The *"Telco Giant Saves $5M on Purchase to Pay (P2P) with Process Mining"* case study details how a telecommunications business utilized process mining to streamline their `purchase-to-pay` process, which encompasses all the stages from placing orders for goods or services to making payments to vendors. The P2P cycle affects numerous departments and systems, ranging from straightforward purchases like office supplies to intricate procurement agreements like data access. P2P is ideal for process mining due to its nature as a multi-departmental process. The Telco was able to pinpoint inefficiencies in their P2P process, such as costly rework, lengthy lead times, high error rates, and implement modifications to lower costs and boost efficiency. The case study demonstrates how process mining has the potential to significantly reduce costs and enhance processes in large, complicated businesses.

A logistics firm employed process mining to streamline its claims handling procedure, as shown in the case study *"Logistics Company Saves €466K by Reducing Case Duration by 3 Hours"*. This case study explains how `Robotic Process Automation (RPA)` can save hundred thousand of dollars by minimizing human intervention in trivial tasks. In

IT systems, robotic process automation (RPA) refers to the use of software to simulate human behavior. Frustration should not be a factor when deciding whether to automate a procedure.It is impossible to automate the task if human judgment is required. The case study shows that by analyzing the data process mining helped the client automate its key processes and lead to a saving of *€466K* over a 5 month period.The company was also able to identify inefficiencies in their process, such as delays and errors, and make improvements to reduce the duration of claims cases by three hours.

These case studies highlight the potential of combining SAP and process mining to improve business processes, increase transparency, and ensure compliance. They also demonstrate the versatility of process mining as a technology and its ability to be applied to a wide range of SAP-based business processes. Overall, the combination of SAP and process mining has the potential to provide organizations with a powerful tool for managing and optimizing their business processes.

2.2 Organizational Mining

`Organizational mining` refers to the process of extracting knowledge from organizational data using `data mining` techniques. It involves analyzing data related to an organization's structure, behavior, and performance to identify patterns, trends, and insights that can be used to improve the organization's efficiency, productivity, and overall performance. In the context of business processes, organizational mining can involve analyzing data related to the process flows, resource utilization, and other metrics to optimize the process and reduce waste.The insights gained through organizational mining can be used to drive positive change and improve performance in different areas of the company. For example, they can be used to identify areas for improvement, optimize processes, and enhance communication and collaboration between departments.

2.2.1 Resource Utilization and Working Behaviour

In most business processes, several activities need to be executed by human resources and cannot be fully automated. To evaluate resource performance and identify best practices as well as opportunities for improvement, managers need objective information about resource behaviors. Efficient resource management is a key success factors for all businesses. A comprehensive understanding of the complex relation between resources and activities enables efficient resource allocation and potential cost reductions[1][2].

`Resource behavior` in process mining refers to the behavior of the resources involved in a business process, such as employees, machines, or systems. It involves analyzing how resources are utilized within a process, including their workload, availability, and utilization. In process mining, resource behavior is usually represented by performance indicators such as resource utilization, cycle time, waiting time, and idle time. The analysis of resource behavior can help organizations to understand how their resources are being used and identify areas for improvement. For example, if a resource is frequently idle, this may indicate an opportunity to redistribute workload or allocate additional resources. Understanding resource behavior is crucial for optimizing process performance and ensuring that resources are being used efficiently and effectively.

Early process mining techniques were developed to discover process models from event logs[3]. The `control-flow` perspective of a process, or the temporal relationships between different business process activities, is captured by a process model. Recently, methods to find models reflecting additional process views, such as the `data-flow` perspective [4] (It demonstrates how data was ingested and transformed) and the organizational perspective [5][6][7][8][9][10] (Resource involvement during process execution), have been added to process mining algorithms.

`Song` and `van der Aalst` [8] were some of the earliest to investigate topic related to resources in the context of Process Mining. Their primary objective was to uncover social networks and organizational structures from event logs. They accomplished this by utilizing task-based measurements that depended on shared activities. Even today, these concepts are still in use, as seen in the work of [11], who used them to identify groups of resources with similar tasks in their tool. `Berti` [12] proposes a method for identifying the key actors in an organizational social network using process mining techniques.The focus is on identifying two types of key actors. *Weighted leaders* are individuals who have a significant impact on the performance of the organization, while *peripheral workers* are individuals who are not directly involved in key activities but have some influence on the network.The main practical goal of the proposed methods is to enable the efficient identification of high and low performing workers based on event logs, thus facilitating the evaluation of employees.

Organizational mining research has primarily addressed problems relating to resource performance. `Song et al.` proposed techniques for mining organizational models [8]. Regression analysis is used by `Nakatumba` and `van der Aalst` to study the effects of employee workloads on service times [9]. Measures for resource availability, competency and cooperation are proposed by `Huang et al.`, who also demonstrate how to find them from logs [10]. `Kim et al.` [13] provide a method to build a decision tree (built from historical performances of resources) that may be used at run-time to determine the optimum resource to be assigned a specific work-item. `Pika et al.` propose a broad framework in [14] for using time series analytic tools to identify changes in resource behavior across time. To mine resource scheduling algorithms for service systems, `Senderovich et al.` provide two different approaches: data mining with decision trees and queuing heuristics [15]. `Suriadi et al.` [16] works also manipulates the time intervals between different events to create multiple representations of resource performance, including throughput, workload, and idle times. RALph Miner, a program to find graphical resource-aware process models that combine multiple job assignment criteria, was created by `Cabanillas et al.` [17].

`Akhavian` and `Behzadan` [18] propose a technique to learn the queuing discipline employed by resources. They initially collect data from a sensor and learns queuing discipline in a batch manner, with evaluation of the queuing discipline used taking place once the entries in the log have been processed. The research of `Senderovich et al.` [19][20][21]aims to learn queue lengths for the purpose of online delay prediction, assuming a specific sort of queuing discipline is used. [22] focuses on learning the way resources select and prioritize their work. It suggests methods for determining the sequence in which work items are conducted by a resource (i.e., the queuing discipline adopted by resources). Instead of estimating the size of the queue or the length of the wait, it aims to find out what kind of queuing discipline the resources use. The work prioritization approach proposed in this article try to determine from data how a resource prioritizes the tasks that are given to him or her, whether in a *FIFO* or *LIFO* manner or using another prioritization method. [23][24][25] focuses on the performance of systems/processes based on the choice of queuing discipline. However, it always assumes a particular type of queuing discipline to start with.

2.2.2 Batch Detection

`Batch detection` in process mining refers to the process of identifying groups of similar activities in a business process that are executed in a batch manner. Batch processing is a common practice in many organizations where a large number of similar tasks are performed in a single batch, rather than individually. Batch detection in process mining is used to identify and analyze these batches of activities, to gain insights into their characteristics, such as their size, duration, and frequency. The information obtained from batch detection can be used to optimize the process, for example by reducing the batch size, shortening the batch duration, or improving the batch scheduling. Batch detection can also help to identify the causes of bottlenecks or inefficiencies in the process, and support continuous process improvement.

Batch processing is an approach that is studied in several domains such as computer science, operations management (manufacturing, and queuing). In computer science it refers to programs processing a series of jobs without any manual intervention, such program can run in the background without interrupting any important process. In operations management it is employed to address a variety of scheduling issues, including those related to the `supply chain` [26], `order picking` [27], and `machine scheduling` [28], with the goal of lowering holding costs, inventory expenses, and machine setup costs in exchange for longer delivery and service times.

Over the past ten years, batch processing has also drawn more attention in the field of business process management. `Pufahl` and `Weske` [29] were the first to bring the concept of batch activities to process modeling and execution. They define the idea of batch operations and describe specification criteria, such as batch size. While [29] concentrate on a single batch operation, [30] expand these ideas to `batch regions`.[31] then developed a strategy to assess the performance of the batch activity. As [29][30][32] focus on the activity level they do not consider how allocation of resources can affect the organization of work for a particular activity. A requirements framework was presented in [29] to provide an overview of additional factors to consider when integrating batch processing into a business process, such as resource involvement.

Although some aspects of batch processing have been well investigated, process mining still has a lot of unexplored territory. `Wen et al.` [33] have studied the issue of extracting batch process models from event logs. They provide an approach that modifies the event log such that a discovery algorithm can use it however the suggested algorithm needs to know where batching occurs in advance; this information must be provided during a pre-processing stage. In contrast, `Nakatumba` [34] investigates the issue of batch processing from a resource perspective to increase the resource awareness of business processes. Here, it is assumed that resources frequently operate in groups, which are identified by mining resource availability. This approach does detect batches, but it does not take into account the many types of batch processing or offer a way to understand batching behavior in great depth.

The most recent development of batch processing in process mining, is a method proposed by `Martin et al.` [35]. Based on the input of an event log,[35] suggests a technique to systematically examine the batching behavior. A `resource-activity matrix` is created by first converting an event log into an activity log. This matrix includes each occurrence of an activity that makes up a resource-activity pair. Following that, a batching matrix is made using this, in which the activity instances are organized into batches. Allocation of activity instance to batches is based on a set of guidelines forming three distinct categories. Metrics can be calculated to determine the impact of batching behavior on process performance. The distinction between simultaneous, concurrent, and sequential batching is the foundation of [35]. While [36] and [29] distinguish between the execution of activities in `parallel` and in `sequential` order, [31] and [33] only take `simultaneous batch processing` into account.

2.3 Goal Oriented Process Mining

`Goal-Oriented Process Mining` refers to a specific approach within the field of process mining that focuses on the evaluation of business processes with respect to a predefined set of goals or objectives. This approach aims to provide a clear and comprehensive view of how well a business process is aligned with its goals and what improvements can be made to better meet them. It involves the use of process mining techniques, such as `process discovery` and `process conformance checking`, to extract data from event logs and create process models that can be analyzed and compared to the predefined goals. The goal-oriented approach can be applied to various types of business processes, including IT systems, customer services, and supply chain management. The goal is to provide valuable insights for process improvement and optimization, and help organizations achieve their desired outcomes more effectively.

Goal modeling is a technique for requirements engineering that supports different types of reasoning, such as heuristic, qualitative, or formal schemes[37]. A goal is an objective that a system or its stakeholders aim to achieve. Typically, goals are represented by their type, attributes, and their connections to other goals and elements within a requirements model[38]. In recent years, various methods have been developed that utilize goal-oriented requirements engineering techniques to aid business process management activities, specifically business process modeling [39][40]. While process-oriented modeling languages emphasize questions like *"how," "what," "where," "who,"* and especially *"when,"* goal-oriented modeling focuses on answering *"who," "what,"* and particularly *"why"* questions. This approach provides a means of documenting intentions and rationales. By adopting a goal-oriented approach, modelers consider both the opportunities that stakeholders seek and the vulnerabilities they seek to avoid. Answering *"what"* questions helps identify the capabilities, services, and architectures required to satisfy stakeholder goals[41].

The combination of a goal-oriented approach with process mining activities shows potential not just for identifying necessary capabilities, services, and architectures, but also for documenting and utilizing goals and rationales to enhance process models and their implementation.The approach helps to ensure that the processes being analyzed are aligned with the overall objectives of the organization. It encourages collaboration between different stakeholders, promoting effective communication and ensuring that everyone is working towards the same goals.By analyzing the relationships between the goals and the processes, the approach can reveal inefficiencies and areas where improvements can be made to better achieve the desired outcomes, hence enabling organizations to make more informed choices about how to improve their processes.

Chapter 3

Preliminaries

3.1 P2P Process

The `Purchase to Pay (P2P) process`, also known as the Procure-to-Pay process, is a critical workflow in organizations that involves the acquisition of goods or services and the subsequent payment to suppliers. It encompasses a series of steps that ensure the seamless and efficient handling of purchasing activities. In the P2P process, it begins with the need for a purchase, such as acquiring raw materials, equipment, or services. The process typically starts with the creation of a `purchase requisition`, which outlines the details of the required item or service, including quantity, specifications, and delivery requirements. This requisition is then submitted for approval, ensuring that the purchase aligns with budgetary and procurement guidelines. Once the purchase requisition is approved, it moves forward to the creation of a `purchase order (PO)`. The PO serves as a formal agreement between the buyer and the supplier, specifying the items or services to be purchased, quantities, agreed prices, and delivery dates. The PO is then sent to the supplier for acknowledgment and acceptance. After the `goods or services` are received from the supplier, the next phase of the P2P process involves the `verification` of the received items against the purchase order. This step ensures that the goods or services match the specified requirements and are in acceptable condition. Any discrepancies or issues are addressed through a process of reconciliation and resolution. Once the received goods or services have been verified, the organization initiates the `invoicing process`. The supplier submits an invoice for payment, typically referencing the purchase order number. The invoice undergoes a series of `approval steps`, which may involve multiple stakeholders or departments depending on the organization's internal control procedures. Each approval step ensures the accuracy, completeness, and legitimacy of the invoice. If an invoice is approved at each step, it progresses towards final payment. The organization processes the payment based on agreed terms, which may include a specific payment period or payment conditions. `Payment` can be made through various methods, such as electronic funds transfer (EFT), checks, or online payment platforms. It is important to note that throughout the P2P process, organizations strive to maintain transparency, accuracy, and compliance with financial regulations and internal policies. This includes maintaining proper documentation, recording transactions, and ensuring appropriate segregation of duties to prevent fraud or unauthorized actions. The P2P process plays a vital role in managing an organization's procurement activities, ensuring that goods and services are acquired in a timely manner, at competitive prices, and with proper financial control. By optimizing and streamlining the P2P process, organizations can achieve cost savings, enhance supplier relationships, and improve overall operational efficiency.

In our thesis, we specifically concentrate on the `manual steps` within the Purchase to Pay (P2P) process. The diagram 3.1 presents a comprehensive illustration of a P2P process that involves manual interventions. Once a purchase order is created, it undergoes a series of `five approval steps` before it can be released. These approval steps serve as manual verification checkpoints to ensure the accuracy and correctness of the order. Similarly, when it comes to releasing the invoice, a `three-step approval` strategy is followed before making the payment. This emphasizes the crucial role of manual intervention in the P2P process. However, manual steps can introduce challenges such as delays and errors, which can hinder the overall efficiency of the process. To address this, our thesis aims to delve into understanding the working behavior of resources involved in these manual steps. By analyzing resource behavior, we seek to optimize their performance and enhance the efficiency of the P2P process as a whole. Through our research, we aim to uncover insights and strategies to improve resource behavior, streamline manual interventions, and ultimately boost the overall operational efficiency of the P2P process. By focusing on the optimization of resource behavior, we aspire to minimize delays, mitigate errors, and enhance the accuracy and speed of manual steps within the P2P process. By leveraging this understanding, organizations can achieve improved process outcomes, reduced costs, and enhanced customer satisfaction. Our thesis aims to contribute to the body of knowledge by providing valuable insights and recommendations for optimizing resource behavior and maximizing the efficiency of manual steps within the P2P process.

3.2 Event Logs

`Event logs` serve as a fundamental source of data in process mining, providing a comprehensive record of activities and events that occur within a process. These logs capture essential information such as the sequence of `activities`, `timestamps`, `case identifiers`, `resources` involved, and other relevant attributes. By analyzing event logs, researchers and practitioners gain valuable insights into process behavior, performance, and patterns.

3.2.1 Structure of Event Logs

Event logs are structured collections of recorded events, each representing a specific activity or event that took place at a particular time during the execution of a process. Each event in the log is associated with a case identifier, which groups related events together to form a process instance. This allows for the analysis of individual process instances as well as aggregated statistics across multiple instances.

The structure of an event log typically consists of several key components:

- **Case Identifier**: A unique identifier assigned to each process instance, enabling the grouping of related events.
- **Timestamp**: The timestamp indicates the time at which an event occurred, providing temporal information for analyzing process dynamics and performance over time.

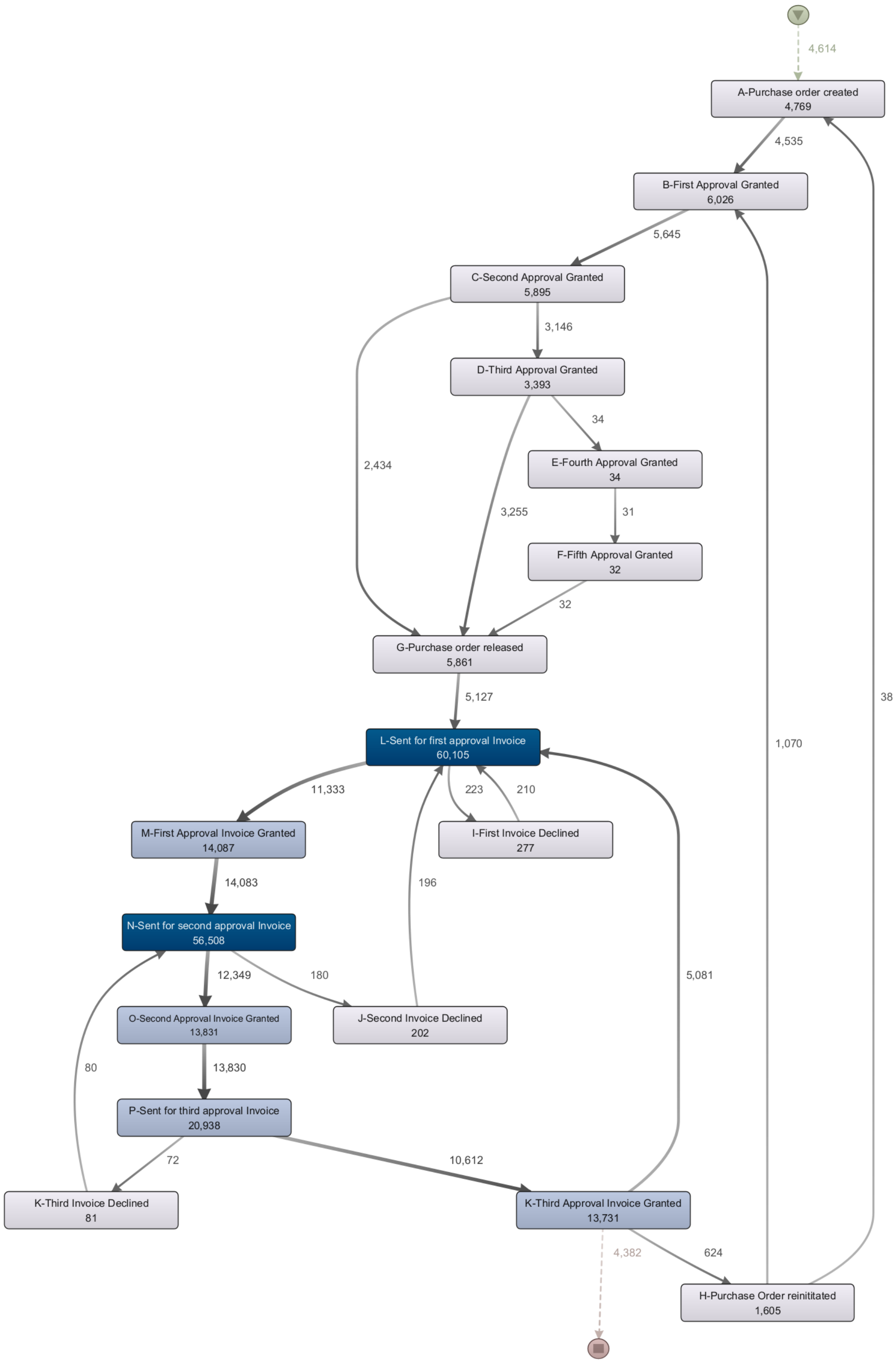


Figure 3.1: P2P Process - Manual Approval Steps

Patient	Activity	Resource	PatientName	Age	Insurance	start_timestamp	Timestamp	@@duration
1	Register	Alexander	Hermann the 1.	51	STAT	6/1/2020 6:00	6/1/2020 6:08	533
1	Initial Exam	Anna	Hermann the 1.	51	STAT	6/1/2020 6:10	6/1/2020 6:25	895
1	Decide Treatment	Ben	Hermann the 1.	51	STAT	6/1/2020 8:32	6/1/2020 8:42	574
1	Treatment B	Benjamin	Hermann the 1.	51	STAT	6/1/2020 9:07	6/1/2020 9:20	796
1	Check Treatment B	Benedikt,Benjamin	Hermann the 1.	51	STAT	6/1/2020 9:23	6/1/2020 9:28	297
1	Discharge	Brigitte	Hermann the 1.	51	STAT	6/6/2020 10:58	6/6/2020 11:04	381
2	Register	Alexander	Georg the 2.	32	STAT	6/1/2020 9:08	6/1/2020 9:18	562
2	Initial Exam	Ava	Georg the 2.	32	STAT	6/1/2020 9:19	6/1/2020 9:35	936
2	Decide Treatment	Benedikt	Georg the 2.	32	STAT	6/1/2020 11:46	6/1/2020 11:56	633
2	Treatment A1	Brigitte	Georg the 2.	32	STAT	6/1/2020 12:17	6/1/2020 12:37	1191
2	Check Treatment A1	Ben,Brigitte	Georg the 2.	32	STAT	6/1/2020 12:39	6/1/2020 12:43	273
2	Discharge	Benjamin	Georg the 2.	32	STAT	6/12/2020 12:17	6/12/2020 12:25	439
3	Register	Alexander	Hermann the 3.	19	STAT	6/1/2020 9:18	6/1/2020 9:28	609
3	Initial Exam	Amalia	Hermann the 3.	19	STAT	6/1/2020 9:30	6/1/2020 9:46	992
3	Decide Treatment	Ben	Hermann the 3.	19	STAT	6/1/2020 12:10	6/1/2020 12:18	494
3	Treatment B	Bernhard	Hermann the 3.	19	STAT	6/1/2020 12:42	6/1/2020 12:56	843
3	Check Treatment B	Benedikt,Bernhard	Hermann the 3.	19	STAT	6/1/2020 13:00	6/1/2020 13:04	278
3	Discharge	Bettina	Hermann the 3.	19	STAT	6/6/2020 14:31	6/6/2020 14:38	433

Figure 3.2: An Example of Event Log

- **Activity**: The activity represents a specific action or task within the process. It provides insights into the sequence of activities performed and the flow of the process.
- **Resource**: The resource associated with an event represents the entity (e.g., person, department, system) responsible for executing the activity. Analyzing resource assignments can help identify bottlenecks, resource utilization patterns, and workload distribution.
- **Other Attributes**: Additional attributes can be included in event logs to capture additional information relevant to the process, such as the duration of an activity, the outcome of an event, or any contextual data that may influence process execution.

Figure 3.2 illustrates an example event log extracted from a hospital case study. The event log captures the sequence of activities performed during the patient journey, including registration, treatment, and discharge. Each entry in the event log contains essential information such as the case ID, the resource involved, timestamps indicating when the activity occurred, and the specific activity performed. The combination of these components provides a rich and detailed representation of the process execution, enabling comprehensive analysis and exploration of process behavior.

3.2.2 Data Collection and Preprocessing

Collecting event logs involves recording activities and events as they occur during the execution of a process. This can be done through various data collection methods, such as system logs, databases, or manual data entry. The chosen method depends on the availability of data sources and the specific context of the process being analyzed. Once collected, event logs may undergo preprocessing steps to ensure data quality and usability. `Preprocessing` tasks may include data cleaning, filtering, and transformation to remove inconsistencies, noise, or irrelevant information. It is essential to ensure that event logs are reliable and accurately reflect the executed process.

`Event logs` play a pivotal role in process mining, serving as a foundation for understanding and improving processes. They provide a detailed and objective record of process executions, enabling comprehensive analysis and valuable insights into process behavior, performance, and resource utilization. By leveraging event logs and applying process mining techniques, organizations can enhance their understanding of processes, optimize resource allocation, reduce inefficiencies, and achieve improved operational outcomes.

3.3 Key Performance Indicators (KPIs)

`Key Performance Indicators (KPIs)` are quantitative or qualitative metrics used to measure the performance and effectiveness of processes, activities, or resources within an organization. They provide insights into various aspects of performance, helping organizations assess their progress towards specific goals and objectives. In the context of process mining, KPIs serve as essential indicators for evaluating process efficiency, resource utilization, and overall process performance.

3.3.1 Types of KPIs

There are several types of KPIs that can be used to measure different aspects of process performance. Here are some common types of KPIs used in process mining:

- **Time-related KPIs**: These KPIs focus on measuring the `time taken` to complete activities, tasks, or the entire process. Examples include cycle time, throughput time, waiting time, and lead time. Time-related KPIs provide insights into process efficiency, bottlenecks, and opportunities for improvement.
- **Quality-related KPIs**: `Quality-related KPIs` assess the accuracy, reliability, or compliance of process outputs or outcomes. Examples include error rates, rework rates, customer satisfaction scores, and compliance levels. Quality-related KPIs help identify areas for quality improvement and ensure that process outputs meet the desired standards.
- **Cost-related KPIs**: `Cost-related KPIs` focus on measuring the financial aspects of process performance, such as costs per activity, costs per case, or overall process costs. Examples include cost per order, cost per transaction, and cost per unit produced. Cost-related KPIs help organizations assess the efficiency of their processes in terms of resource utilization and cost-effectiveness.
- **Productivity-related KPIs**: `Productivity-related KPIs` measure the output or value generated by resources or processes relative to the input or effort invested. Examples include output per resource, tasks completed per hour, or revenue generated per process. Productivity-related KPIs help organizations assess the efficiency and effectiveness of their resources and processes.
- **Compliance-related KPIs**: `Compliance-related KPIs` measure the adherence to specific regulations, standards, or guidelines within a process. Examples include compliance rates, adherence to SLAs (Service Level Agreements), or adherence to regulatory requirements. Compliance-related KPIs ensure that processes operate within the defined legal and regulatory boundaries.

3.3.2 Selection and Analysis of KPIs

Selecting the most relevant KPIs for process mining depends on the specific goals and objectives of the analysis. It is important to identify KPIs that align with the organization's overall strategy, process improvement initiatives, and stakeholder requirements.

KPIs should be specific, measurable, attainable, relevant, and time-bound (SMART) to ensure their effectiveness in driving process improvement.

Once selected, KPIs are analyzed using process mining techniques to assess process performance, identify patterns, and discover areas for improvement. Process mining enables the extraction of valuable insights from event logs and the correlation of KPIs with process behavior, resource utilization, and overall performance. By analyzing KPIs in conjunction with process data, organizations can identify bottlenecks, inefficiencies, and opportunities for optimization.

3.3.3 Benefits of KPI Analysis

Analyzing KPIs in process mining provides several benefits for organizations:

- **Performance Measurement**: KPI analysis allows organizations to measure and monitor `process performance` objectively. It provides a clear understanding of how well processes are performing in relation to defined goals and targets.
- **Process Improvement**: By analyzing KPIs, organizations can identify areas for `process improvement` and optimization. KPI analysis helps in identifying bottlenecks, inefficiencies, and deviations from desired performance levels, enabling organizations to take corrective actions.
- **Resource Optimization**: KPI analysis helps organizations `optimize resource` allocation and utilization. By analyzing resource-related KPIs, organizations can identify underutilized or overloaded resources, balance workloads, and improve overall resource efficiency.
- **Decision Support**: KPI analysis provides valuable insights for `decision-making`. It enables data-driven decision-making by providing objective information on process performance and identifying areas that require attention or intervention.
- **Continuous Monitoring**: KPI analysis facilitates `continuous monitoring` of process performance. By regularly monitoring KPIs, organizations can track their progress, detect performance trends, and ensure sustained process improvement over time.

By leveraging the power of KPI analysis in process mining, organizations can gain deep insights into process performance, resource utilization, and overall operational effectiveness. This knowledge enables data-driven decision-making, fosters process improvement, and supports the achievement of organizational goals.

3.4 Work Prioritization Pattern

In process mining, `work prioritization pattern` can be analyzed to gain insights into how work is being managed and prioritized within a process by the resource. These patterns can provide valuable information for process improvement. Here are some common types of the prioritization patterns

- **First-In-First-Out (FIFO):** This is a common prioritization pattern where work is processed in the order it was received. This pattern is often used in processes where fairness and equality are important, such as customer service or IT support.

By following the `FIFO pattern`, resources ensure that tasks are handled in a timely and equitable manner.

- **Last-In-First-Out (LIFO):** This prioritization pattern is the opposite of FIFO. In the `LIFO pattern`, the last item to enter the process is given priority and processed first. This pattern is often employed in manufacturing processes where the most recently produced items are considered more valuable or perishable. By adopting the LIFO pattern, resources can prioritize the most recent tasks to meet time-sensitive requirements or optimize resource utilization.
- **Random :** `Random queuing` discipline refers to a queuing system that assigns an equal probability to each piece of work for being selected by a resource for processing, instead of assigning a fixed order or priority. An instance of such a queuing system is its implementation in a call center where the next call in the queue is randomly chosen to be processed when a representative is available, thereby enabling them to handle various call types and not just a specific one. The goal is to ensure fairness in resource utilization and avoid bottlenecks or under utilization.

When a process is executed, the details are logged in as event logs. Our approach utilizes this information from event logs to determine the order of prioritization that resources follow while performing their tasks.In figure 3.3 each task is represented by the activity name and the corresponding case to which it belongs. For instance, a task of *create PO* assigned to resource *R1* that needs to be completed for a case identified as *c1* is represented as `[c1, CPO]`. Resource *R1* follows the `First In First Out (FIFO)` method for task prioritization, which means that tasks are carried out in the same sequence in which they are assigned. On the other hand, Resource *R2* uses the `Last In First Out (LIFO)` method, which implies that the tasks assigned at the end are given priority and completed first. Resource *R3* uses a `Random` order to complete tasks, which means that tasks are completed in an unordered manner without a specific pattern or sequence. By understanding and analyzing the work prioritization patterns of resources, organizations can gain valuable insights into resource behavior and task management. This knowledge can be utilized to optimize resource allocation, improve task scheduling, and enhance overall process efficiency. It enables organizations to make informed decisions regarding workload distribution and task prioritization, resulting in improved productivity, reduced delays, and enhanced operational outcomes.

3.5 Batch Detection

In an organizational process, cases are typically passed from one activity to another until the case is completely handled and the process is finished. Both cases and activities within a process have their own unique properties that can affect the order in which cases are processed and whether they can be processed simultaneously in batches. Cases that share common properties may be processed together in batches, while cases that are different require a customized approach. Additionally, activities can be carried out periodically, meaning cases waiting for that specific activity are held until a particular moment in time when they are then processed in a batch. The handling of cases and activities can be optimized by identifying opportunities for batch processing. `Batch processing` involves grouping together similar cases or activities to be processed simultaneously, which can lead to improved efficiency and resource utilization. The different types of batch processing

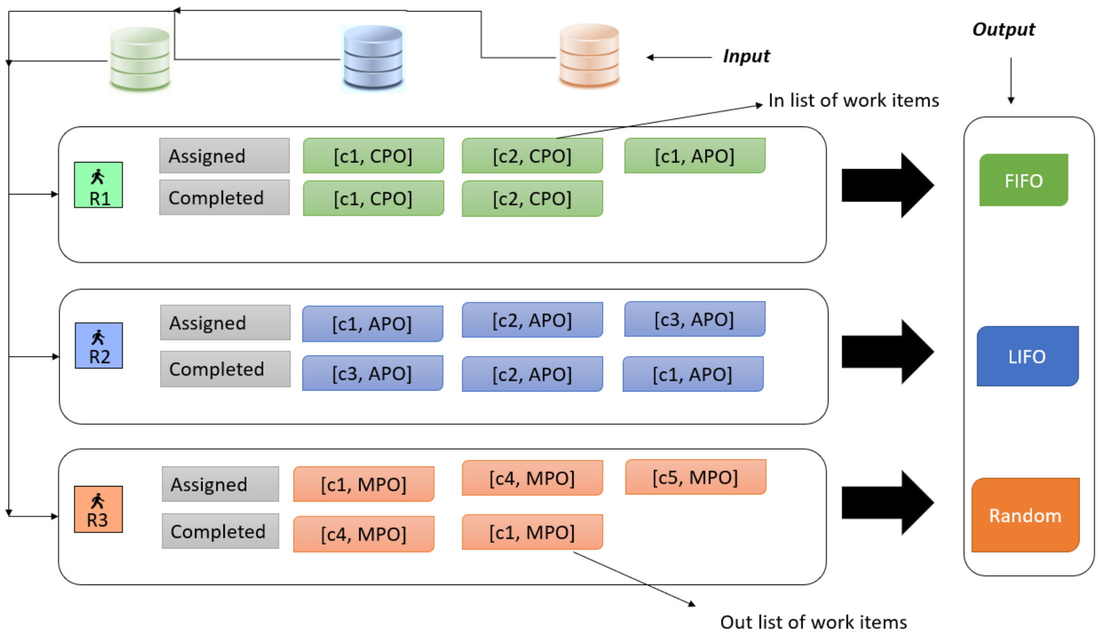


Figure 3.3: Explanation of the Work Prioritization Patterns

can be classified as follows:

- **Simultaneous Batch Processing:** `Simultaneous batch processing` occurs when the same resource executes instances of an activity for separate cases at the exact same time. This approach is suitable when cases share common characteristics or require the same type of processing. For example, in a manufacturing setting, multiple car parts requiring the same color paint can be processed together in a spray booth. This simultaneous batch processing reduces the setup time and maximizes resource efficiency.

- **Sequential Batch Processing:** `Sequential batch processing` involves a single resource executing instances of an activity for separate cases one after another, with minimal or no time delay between them. This pattern is useful when cases can be processed quickly and efficiently in a sequential manner. For instance, an employee may handle their emails twice a day in a sequential batch, where there may be a slight delay of a few seconds between processing different emails. This approach allows for efficient task switching and reduces time spent on context switching.

- **Concurrent Batch Processing:** `Concurrent batch processing` occurs when instances of an activity are executed by the same resource for separate cases that partially overlap in time. This approach allows for the efficient utilization of resources and can help reduce idle time. For example, a clerk can start processing a second invoice while additional information is being gathered to complete the processing of the first one. Concurrent batch processing ensures a continuous flow of work and minimizes delays in case processing.

Figure 3.4 displays the different types of batching patterns. By detecting and leveraging batch processing opportunities, organizations can streamline their processes, reduce idle time, and improve overall operational efficiency. The identification of batch processing pat-

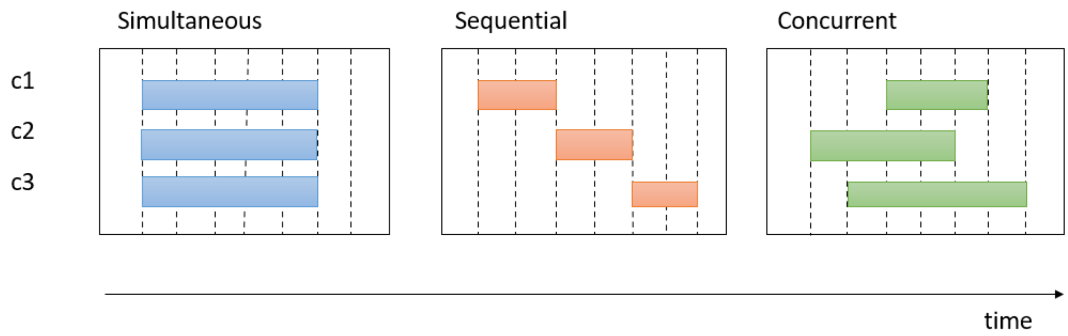


Figure 3.4: Illustration of Batching Patterns

terns allows for better resource planning and scheduling, leading to enhanced productivity and reduced processing time for cases within the organizational process.

3.6 Workload

The `workload` of resources plays a crucial role in the efficiency and performance of organizational processes. It refers to the amount of work assigned to and executed by each resource within a given time frame. Understanding and managing the workload of resources is essential for optimizing resource utilization, balancing work distribution, and ensuring timely task completion.

In the context of organizational processes, workload management involves:

- **Workload Allocation:** Determining the appropriate distribution of tasks and activities among available resources based on their skills, capabilities, and availability. `Workload allocation` aims to ensure that each resource has a fair and manageable workload without being overwhelmed or underutilized.
- **Workload Monitoring:** Regularly tracking and monitoring the workload of resources to identify potential bottlenecks and imbalances. `Workload monitoring` allows for timely intervention and adjustments to prevent excessive workloads, mitigate risks, and optimize resource allocation.
- **Workload Balancing:** Adjusting the distribution of work among resources to achieve a more balanced and equitable workload distribution. `Workload balancing` helps prevent overburdening certain resources while ensuring that all resources are effectively utilized and their skills are optimized.
- **Workload Optimization:** Analyzing the workload patterns and resource capacities to identify opportunities for `workload optimization`. This involves identifying workload peaks and valleys, optimizing task sequencing, and identifying areas where automation or resource reallocation can improve overall workload management.

Effectively managing the workload of resources ensures that tasks are completed efficiently, deadlines are met, and resource capacity is utilized optimally. It contributes to improved productivity, reduced delays, and enhanced overall process performance. By understanding the workload patterns of resources and implementing appropriate workload management strategies, organizations can achieve better resource utilization, mitigate risks, and achieve their process objectives more effectively.

Chapter 4

Method

This section outlines the approach and methodology employed in this study to evaluate the impact of resource working behavior on process performance. By following a systematic and rigorous approach, we aim to gain a comprehensive understanding of resource behavior and its influence on overall process efficiency.

4.1 Overview of the Proposed Approach

Our approach incorporates an `event log`, capturing detailed activity information, along with a carefully chosen `KPI` from a set of available options. By leveraging the event log data and the selected KPI, we conduct a comprehensive analysis of resource working behavior, examining factors such as `work prioritization pattern`, `batching strategies`, and `workload` considerations. The outcomes of this analysis are synthesized into concise `synthetic tables`, enabling a clearer understanding of how resource behavior impacts overall process performance.

4.1.1 Event Log and KPI

Our framework is built upon two essential inputs: the `event log` and a `KPI` used to evaluate and measure the success of process instances. The `event log` forms the foundation of our analysis, providing valuable insights into the activities performed, timestamps, resource utilization, and the expected and actual completion dates for each case. The `KPI` serves as a critical criterion for assessing process success, allowing us to gauge performance and make informed decisions based on the desired outcomes.

4.1.2 Work Prioritization Pattern Analysis

The first step of the approach focuses on analyzing the `work prioritization patterns` adopted by resources. By understanding how resources prioritize their tasks, valuable insights can be gained regarding the efficiency and effectiveness of their decision-making process. This analysis involves examining the strategies employed by resources to determine the task order. To measure the impact of work prioritization patterns on process performance, a correlation analysis is conducted. The observed patterns are evaluated against KPIs to determine their success rates. This analysis helps identify effective patterns that lead to higher success rates and inefficient patterns that may hinder process performance.

4.1.3 Batching Evaluation

The second step involves evaluating the batching strategies applied by resources. `Batching` refers to the grouping of similar tasks for more efficient processing. By examining how resources group and process tasks, insights can be gained into the impact of batching on process performance. The correlation between batching patterns and process success is assessed. This evaluation helps identify the impact of batching strategies on the overall success of the process, which contributes to improved process performance and enables the identification of potential bottlenecks or inefficiencies.

4.1.4 Workload Consideration

The approach also considers the workload of resources in the analysis. `Workload` refers to the volume and complexity of tasks assigned to individual resources. By assessing resource success rate across various workload ranges, insights can be gained into the impact of workload on process performance. This evaluation helps identify resource imbalances, potential overloads, and areas where workload distribution can be optimized.

4.1.5 Display of Results

To facilitate a comprehensive understanding of the findings, the approach includes the display of results in synthetic tables. `Synthetic tables` provide a concise and visually appealing representation of the analyzed data, making it easier to interpret and compare different aspects of resource working behavior and their impact on process performance. These tables summarize the correlation between work prioritization patterns, batching strategies, workload, and process success rates, enabling stakeholders to quickly identify patterns and trends.

Figure 4.1 presents a high-level overview of our framework. The framework takes an `event log` and a `KPI` as inputs. The event log undergoes preprocessing to prepare it for analysis. Subsequently, we analyze the behavior of resources, focusing on three key aspects: `work prioritization patterns`, `batching patterns`, and `workload patterns`. These behavior patterns are carefully evaluated using relevant `metrics` and the selected `KPI`. The evaluation process involves assessing the `success rate` of each behavior pattern based on the chosen KPI and metric. This analysis provides valuable insights into the effectiveness and efficiency of resource working behavior in relation to process performance. To effectively communicate our findings, we present the results in the form of `synthetic tables` that cover three main aspects: the observed `working behavior patterns`, the corresponding `evaluation results`, and the `success rates` associated with each behavior pattern.

By using this comprehensive approach, stakeholders can understand how the way resources work affects the performance of a process. By analyzing patterns in work prioritization, strategies for grouping tasks, workload considerations, and presenting the results in easy-to-understand tables, we can establish a strong framework for improving how resources are allocated, tasks are ordered, and overall process efficiency is achieved.

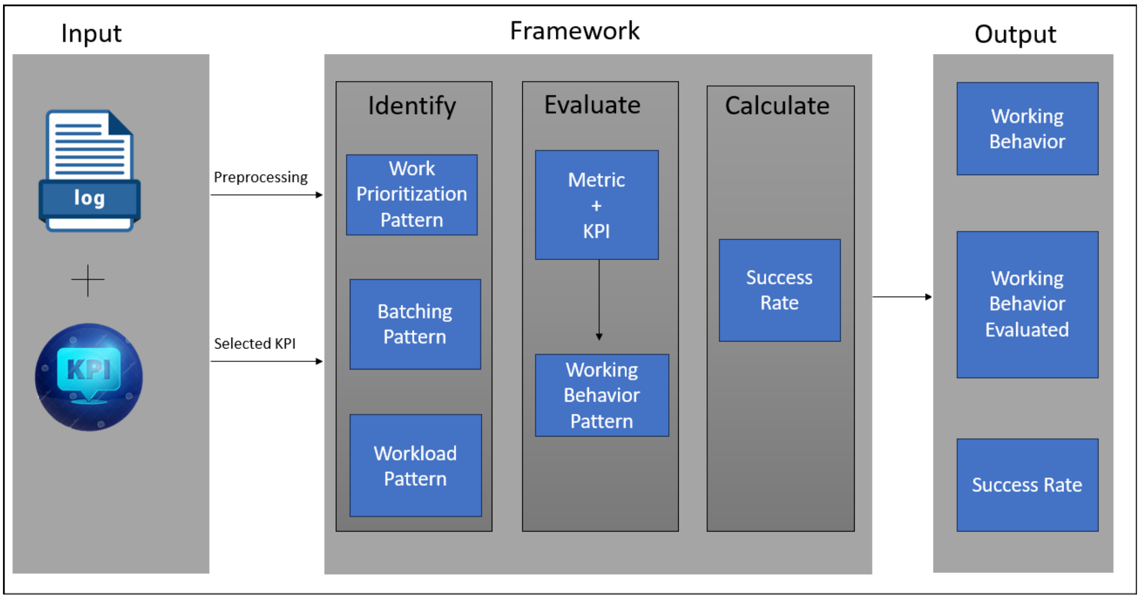


Figure 4.1: Overview of the framework

4.2 Event Log, Data Preparation and KPI Selection

To ensure the accuracy and reliability of the analysis, the `event log` data is cleaned and pre-processed. This step involves removing redundant rows and columns that do not significantly contribute to the analysis. Additionally, it addresses errors, inconsistencies, and missing data in the event log, which can arise from factors such as data entry mistakes or discrepancies between different data sources. The cleaned data is then transformed into a suitable format for analysis.

`Event logs` can come in different types, event logs that record a single event for each activity, with the start and completion timestamps being the same, event logs that record separate events for the start and completion timestamps of each activity. The approach is designed to effectively handle both types of event logs. The analysis methods and preprocessing techniques are adaptable to accommodate the characteristics of each type of event log. By considering the unique attributes and data structures of different event log types, the approach ensures accurate and meaningful results across various scenarios.

In our evaluation of resource working behavior with the success of the process instance, we employ `three KPIs` to measure the impact of working behavior on the overall performance of the process.

- **Lead Time**

  `Lead Time` assesses the total time required for a process or task from initiation to completion. It encompasses all the necessary steps and activities involved, including any waiting or idle periods. To determine the lead time, we analyze the event log data and identify the minimum and maximum timestamps for each case. By calculating the difference between these timestamps, we obtain the lead time for each case. This information is then added as a dedicated column in the event log. To establish a benchmark for being successful, we calculate the mean lead time for the entire event log. This mean value serves as a threshold for our KPI. If the lead time of a particular case exceeds the threshold, it is considered *bad case*, indicating a longer-than-average duration. Conversely, if the lead time falls below or equals the threshold, the case is deemed *good case*, denoting an efficient and timely completion.

By utilizing this KPI, we gain valuable insights into the performance of the process, enabling us to identify areas for improvement and potential bottlenecks. The lead time measurement provides a comprehensive view of the process efficiency and effectiveness, allowing us to make informed decisions and optimize resource utilization accordingly.

- **Rework**

  `Rework` refers to the additional work or corrective actions that are required on a process or task after its initial completion due to errors, deficiencies, or inconsistencies. It involves revisiting and modifying previously completed work to align it with the desired standards or requirements. Rework is typically necessary when the initial output does not meet the expected quality, specifications, or objectives. In our event log, we capture all the activities associated with each case ID. If we observe that a previously performed activity needs to be repeated for a particular case ID, it indicates the occurrence of rework. Consequently, when such a case arises, we classify it as a *bad case* since it has undergone rework. On the other hand, if a case does not involve any of these rework-indicating activities, it is considered a *good case* as it has progressed smoothly without the need for additional work. This classification serves as the foundation for evaluating the process. Cases that meet the rework criteria are part of the *bad log*, highlighting the negative impact of rework on process performance. Conversely, cases that avoid rework are part of the *good log*, indicating efficient and effective handling of the process.

  By employing this approach, we can assess the impact of rework on process performance and identify opportunities for process improvement. It allows us to focus on reducing rework instances and optimizing resource utilization to achieve more efficient and effective outcomes.

- **Due Date Compliance**

  One crucial KPI involves evaluating whether a case was completed before its planned end date. This KPI holds significant importance as delays in case completion can result in substantial penalties for the company. To assess this, we compare the `planned end date` of each case with its `actual end date`. During our event log analysis, we examine the columns that capture the planned end date and the actual end date of each case. If the planned end date is later than the actual end date, it indicates that the case was completed on time. Conversely, if the planned end date is earlier than the actual end date, it signifies a delay in case completion. To track this information systematically, we introduce an additional column in the data set that records whether each case was completed on time or not. By incorporating this KPI into our analysis, we can categorize cases into two distinct groups: the *good log* and the *bad log*. Cases that are completed within the defined time limit are classified as part of the *good log*, indicating successful adherence to deadlines. On the other hand, cases that are completed after the planned end date are categorized as part of the *bad log*, signifying delayed completion.

This KPI allows us to identify cases that may require improvement or optimization to ensure timely completion and mitigate potential penalties. By monitoring the percentage of cases completed on time over time, we can track the effectiveness of process enhancements and measure the overall case performance of the organization.

4.3 Exploring the Relationship between Work Prioritization and Process Instance Success

Prioritizing work is a critical aspect that has a significant impact on process performance. The way a resource chooses to prioritize their work can profoundly influence the overall success of the process. To address this, the proposed method aims to identify the `prioritization patterns` employed by the resource. These patterns will then be evaluated based on a predetermined set of `KPIs`. By assessing these patterns, we can determine the `success rate` associated with each pattern. Understanding and evaluating the prioritization patterns employed by resources can provide valuable information for process improvement. It enables us to identify effective strategies and areas for optimization. By analyzing the impact of different prioritization patterns on task completion and overall success, we can make informed decisions to enhance productivity, streamline workflows, and achieve better outcomes.

4.3.1 Defining Prioritization Pattern

A `prioritization pattern` refers to the specific approach or strategy used by a resource to determine the order and importance of their tasks or activities. It involves how the resource assigns priority levels or considers specific criteria to determine which tasks should be addressed first or given higher importance in terms of completion. The prioritization pattern reflects the resource's decision-making process in managing their workload. It encompasses factors such as urgency, deadlines, dependencies and the overall impact of each task on the desired outcome. By examining the prioritization pattern, we gain insights into how the resource organizes and prioritizes their tasks to achieve optimal efficiency and effectiveness.

We will categorize work prioritization patterns into `three` main types:

- **FIFO (First-In-First-Out):** When a resource follows a `FIFO` approach, they prioritize tasks based on their arrival order. In other words, the resource handles the tasks in the order they were received. This strategy ensures that older tasks are addressed first, maintaining a sense of fairness and preventing potential bottlenecks caused by task accumulation. FIFO prioritization can help maintain a smooth workflow and ensure timely completion of tasks. It provides a structured approach that minimizes task delays and potential dissatisfaction from stakeholders waiting for their tasks to be addressed.

- **LIFO (Last-In-First-Out):** In contrast to FIFO, a `LIFO` approach prioritizes the most recently received tasks over older ones. The resource focuses on completing the newest tasks first before moving on to older ones. This strategy can be useful in situations where there is a need to address urgent or time-sensitive tasks immediately. LIFO prioritization allows resources to quickly respond to new requests or changes in priorities. However, it may result in a backlog of older tasks, potentially leading to delays and a sense of unfairness among stakeholders with pending tasks.

- **Random Prioritization:** `Randomly prioritizing work` involves selecting tasks without any particular order or pattern. This approach introduces an element of unpredictability and can be useful in situations where all tasks hold equal importance or when resources need to introduce variety in their work routine. Random prioritization can prevent biases and ensure a fair distribution of attention to different tasks. However, it may also lead to potential inefficiencies, as critical or time-sensitive tasks could be delayed or overlooked due to the random selection process.

4.3.2 Identification of the Prioritization Pattern

In order to identify the `prioritization pattern` opted by the resources, a filter will be implemented first on the `resource` attribute of the event log to generate a list of all the resources that carried out distinct activities in the event log. To determine a resource's work prioritization pattern, we must determine the order in which the case IDs were assigned to the resource and the order in which he performed the activities for these case IDs. We assess each resource separately and obtain all the activities performed by that resource and their respective case IDs. This information is stored in a data-set. Then we scan through the event log for the case IDs of the activities on which the resource worked and extract the `completion time` of their `preceding activity`. The completion time of the previous activity for these case IDs is recorded in another data-set. We make an `assumption` that as soon as the preceding activity for these case IDs is completed, the case ID is assigned to the resource to perform its next activity.

Figure 4.2 shows a list of activities that were completed by the resource *"Z0005C7T"* this information is stored in the first data set, and Figure 4.3 shows the list containing the `preceding activity` and its completion time of the case IDs that occurred in the first data set.

	Case ID	Activity	Resource	Complete Timestamp
0	3313406486	H-Purchase Order reinititated	Z0005C7T	2018-01-12 14:48:55.000
1	3313408420	H-Purchase Order reinititated	Z0005C7T	2018-01-17 15:13:04.000
2	3313406495	M-First Approval Invoice Granted	Z0005C7T	2018-01-18 08:46:06.000

Figure 4.2: Cases Where an Activity Was Performed by the Resource

We now have two data sets at hand for all the cases. The first data set contains information about all the case IDs for which an activity was performed by the resource. The second data set contains information about the completion time of the preceding activity of these case IDs. The two data sets are modified by adding a new column named *Execution*. For the first data set, the Execution column is updated with the value *"Completed"* as it holds all the activities that were performed/completed by the resource being evaluated. For the second data set, the *Execution* column is updated with the value *"Assigned"* as it holds the information regarding when these activities were assigned to the resource being evaluated. After adding the Execution column and updating the values, the two

	Case ID	Activity	Resource	Complete Timestamp
0	3313406486	D-Third Approval Granted	Z001723K	2018-01-12 10:09:52.000
1	3313408420	A-Purchase order created	Z0005C7T	2018-01-15 10:02:06.000
2	3313406495	L-Sent for first approval Invoice	Z0005C7T	2018-01-18 06:45:34.000

Figure 4.3: Preceding Activity of Cases on Which the Resource Worked

data sets are combined to obtain a comprehensive data set. Once we have collected this information, we can arrange it in increasing order of timestamps. By doing so, we can have a chronological summary of when the case IDs were assigned and when they were completed.

Figure 4.4 shows an example merged list for a `resource`. Whenever a case ID is assigned to the resource, it is marked in the *Execution* column as *"Assigned"*, and whenever the resource completes an activity, it is marked as *"Completed"* in the *Execution* column. The list is arranged in ascending order of timestamps to get a better understanding of how the tasks were assigned and how they were completed by the resource.

	Case ID	Activity	Resource	Complete Timestamp	Execution
0	3313406486	D-Third Approval Granted		2018-01-12 10:09:52.000	Assigned
1	3313406486	H-Purchase Order reinititated	Z0005C7T	2018-01-12 14:48:55.000	Completed
2	3313408420	A-Purchase order created		2018-01-15 10:02:06.000	Assigned
3	3313408420	H-Purchase Order reinititated	Z0005C7T	2018-01-17 15:13:04.000	Completed
4	3313406495	L-Sent for first approval Invoice		2018-01-18 06:45:34.000	Assigned
5	3313406495	M-First Approval Invoice Granted	Z0005C7T	2018-01-18 08:46:06.000	Completed
6	3313412998	J-Second Invoice Declined		2018-02-08 09:10:17.000	Assigned
7	3313422913	A-Purchase order created		2018-02-09 14:40:22.000	Assigned
8	3313416565	L-Sent for first approval Invoice		2018-02-13 03:03:36.000	Assigned
9	3313416565	M-First Approval Invoice Granted	Z0005C7T	2018-02-13 10:39:50.000	Completed

Figure 4.4: Resource Work List

The subsequent stage involves recognizing the `work prioritization pattern`. We will traverse through the previously acquired data set and maintain two distinct lists. The first list will document when the case IDs were *"Assigned"*, while the second will record when the case IDs were *"Completed"*. As soon as an assigned event is observed, we will add its relevant case ID to `list1`. We will continue to do this until a `completion event` arises, in which case we will add the corresponding case ID of the completed activity to `list2`. We will continue adding the case IDs of completed activities to list2 unless we come across a subsequent event indicating an assigned event. If such an event is encountered, the case IDs already present in list2 will be assessed.

We will compare the first indices of both list1 and list2, and if they match, we will label it as a `First-In-First-Out (FIFO) pattern`, increment the counter for FIFO, and store the case ID in a list called `FIFO cases`. If the first index of list2 is the last index of list1, we will call it a `Last-In-First-Out (LIFO) pattern`, increment the counter for LIFO, and store the case ID in a list called `LIFO cases`. If the match occurs somewhere in between, we will call it a `random pattern`, increment the counter for Random, and store the case ID in a list called `Random cases`. After evaluating a case ID, it will be `removed` from both lists. Finally, we will examine the counters for FIFO, LIFO, and Random to determine how a particular resource prioritizes the tasks assigned to them.

On the left-hand side of Figure 4.5, an example of a resource's `work list` is displayed in ascending order of timestamps. The work list indicates when a task is assigned to the resource by marking it as *"Assigned,"* and when the resource completes a task, it's marked as *"Completed."* As we traverse through the work list, we add the case IDs of assigned tasks to list1 unless a completed event is observed in the log. In that case, the corresponding case ID is added to list2. The lists are evaluated if an assigned event is observed after the completed event. In such a scenario, we compare the `expected FIFO/LIFO output` with the `actual output`. If the actual output matches the expected FIFO output, then the case is considered to be treated in a `FIFO` manner by the resource. Conversely, if the actual output matches the expected LIFO output, then the case is considered to be treated in a `LIFO` manner. If neither of the outputs match, then the case is considered to be treated `randomly` by the resource.

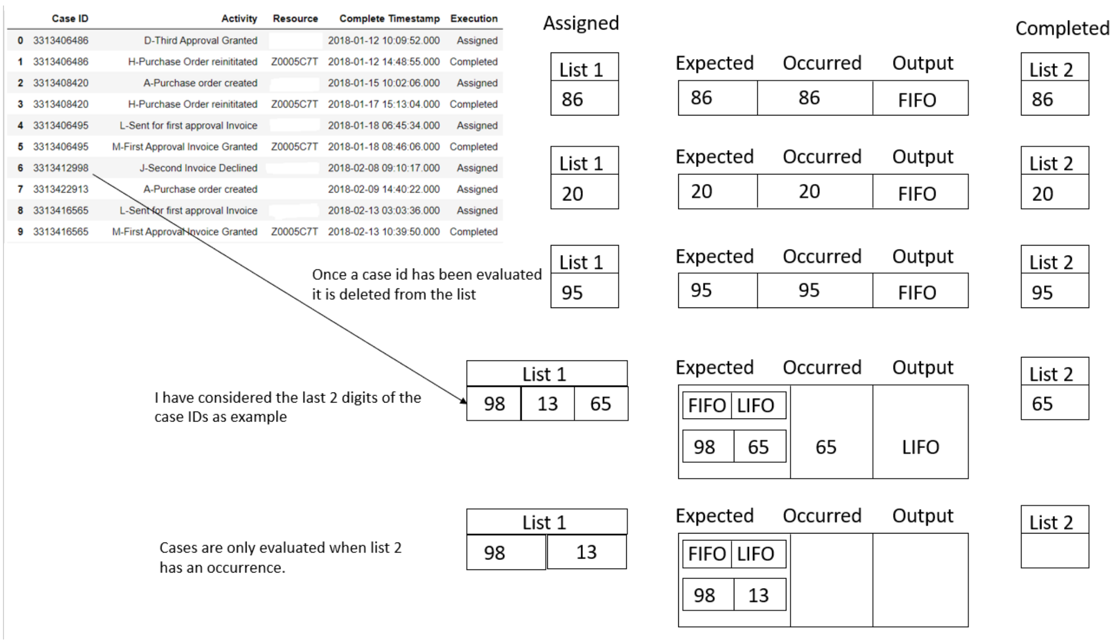


Figure 4.5: Approach for Identifying the Prioritization Pattern

By employing the aforementioned approach, we apply this method for all resources in our event log and determine their working patterns. Figure 4.6 displays an example of the outcome we obtain after implementing the above approach for a list of resources.

Resource	Task Completed	FIFO	LIFO	Random
Z0029XZJ	2894	297	2500	97
Z0003PCQ	2583	46	1187	1350
Z0005X7J	1699	918	456	325
Z0000DLC	1345	213	803	329
Z001NEJE	1273	12	591	670
Z002NUJN	1175	6	543	626
Z00063OV	1099	15	631	453
Z000NHPZ	1069	619	230	220
Z0000D8W	1003	2	534	467
Z0000E9B	864	256	312	296

Figure 4.6: Resource Working Behaviour - Prioritization Pattern

4.3.3 Evaluation of the Prioritization Pattern

To determine the success of an instance, we must establish a set of `KPIs`. The KPIs can aid us in evaluating whether an output is favorable or unfavorable. With knowledge of both types of output, we can identify the best prioritization pattern for the resource. In order to do that, we will make two subsets of our original log. Cases that meet the KPI criteria will be included in the *good log*, while cases that do not meet the KPI criteria will be included in the *bad log*.

Now, in order to evaluate the obtained `prioritization patterns`, we can utilize the lists of cases we maintained in our above approach. The lists of case IDs were generated during the assessment of the working behavior to identify the cases that were prioritized in a specific pattern. We can use these lists to search for the case IDs in the *good log* and *bad log*. The good and bad logs are obtained on the basis of a `KPI` as explained above. For instance, if we consider a case ID from the FIFO list, we will check its appearance in the logs. If it appears in the good log, it is counted as a *"Good FIFO"*, and if it appears in the bad log, it is counted as a *"Bad FIFO"*. Figure 4.7 shows an example where we used our approach to evaluate the prioritization patterns of a set of resources. The KPI used to obtain these results is `Due Date Compliance`.

4.3.4 Calculating Success Rates

To identify the most effective working behavior, we will evaluate the `success rates` associated with different work prioritization patterns. By analyzing these success rates, we can determine which pattern yields the highest rate, indicating the optimal behavior to adopt. The success rate of a specific pattern can be determined by dividing the number of *"Good"* cases that correspond to that pattern by the total number of cases. This calculation

Resource	Task Completed	FIFO	Good_FIFO	Bad_FIFO	LIFO	Good_LIFO	Bad_LIFO	Random	Good_Random	Bad_Random
Z0029XZJ	2894	297	262	35	2500	2160	340	97	93	4
Z0003PCQ	2583	46	44	2	1187	953	234	1350	1261	89
Z0005X7J	1699	918	895	23	456	449	7	325	322	3
Z0000DLC	1345	213	195	18	803	745	58	329	300	29
Z001NEJE	1273	12	11	1	591	544	47	670	607	63
Z002NUJN	1175	6	2	4	543	480	63	626	537	89
Z000630V	1099	15	13	2	631	567	64	453	373	80
Z000NHPZ	1069	619	569	50	230	225	5	220	218	2
Z0000D8W	1003	2	2	0	534	477	57	467	435	32
Z0000E9B	864	256	226	30	312	274	38	296	281	15

Figure 4.7: Evaluated Working Behaviour - Prioritization Pattern

provides us with a percentage that represents the success rate for that particular pattern. After calculating the success rates for each pattern, we compare them to find the pattern with the highest success rate. This pattern is considered the optimal behavior because it demonstrates the highest level of success among the options evaluated.

To provide a visual representation, an example of success rates for a set of resources has been displayed in Figure 4.8. This illustration showcases the success rates associated with different work prioritization patterns, allowing for easy comparison and identification of the pattern with the highest success rate. The KPI used to obtain these results is `Due Date Compliance`.

By utilizing this analysis, we can make informed decisions about the most effective working behavior to optimize process performance and achieve greater success in the given context.

4.4 Exploring the Relationship between Workload and Process Instance Success

`Workload` is a critical factor that influences the working behavior of individuals. While it is commonly believed that a high workload leads to reduced success, the relationship between workload and success is not well understood. In many industries and fields, individuals face significant workload demands, and understanding how to manage workload effectively is crucial for achieving success. Therefore, this study aims to explore the relationship between workload and success, with a particular focus on how workload impacts the success rate of the process instance. By examining this relationship, we hope to provide insights that can help individuals and organizations better manage workload and improve overall success.

Resource	Task Completed	Success_Rate_FIFO	Success_Rate_LIFO	Success_Rate_Random
Z0029XZJ	2894	0.8821548821548821	0.864	0.9587628865979382
Z0003PCQ	2583	0.9565217391304348	0.8028643639427128	0.934074074074074 1
Z0005X7J	1699	0.9749455337690632	0.9846491228070176	0.9907692307692307
Z0000DLC	1345	0.9154929577464789	0.9277708592777086	0.9118541033434651
Z001NEJE	1273	0.9166666666666666	0.9204737732656514	0.9059701492537313
Z002NUJN	1175	0.3333333333333333	0.8839779005524862	0.8578274760383386
Z00063OV	1099	0.8666666666666667	0.8985736925515055	0.8233995584988962
Z000NHPZ	1069	0.9192245557350566	0.9782608695652174	0.990909090909091
Z0000D8W	1003	1.0	0.8932584269662921	0.9314775160599572
Z0000E9B	864	0.8828125	0.8782051282051282	0.9493243243243243

Figure 4.8: Success Rate - Prioritization Pattern

4.4.1 Defining Workload and Metrics

`Workload` is a significant factor that can influence the working behavior of a resource in a process. It encompasses the volume of work assigned to an individual, which they are responsible for completing. To better understand the relationship between workload and success, we will measure workload based on the number of events performed in a month, quarter, and year. These metrics will help us track the amount of work completed by individuals over different time frames and enable us to assess consistency across these periods.

We will measure workload using `three` key metrics:

- **Number of tasks performed per month**: This metric will allow us to gauge the workload on a `monthly` basis and evaluate the level of productivity achieved within that time frame.
- **Number of tasks performed per quarter**: By examining workload at the quarterly level, we can observe work patterns and identify any variations or trends that may arise over a `three-month period`.
- **Number of tasks performed per year**: This metric will provide a broader view of workload and allow us to assess long-term productivity. It will help us understand how workload fluctuates over the course of a `year` and identify any seasonal or cyclical patterns that may exist.

By utilizing these workload metrics across different time intervals, we aim to gain a comprehensive understanding of how workloads impact success and determine if there are any consistent patterns or correlations between workload and success of a process instance.

4.4.2 Identifying Resource Workload across Different Metrics

The study utilizes three different metrics `year,quarter,month` - to calculate the workload of a resource. The analysis begins by identifying the minimum and maximum timestamps present in the event log, which determine the ranges for months, years, and quarters. Next, the evaluation of each resource is conducted individually. This involves extracting all activities performed by the resource under consideration, each associated with a specific timestamp.To determine the workload within a given time range (e.g., per month, per year, per quarter), the number of activities performed by the resource is counted. This counting process involves tallying the activities that fall within the defined time range for the corresponding metric. For example, when assessing workload per month, the study identifies the start and end timestamps for each month within the data-set. The activities performed by the resource within each month are examined, and the count of activities falling within that specific month's time range is calculated. The same process is repeated for workload evaluation per year and per quarter. The ranges for each year and quarter are determined based on the minimum and maximum timestamps in the event log. Then, the activities performed by the resource are evaluated within the defined time ranges for years and quarters, respectively, to calculate the corresponding workload. By analyzing the workload of each resource across different time intervals, this approach provides insights into the distribution and volume of tasks performed, enabling a comprehensive assessment of the resource's workload and its potential impact on success.

Figure 4.9 showcases the workload of a group of resources over various years. In the chart, the presence of a *"NaN"* value indicates that a particular resource did not engage in any purchase order-related tasks during that specific year. By observing this visualization, we can discern the workload distribution among the resources over time. The *"NaN"* values help us identify periods in which certain resources were not involved in purchase order activities, providing a comprehensive overview of their level of contribution throughout the years.

4.4.3 Evaluating Resource Workload across Different Metrics

To evaluate the workload of resources, we employed a meticulous approach based on key performance indicators `KPIs` and the classification of event logs into two subsets: *good log* and *bad log*. The division was determined by evaluating the KPI for each case ID. If a case ID exhibited a successful KPI outcome, it was classified as part of the good log; otherwise, it was classified as part of the bad log. Building upon the previously identified workloads of resources, we further examined the associated case IDs of the tasks performed across different time frames. Tasks linked to case IDs within the good log were deemed *successful*, while those associated with the bad log were considered *unsuccessful*. For example, let's consider a resource that performed 60 tasks in the month of January. Out of these 60 tasks, if 55 of them had case IDs that were part of the good log, we would determine that the resource successfully performed 55 out of the 60 tasks in January. To gain a comprehensive understanding of resource working behavior on the performanc eof the process, we conducted the evaluation process for each resource, taking into account `workload per month, year, and quarter`. This meticulous analysis involved examining the tasks performed by each resource and their associated case IDs within these specific time frames, considering relevant Key Performance Indicators (KPIs). By conducting this evaluation across different time frames, we obtained valuable insights into the number

Resource	2018	2019	2020	2021	2022
Z001NEJE	739	166	6	0	nan
Z0005C7T	559	112	51	30	nan
Z000DLC	459	144	57	30	nan
Z0003PCQ	402	221	22	3	nan
Z00630V	412	108	23	7	nan
Z002NUJN	393	150	2	0	nan
Z0001M0E	245	113	40	24	nan
Z000D8W	237	96	32	33	nan
Z0029XZJ	124	127	66	39	nan
ZZZZZ5C9	303	39	1	1	nan

Figure 4.9: Resource Working Behaviour - Workload

of successful cases achieved by each resource. This approach allowed us to identify patterns and trends in resource behavior over time, providing a deeper understanding of their contributions to overall process success. This evaluation serves as a foundation for identifying resource-specific patterns, optimizing workload allocation, and ultimately enhancing overall process efficiency and success.

Figure 4.10 depicts the evaluated `workloads` of a group of resources over different years. The adjacent success column for each year represents the number of tasks performed by the resources that align with the good log, based on the `Due Date Compliance` KPI. By examining this representation, we can gain insights into the distribution of workloads among the resources across multiple years. Furthermore, the success column allows us to gauge the effectiveness of the resources in meeting the criteria defined by the Due Date Compliance KPI.

4.4.4 Calculating Success Rate across Different Metrics

To measure the `success rates` of resources per year, month, and quarter, a straightforward calculation was employed. The total number of successful tasks performed within each respective time frame was divided by the total number of tasks performed during that period. This calculation provided a percentage that represented the success rate of each resource within the specific year, month, or quarter. For instance, to determine the success rate of a resource in a given month, we summed up the total number of successful tasks performed by that resource during that month. Then, we divided this sum by the total number of tasks performed by the resource in the same month. The resulting quotient represented the success rate for that particular resource in that specific month. The same calculation method was applied for evaluating success rates per year and per

Resource	2018	2018 success	2019	2019 success	2020	2020 success	2021	2021 success
Z001NEJE	739	675.0	166	147.0	6	4.0	0	0.0
Z0005C7T	559	419.0	112	95.0	51	45.0	30	27.0
Z0000DLC	459	412.0	144	129.0	57	55.0	30	28.0
Z0003PCQ	402	313.0	221	179.0	22	19.0	3	2.0
Z000630V	412	330.0	108	90.0	23	22.0	7	7.0
Z002NUJN	393	312.0	150	128.0	2	2.0	0	0.0
Z0001M0E	245	203.0	113	97.0	40	35.0	24	24.0
Z0000D8W	237	203.0	96	86.0	32	29.0	33	31.0
Z0029XZJ	124	101.0	127	106.0	66	57.0	39	36.0
ZZZZZ5C9	303	244.0	39	28.0	1	1.0	1	1.0

Figure 4.10: Evaluated Working Behaviour - Workload

quarter. By utilizing this calculation approach across different metrics, we obtained a comprehensive understanding of the success rates achieved by each resource throughout various time periods. These success rate metrics serve as valuable indicators for evaluating resource behavior and assessing the effectiveness of workload allocation strategies.

Figure 5.16 illustrates the `success ratio` of a group of resources across different years. A higher success ratio indicates positive impact on the performance of the process during those specific years. By analyzing this ratio, we can determine the relative effectiveness of each resource over time and identify the years in which they performed exceptionally well. The KPI used to obtain these results is `Due Date Compliance`.

4.4.5 Identifying the Optimal Workload for Resources

To determine the `optimal workload` for each resource, we employ a methodology that involves calculating the `mean success rate per workload range`. The goal is to identify the workload range that yields the highest mean success rate, indicating the most favorable level of work for each resource. We establish 10 distinct workload ranges, starting from 0-10 and extending up to 100 or beyond. These ranges are designed to cover a broad spectrum of workloads that resources may encounter. For each resource and corresponding time period, we collect data on the number of activities performed and their associated success rates. Let's consider *Resource A* as an example. In `February`, *Resource A* completed `22` activities with a success rate of `0.79`, in `March`, `28` activities with a success rate of `0.64`, in `April`, `46` activities with a success rate of `0.94`, and in `May`, `98` activities with a success rate of `0.74`. Using this data, we calculate the mean success rate for each workload range. For instance, if the workload range is set to `21-30` activities, we compute the mean success rate by averaging the success rates of *Resource* A for the corresponding time period within that range. In this case, the mean success rate for

Resource	2018	2018 success	2018 success ratio	2019	2019 success	2019 success ratio	2020	2020 success	2020 success ratio
Z001NEJE	739	675.0	0.9133964817320703	166	147.0	0.8855421686746988	6	4.0	0.6666666666666666
Z0005C7T	559	419.0	0.7495527728085868	112	95.0	0.8482142857142857	51	45.0	0.8823529411764706
Z0000DLC	459	412.0	0.8976034858387799	144	129.0	0.8958333333333334	57	55.0	0.9649122807017544
Z0003PCQ	402	313.0	0.7786069651741293	221	179.0	0.8099547511312217	22	19.0	0.8636363636363636
Z000630V	412	330.0	0.8009708737864077	108	90.0	0.8333333333333334	23	22.0	0.9565217391304348
Z002NUJN	393	312.0	0.7938931297709924	150	128.0	0.8533333333333334	2	2.0	1.0
Z0001M0E	245	203.0	0.8285714285714286	113	97.0	0.8584070796460177	40	35.0	0.875
Z0000D8W	237	203.0	0.8565400843881856	96	86.0	0.8958333333333334	32	29.0	0.90625
Z0029XZJ	124	101.0	0.8145161290322581	127	106.0	0.8346456692913385	66	57.0	0.8636363636363636
ZZZZZ5C9	303	244.0	0.8052805280528053	39	28.0	0.717948717948718	1	1.0	1.0

Figure 4.11: Success Rate - Workload

the `21-30` workload range would be `(0.79 + 0.64) / 2 = 0.715`. By performing this calculation for each workload range and resource, we obtain a singular mean success rate value for each range. The optimal workload range for each resource is then determined by selecting the range with the **`highest mean success rate`**. This range represents the level of workload that maximizes the resource's potential and performance of the process. Identifying the optimal workload for resources is crucial for achieving an optimal balance between workload allocation and process performance. It allows us to effectively leverage the resources' capabilities, avoiding under-utilization or overburdening. By selecting the workload range with the highest mean success rate, we can optimize resource allocation and enhance the overall success and efficiency of the process.

Figure 4.12 presents the success rate of a group of resources across different workload ranges. This visualization allows for easy identification of the `optimal workload` range for each resource by identifying the workload range associated with the highest success rate. By examining the corresponding workload range, we can determine the level of tasks or workload at which each resource performs most effectively, leading to a higher rate of success.

4.5 Exploring the Relationship between Batching and Process Instance Success

The efficient execution of batches is crucial for many organizations that handle large volumes of data or have frequent repetitive tasks.To achieve optimal results in any process, it is crucial to understand the impact of `batch executions` on process performance. By examining batch executions, we can determine whether they contribute positively or negatively to the success of the process, and how they can be optimized for better process performance.Therefore, this study aims to explore the impact of batch executions by resources on process performance and identify ways to improve the process by analyzing and optimizing batch executions.

Resource	1-20	21-40	41-60	61-80	81-100
Z0003PCQ	0.3501812489701763	0.8571428571428571	0.8846691775692701	0.8820071742786935	0.8974234907533807
Z0029XZJ	0.8643555328400528	0.9203720030925915	0.915537151004049	0.9126870990734142	nan
Z002NUJN	0.9296398046398047	0.8	0.875	0.8609035409035409	0.843070393658629
Z001NEJE	0.7986394557823129	0.9285714285714286	0.9301108374384237	0.9210498128631932	0.9046233127397301
Z0000DLC	0.9581336448209512	0.9128535813757491	0.8912806739181689	0.909197651663405	0.9302325581395349
Z0005C7T	0.7867016617016619	0.8381505270092227	0.7627981576910504	0.7760837502496853	nan
Z0027ARY	0.967361111111112	0.9391217034599387	0.8887016704295738	0.8871369152298092	0.9394651777843079
Z0000D8W	0.9105005413177159	0.9127518923571556	0.9525219986333654	0.8767123287671232	nan
Z00630V	0.9881288156288156	0.8658736279037031	0.8178616352201259	0.7960127931769724	0.8148148148148148
Z0005X7J	0.9757061157796452	0.9562012975557868	0.9695045695045694	nan	nan
Z000NHPZ	0.9284043328160974	0.9551093812296477	nan	nan	nan
ZZZZZ5C9	0.9066666666666666	0.7280604314113769	0.7620232475111387	0.7810826428473487	0.7777777777777778

Figure 4.12: Success Rate for Different Workload Ranges

4.5.1 Defining Batching and Metrics

`Batching` in event logs refers to the practice of executing activities in groups or batches rather than individually. It involves grouping multiple activities together and performing them simultaneously, which can improve efficiency and streamline processes. In this study, we explore different types of batching and define metrics to identify them.

We define `three` metrics to identify different types of batching in event logs:

- **Activity-based Batching**: This metric focuses on the consecutive execution of the same activity by the same resource for different case IDs. When a resource performs multiple instances of the same activity in succession, it indicates `activity-based batching`. This suggests that the resource is performing similar activities as a batch, potentially optimizing their workflow and improving efficiency.
- **Time-based Batching**: This metric examines the time intervals between activities performed by a resource. If a resource executes multiple activities within a specific time window, such as five minutes, it indicates `time-based batching`. This type of batching implies that the resource is grouping activities together and implementing them in a group, potentially to leverage shared resources or maximize productivity within a given time frame.
- **Size-based Batching**: This metric considers the number of activities performed by a resource within a certain time period. If a resource performs tasks exceeding a predefined threshold, it indicates `size-based batching`. This suggests that the resource is executing activities in larger batches, potentially to take advantage of economies of scale, reduce setup time, or optimize resource allocation.

By applying these metrics to the event logs, we aim to identify different types of batching and gain insights into how batching practices impact process efficiency and resource utilization. The analysis of batching can provide valuable information for process optimization, workload management, and resource allocation strategies.

4.5.2 Identifying Batch Executions by Resources across Different Metrics

In order to identify batch executions performed by a resource in an event log, we employ three metrics: activity-based batching, time-based batching, and size-based batching.

To identify batches within the activity sequence for `activity-based batching`, we carefully analyze the consecutive activities performed by a resource. The process commences by examining the first activity in the sequence. We then proceed to compare each subsequent activity to the current one. If they are identical and being performed for a different case ID, we identify these activities as part of a batch. This indicates that the resource is executing a series of similar activities for different cases, which signifies a batching scenario. When encountering a different activity, we start considering the possibility of a new batch. Subsequently, we evaluate the next activities to determine if they are the same as the current one and being performed for a different case ID. If this condition holds true, we mark these activities as part of the ongoing batch. However, if we observe a different subsequent activity while contemplating the possibility of a new batch, we classify it as a single activity. This means that the previous batch has ended, and the resource has moved on to executing a distinct task. By following this logic, we can effectively identify and track batches within the resource's activity sequence, enabling us to gain insights into how tasks are grouped and processed for improved process performance.

In the case of `time-based batching`, we employ a predefined time threshold to determine the temporal proximity of activities. For instance, let's consider a batch time limit of 5 minutes. The evaluation process begins by analyzing the first activity performed by the resource. If another activity occurs within the specified time threshold, we classify both the current and next activities as part of a batch. We then proceed to evaluate the next activity in the sequence. However, if no activity occurs within the defined time threshold, the current activity is marked as a single execution. This indicates that there is a gap between activities beyond the specified time limit, suggesting that the current task was performed independently. We continue this evaluation process for each subsequent activity, applying the same logic. As a result, every activity is identified as either part of a batch or as a single execution based on their temporal proximity. By using this approach, we can effectively identify batching patterns by considering the time intervals between activities. This metric provides insights into how resources group and execute tasks within a specific time frame, aiding in the analysis of temporal dependencies and optimizing process efficiency.

`Size-based batching` is an enhancement of time-based batching that considers both time and activity count thresholds. In addition to tracking the time interval, such as 5 minutes, we also monitor the number of activities executed by a resource within that time window. For example, let's assume a size threshold of 10 activities. We evaluate each activity and determine if there are 10 subsequent activities occurring within the specified time frame. If this condition is met, we classify all these activities as executed in a batch. By incorporating both time and activity count thresholds, we gain a deeper understanding of the batching patterns employed by resources. This approach allows us to identify and analyze different types of batching based on the number of activities performed within a given time frame. By applying these metrics, we gain insights into the occurrence of different types of batching executed by resources. Each metric provides a unique perspective on how activities are grouped and executed as batches. Understanding these batching patterns helps us analyze process efficiency, resource utilization, and potential areas for optimization.

Figure 4.13 displays the total number of activities performed by a resource, as well as the count of activities that were executed as part of a batch. This visualization provides a clear overview of the resource's activity level and the extent to which batching was utilized. By examining this figure, we can observe the overall volume of activities performed by the resource. The comparison between the total activity count and the number of activities executed in batches offers insights into the resource's batching behavior and its impact on process efficiency.

Resource	Total_Executions	Total_Batch_Execution
Z002NUJN	8914	6962
Z003MCEP	6417	5353
Z0003PCQ	7304	4871
Z0027ARY	5261	4329
Z0000D8W	4799	3346
Z0032Y7N	3807	3252
Z000IB3U	3508	3148
Z0005X7J	3890	2636
Z000M6EM	2750	2587
Z0016C3J	5818	2356
Z000SK7E	4250	2146
Z0016B5E	2713	2018
ZZZZZ5C9	3182	2006

Figure 4.13: Resource Working Behavior - Batching

4.5.3 Evaluating Batch Executions by Resources across Different Metrics

In our evaluation of batching implemented by the resources, we adopted a comprehensive approach that involved analyzing key performance indicators `KPIs` and classifying event logs into two distinct subsets: *good log* and *bad log*. This classification was based on the KPI outcomes observed for each case ID. Case IDs that demonstrated successful KPI results were categorized as part of the *good log*, while those with unfavorable outcomes were assigned to the *bad log*. Expanding on the initial identification of batching within the resources, we delved deeper into the tasks performed in batches and their associated case IDs. By examining the case IDs linked to tasks executed as batches within the good log, we considered these tasks as *successful* instances of batching. Conversely, tasks associated with case IDs in the bad log were regarded as *unsuccessful* cases of batching. This evaluation process was conducted for each resource, taking into account different types of batching, including `activity batching`, `time-based batching`, and `size-based batching`. Through meticulous analysis of the tasks performed as batches and their corresponding case IDs, we were able to gain valuable insights into the count of successful and unsuccessful cases based on each type of batching adopted by the resources. By utilizing this detailed methodology and examining the relationship between batching and count of good and bad batching, we aimed to provide a comprehensive understanding of how different types of batching impact the overall process performance. These insights can then be used to optimize and improve the resource allocation strategies, ultimately enhancing the efficiency and success of the process.

Figure 5.11 illustrates the evaluation of activity-based batching employed by the resources. The figure presents two categories: *"Good Batching"* and *"Bad Batching"*, which reflect the outcomes of batch executions based on the chosen Key Performance Indicator (KPI).The KPI used to obtain these results is `Due Date Comliance`. The *"Good Batching"* category represents the count of case IDs for which activities executed in batches resulted in a positive outcome according to the KPI. These batch executions contributed to successful cases or desired process outcomes. On the other hand, the *"Bad Batching"* category represents the count of case IDs where activities executed in batches resulted in unfavorable outcomes based on the KPI. These batch executions are associated with cases that did not meet the desired performance criteria or encountered issues. By analyzing the data presented in this figure, we can gain insights into the effectiveness of activity-based batching. It allows us to understand the impact of batch execution on overall process performance and identify patterns where batching strategies contribute to positive or negative outcomes.

4.5.4 Calculating Success Rate across Different Metrics

To assess the `success rates` of the batching implemented by the resources using the three different definitions `activity-based, time-based, and size-based`, we utilized a straightforward calculation. We divided the total number of good batches by the total number of tasks performed as batches by that resource. This calculation resulted in a percentage that represented the success rate of batching for each resource and for each specific metric. By employing this methodology, we were able to quantify and compare the success rates of different types of batching implemented by the resources. This provided valuable insights into the effectiveness and efficiency of each batching approach, enabling us to assess resource behavior and optimize batching strategies accordingly.

Figure 5.12 showcases the success rate of a set of resources based on the batching strategies they implemented. The success rate serves as an indicator of how well the applied batching techniques contributed to favorable outcomes for these resources. A higher success rate suggests that the implemented batching strategies positively influenced the overall

Resource	Total_Executions	Total_Batch_Execution	Good_Batching	Bad_Batching
Z002NUJN	8914	6962	5864	1098
Z003MCEP	6417	5353	5347	6
Z0003PCQ	7304	4871	4458	413
Z0027ARY	5261	4329	3778	551
Z000D8W	4799	3346	2949	397
Z0032Y7N	3807	3252	1983	1269
Z000IB3U	3508	3148	2765	383
Z0005X7J	3890	2636	2590	46
Z000M6EM	2750	2587	2111	476
Z0016C3J	5818	2356	1308	1048
Z000SK7E	4250	2146	2063	83
Z0016B5E	2713	2018	1862	156
ZZZZZ5C9	3182	2006	1197	809

Figure 4.14: Evaluated Working Behavior - Batching

behavior and effectiveness of the resources. It indicates that the grouping and execution of activities in batches resulted in successful cases or desired process outcomes.The KPI used to obtain these results is `Due Date Compliance`. By analyzing the data presented in this figure, we can gain insights into the correlation between different batching approaches and the success rates achieved by the resources. This information is valuable for evaluating and optimizing batching strategies, identifying best practices, and guiding resource behavior to enhance overall process performance.

Resource	Total_Executions	Total_Batch_Execution	Good_Batching	Bad_Batching	Success_Rate
Z002NUJN	8914	6962	5864	1098	0.8422866992243608
Z003MCEP	6417	5353	5347	6	0.9988791331963385
Z0003PCQ	7304	4871	4458	413	0.915212482036542 8
Z0027ARY	5261	4329	3778	551	0.8727188727188727
Z0000D8W	4799	3346	2949	397	0.8813508667065152
Z0032Y7N	3807	3252	1983	1269	0.6097785977859779
Z000IB3U	3508	3148	2765	383	0.8783354510800508
Z0005X7J	3890	2636	2590	46	0.9825493171471927
Z000M6EM	2750	2587	2111	476	0.8160030923850019
Z0016C3J	5818	2356	1308	1048	0.5551782682512734
Z000SK7E	4250	2146	2063	83	0.9613233923578751
Z0016B5E	2713	2018	1862	156	0.9226957383548068
ZZZZZ5C9	3182	2006	1197	809	0.5967098703888335

Figure 4.15: Success Rate - Batching

Chapter 5

Implementation

5.1 Resource Behavior Analyzer

In this section, we will discuss the implementation of the tool developed for analyzing event logs by using key performance indicators (KPIs) to identify the relation between resource working behavior and success of the process [1]. The tool is built using the `Django framework` with `Python` as the programming language. The front-end interface is designed using `HTML, CSS`, and `jQuery`, providing a user-friendly and interactive experience. Upon accessing the tool, the user is presented with a `homepage` that offers two main options. The first option allows the user to choose an `event log`, which serves as the input for the analysis. The second option is a drop down menu that provides a `list of available KPIs`. The different types of event logs and KPIs have been explained in the section 4.2. The user can select a specific KPI of interest from the drop down menu. Once the input is inserted, the user can proceed by clicking the submit button, initiating the code execution.

5.1.1 Home Page

Figure 5.1 showcases the home view of our implemented tool, which provides a user-friendly interface. The first drop down menu is accompanied by descriptive text indicating its purpose: to *"Upload a CSV file"* containing the event log data. Similarly, the second drop down menu presents a list of KPIs, accompanied by the text *"Select a KPI"*, allowing users to choose the specific performance indicator they wish to analyze. In the center of the interface, clear instructions are provided, reiterating the process: users need to upload the event log and select a KPI from the drop down menu. Upon clicking the submit button, the framework will be implemented, processing the chosen data and generating the desired results. This intuitive design aims to streamline the user experience and facilitate efficient analysis of resource performance.

5.1.2 Dashboard

Once the analysis is completed, the tool presents a comprehensive and user-friendly dashboard, offering a range of functionalities. Positioned on the left side of the dash-

[1] `https://github.com/faizanhassan94/Thesis`

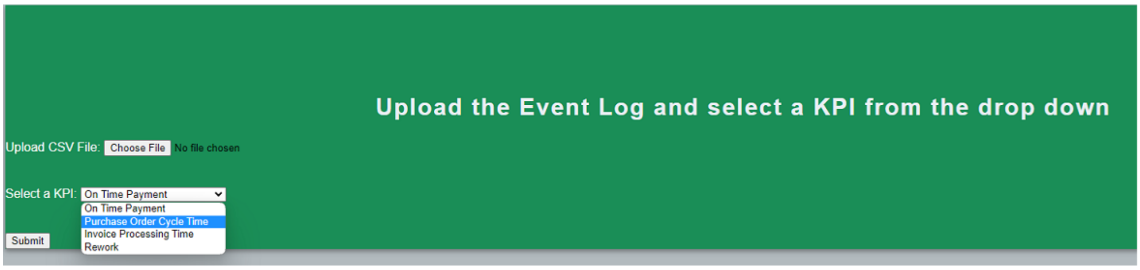


Figure 5.1: Home Page

board, three buttons correspond to the three distinct working behavior patterns implemented in our framework: `the work prioritization pattern, batching pattern`, and `workload pattern`. By simply clicking on the button representing their chosen approach, users can view the results specific to that pattern.

Figure 5.2 illustrates the output dashboard, showcasing a welcoming message along with the user's name at the top left corner. Directly underneath, the three buttons representing the different working behavior patterns are displayed. At this point in the view, the right side of the dashboard remains empty, indicating that no specific pattern has been selected yet. The clear layout and intuitive design of the dashboard aim to facilitate user interaction and make it easy to access and interpret the results generated by the tool.

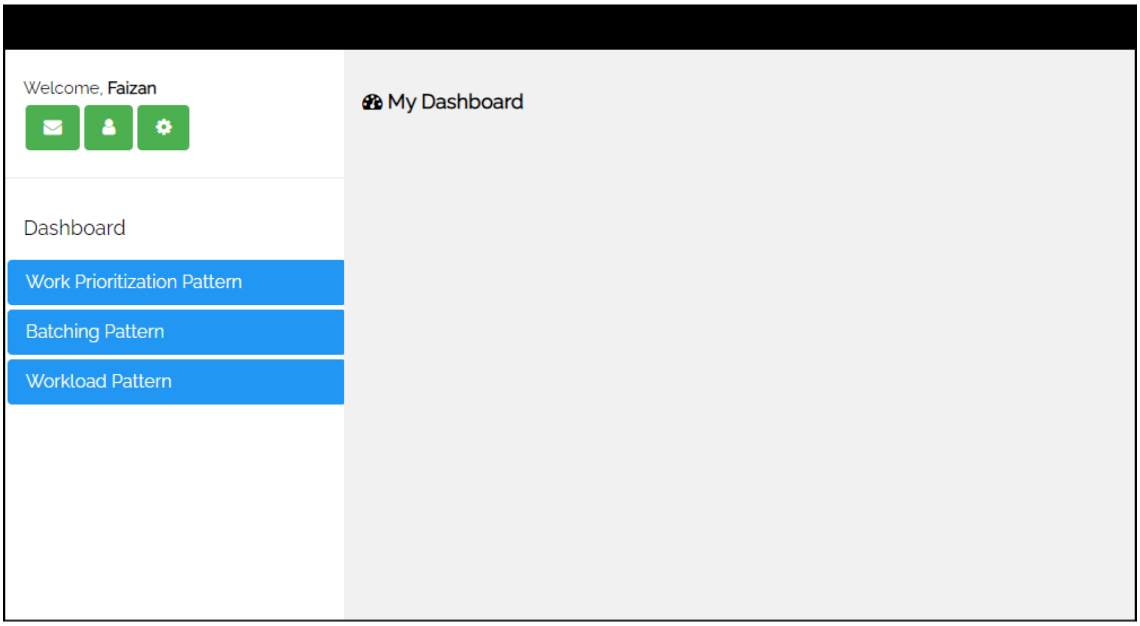


Figure 5.2: Dashboard

Upon selecting a pattern button, the tool opens a dedicated view tailored to that specific pattern. On the right side of the dashboard, `three` additional buttons are presented, offering further insights and analysis options:

- **Working Behavior:** Clicking this button provides users with an overview of the observed working behavior of the resources, derived from the event log data. This view offers a high-level understanding of how the resources have performed in terms of their tasks and activities. This has been explained in detail for each aspect i.e. work prioritization pattern, workload pattern, batching pattern in section 4.3.2 4.4.2 4.5.2 respectively.

- **Working Behavior Evaluated:** This button leads to a comprehensive evaluation of the working behavior patterns. It delves into the details, examining individual cases and highlighting both successful and unsuccessful instances. This evaluation aids in identifying patterns, trends, and potential areas of improvement or concern. This has been explained in detail for each aspect i.e. work prioritization pattern, workload pattern, batching pattern in section 4.3.3 4.4.3 4.5.3 respectively.
- **Success Rate:** Clicking this button reveals the success rate of the resource working behavior. This metric provides an objective measure of the effectiveness of their performance. It quantifies the percentage of cases or tasks that have been successfully executed, enabling users to gauge the overall efficiency and reliability of the resources. This has been explained in detail for each aspect i.e. work prioritization pattern, workload pattern, batching pattern in section 4.3.4 4.4.4 4.5.4 respectively.

5.1.3 Buttons and Drop Downs

These three buttons on the right side of the dashboard provide users with focused perspectives and actionable insights, empowering them to make informed decisions and optimize resource management based on the outcomes of the work prioritization pattern analysis.

Figure 5.3 showcases the `three` buttons that appear after selecting any of the analysis patterns. Clicking on each button reveals specific and relevant information related to the chosen pattern. These buttons serve as navigation options for users to access different aspects of the analysis results. By clicking on each button, users can dive deeper into the insights and details pertaining to the particular working behavior pattern they have selected. This intuitive design ensures that users can easily explore and understand the findings within their area of interest.

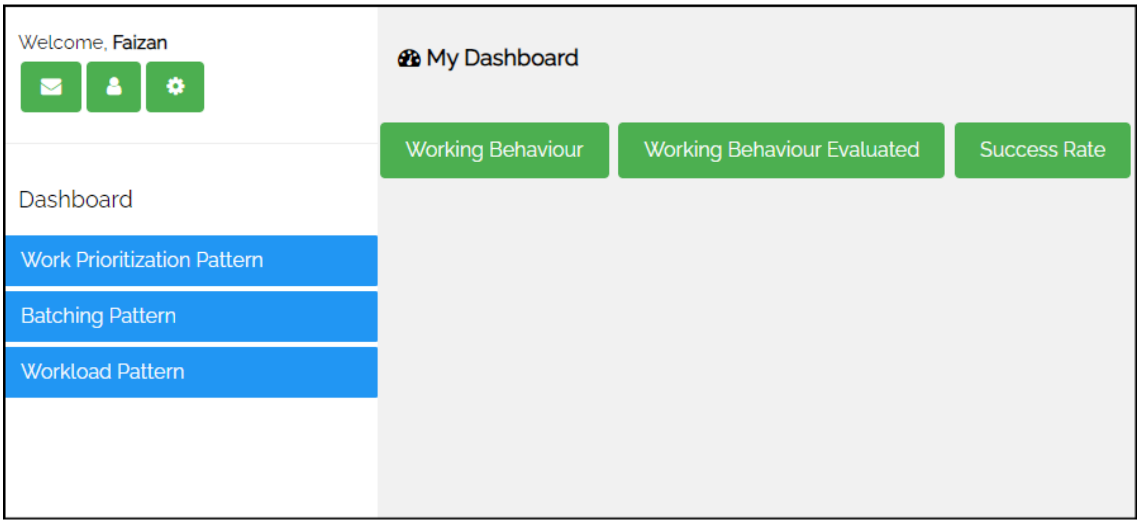


Figure 5.3: Dashboard Buttons

The dashboard features not only buttons but also incorporates `two drop down` menus to enhance user flexibility. The first drop down menu offers a range of view options, including the ability to examine results for `all resources` or selectively focus on the `top 20 resources` based on task performance. The second drop down menu complements the chosen pattern by enabling the selection of a specific `metric`. These drop down menus provide users with additional customization choices, facilitating a comprehensive analysis of resource behavior and enabling them to align their investigation with specific patterns and metrics that align with their objectives.

Figure 5.4 showcases the two drop down menus that facilitate further customization of the analysis results. The first drop down menu, labeled *"Resources"* allows users to choose between viewing the results for either all resources or the top 20 resources. The top 20 resources are determined based on their maximum working behavior within the specific pattern being analyzed. This feature allows users to focus on a subset of resources that exhibit notable performance. The second drop down menu enables users to *"select the evaluation metric"* for the pattern under investigation. This metric serves as a benchmark for assessing the effectiveness and efficiency of the working behavior pattern. Users can choose from a range of available metrics tailored to the specific pattern, providing them with flexibility in evaluating and comparing different aspects of resource performance. These drop down menus enhance the tool's versatility, allowing users to customize their analysis and explore different perspectives based on their specific needs and interests.

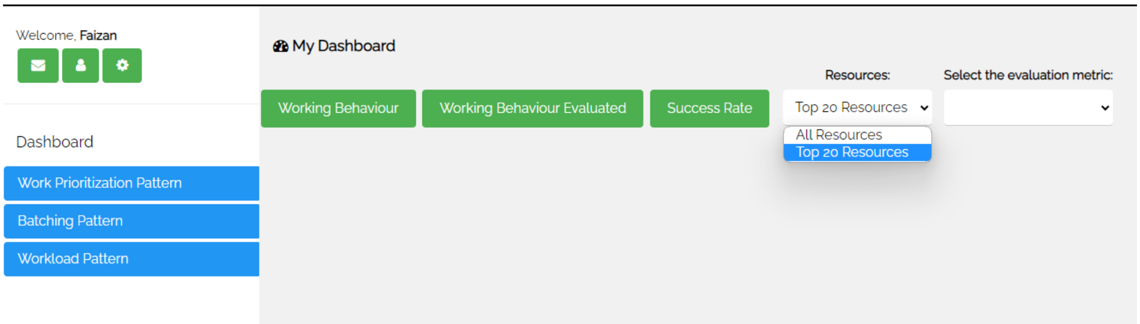


Figure 5.4: Dashboard Drop downs

5.1.4 Work Prioritization Pattern

Within the `work prioritization pattern`, users have the option to explore the different prioritization patterns adopted by the resources. The available metric for this pattern is `FIFO/LIFO/Random` as explained in the section 4.3.1. By selecting this metric, users can analyze how the resources prioritize their tasks based on different approaches: First-In-First-Out (FIFO), Last-In-First-Out (LIFO), or Random order. This insight provides valuable information on the sequencing of tasks and resource decision-making within the work prioritization pattern.

Figure 5.5 showcases the interface after selecting the work prioritization pattern. The drop down menu corresponding to the metric selection displays a single option since we only have one metric available for this pattern. Users can choose the desired metric to explore the work prioritization patterns adopted by the resources. Once the values are selected from the drop down menus, users can simply click on the buttons to generate and view the respective results. This streamlined interface design ensures a user-friendly experience, allowing users to effortlessly navigate and analyze the outcomes of their selected patterns and metrics.

Figure 5.5: Work Prioritization Pattern Dashboard

Figure 5.6 showcases the results of the working behavior analysis based on the prioritization pattern as explained in the section4.3.2. The table within the figure presents the resource names and their corresponding task completion numbers. The columns labeled *"FIFO"*, *"LIFO"*, and *"Random"* indicate the distribution of task prioritization for each resource across the different patterns. The figure provides valuable insights into how the resources prioritize their tasks, allowing for a deeper understanding of their working behavior.

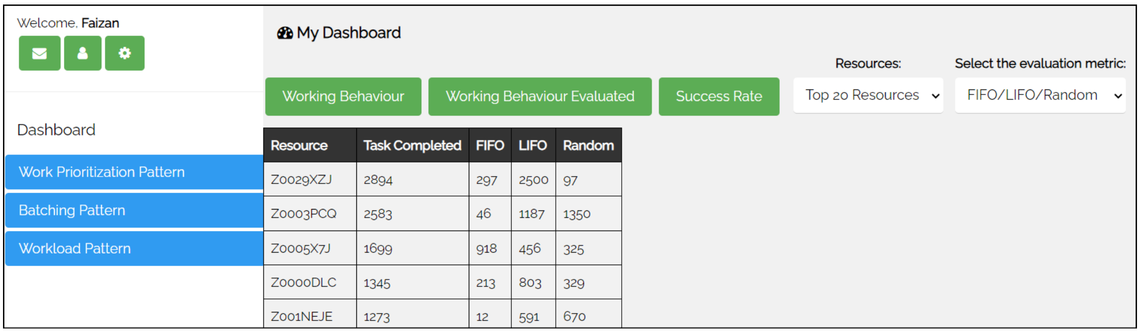
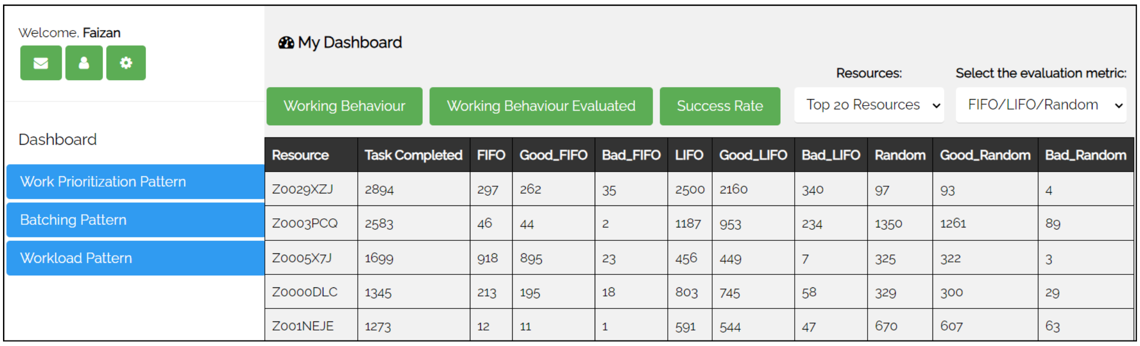


Figure 5.6: Working Behavior - Work Prioritization Pattern

Figure 5.7 displays the evaluated working behavior as explained in the section 4.3.3 based on the selected metric *"FIFO/LIFO/Random"* and the *"Top20 Resources"*. The table within the figure presents the results for each working pattern, including FIFO (Good and Bad), LIFO (Good and Bad), and Random. The columns in the table provide insights into the performance of each working pattern for the evaluated resources. By analyzing these results, valuable information can be obtained regarding the effectiveness and efficiency of different prioritization patterns.

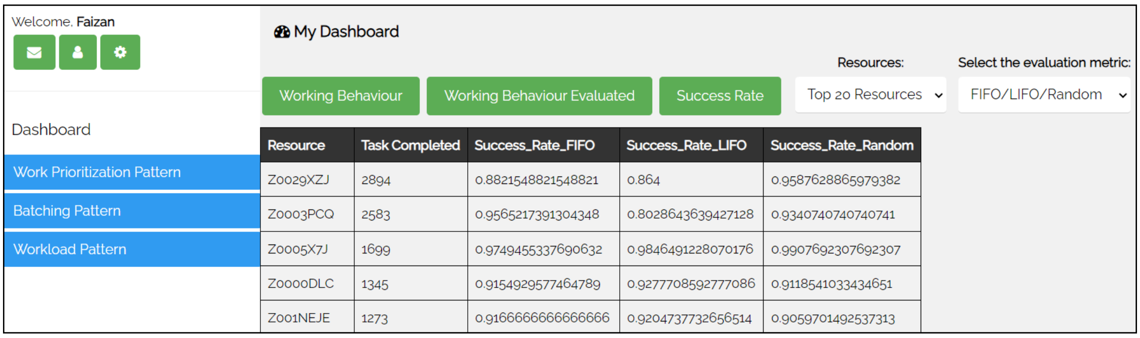
Similarly, Figure 5.8 presents the success rate for each prioritization pattern as explained in the section 4.3.4. The table within the figure provides insights into the performance of each working pattern in terms of the success rate. The columns in the table display the resource, the number of tasks completed by the resource, and the success rate for each pattern, including FIFO, LIFO, and Random. The success rate indicates the percentage of tasks successfully completed for each prioritization pattern. By analyzing these results, we can gain a better understanding of how different working patterns impact the overall



Resource	Task Completed	FIFO	Good_FIFO	Bad_FIFO	LIFO	Good_LIFO	Bad_LIFO	Random	Good_Random	Bad_Random
Z0029XZJ	2894	297	262	35	2500	2160	340	97	93	4
Z0003PCQ	2583	46	44	2	1187	953	234	1350	1261	89
Z0005X7J	1699	918	895	23	456	449	7	325	322	3
Z0000DLC	1345	213	195	18	803	745	58	329	300	29
Z001NEJE	1273	12	11	1	591	544	47	670	607	63

Figure 5.7: Evaluated Working Behavior - Work Prioritization Pattern

success rate of the tasks assigned to resources.



Resource	Task Completed	Success_Rate_FIFO	Success_Rate_LIFO	Success_Rate_Random
Z0029XZJ	2894	0.8821548821548821	0.864	0.9587628865979382
Z0003PCQ	2583	0.9565217391304348	0.8028643639427128	0.9340740740740741
Z0005X7J	1699	0.9749455337690632	0.9846491228070176	0.9907692307692307
Z0000DLC	1345	0.9154929577464789	0.9277708592777086	0.9118541033434651
Z001NEJE	1273	0.9166666666666666	0.9204737732656514	0.9059701492537313

Figure 5.8: Success Rate - Work Prioritization Pattern

5.1.5 Batching Pattern:

The `batching pattern` offers three different available metrics: `activity-based batching`, `size-based batching`, and `time-based batching` as explained in the section 4.5.1. Each metric captures distinct aspects of the batching behavior employed by the resources. Users can choose to explore the results based on these different batching patterns adopted by the resources. By selecting a specific metric from the interface, users can analyze the corresponding batching behavior and its impact on resource performance and efficiency. The interface provides an intuitive way to navigate through the results, allowing users to gain insights into how resources group tasks based on activities, size, or time considerations.

In Figure 5.9, which displays the interface for the batching pattern, users will encounter a drop down menu showcasing the three available metrics: `Activity-Based Batching`, `Size-Based Batching`, and `Time-Based Batching`. Users can select the desired metric from the drop down to explore the corresponding batching behavior of the resources. Similar to the work prioritization pattern, users can click on the buttons on the dashboard to view the results specific to their selected metric within the batching pattern. This user-friendly interface design enables users to effortlessly navigate and analyze the batching behavior and its impact on resource performance and efficiency.

Figure 5.10 showcases the working behavior of resources in terms of batching as explained in the section 4.5.2. The selected metric for this analysis is *"Activity-based Batch"*. The

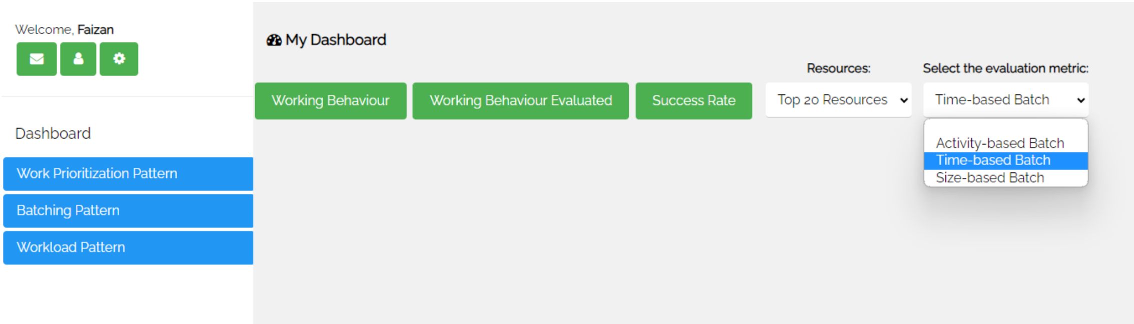


Figure 5.9: Batching Pattern Dashboard

results presented in the figure provide insights into whether the resources performed any activity-based batching. The table within the figure includes columns such as *"Resource"*, *"Total Executions"*, and *"Total Batch Executions"*. These columns represent the resource's involvement in executing activities as part of activity-based batches. By examining these results, we can better understand the working behavior of resources with respect to batching activities.

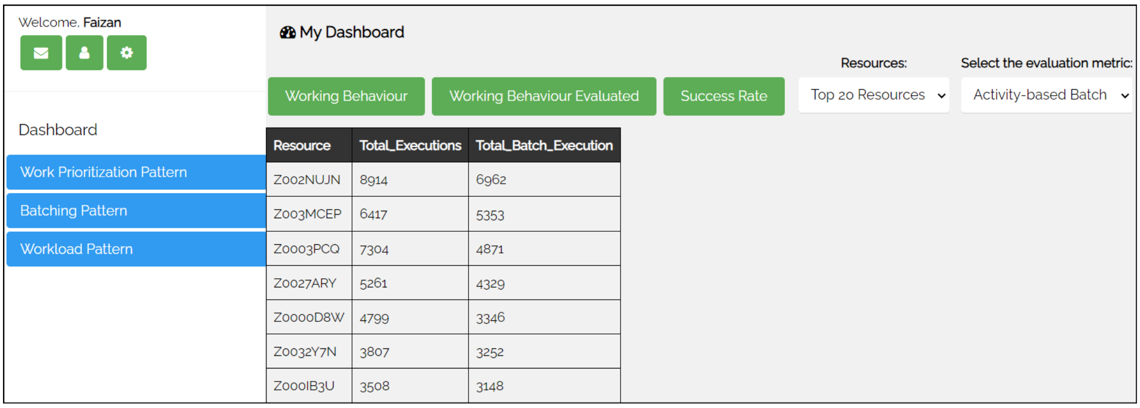
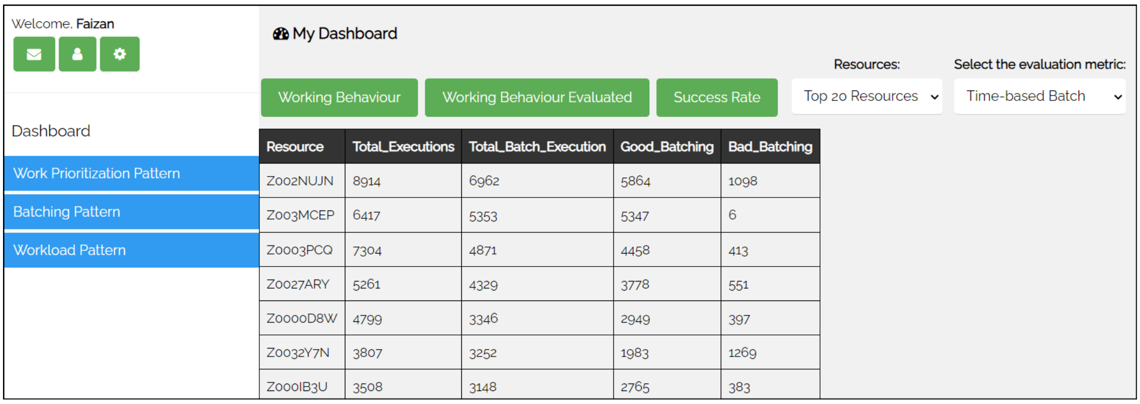


Figure 5.10: Working Behavior - Batching

Figure 5.11 illustrates the evaluated batching behavior as explained in the section 4.5.3, specifically focusing on *Time-based Batch"*. The results shown in the figure correspond to a different metric, with the *"Top20 Resources"* selected from the second drop-down menu. By clicking on *"Working Behavior Evaluated"*, the displayed results provide deeper insights into the batching behavior. Compared to the previous results, two additional columns have been included: *"Good batching"* and *"Bad Batching"*. These columns analyze and evaluate the total batch executions based on a specific KPI, categorizing them as either good or bad. The table within the figure presents these evaluated results, shedding light on the quality of batching behavior exhibited by the resources.

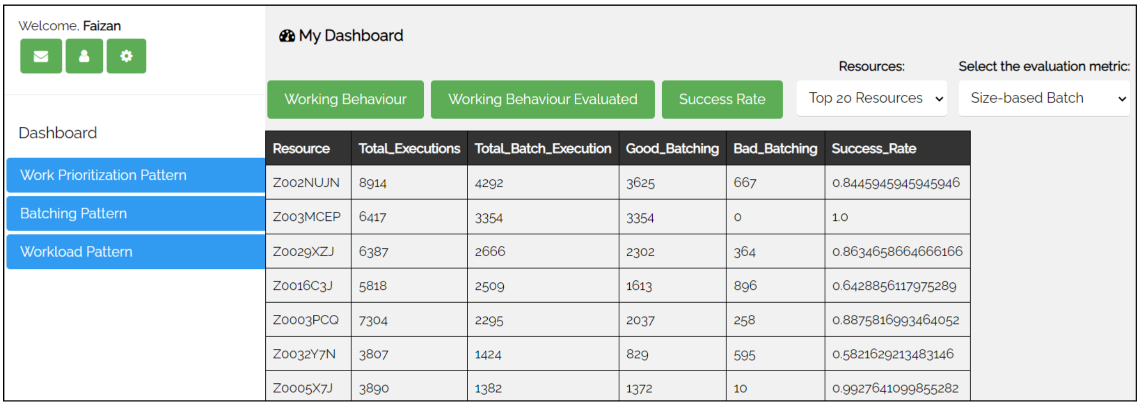
Figure 5.12 showcases the success rate of the employed batching strategy as explained in the section 4.5.4, specifically focusing on *"Size-based Batch"*. The displayed results illustrate the success rate achieved by each resource through the utilization of size-based batching. The success rate column in the table showcases the corresponding success values for each resource's batching approach. In this case, the *"Top20 Resources"* have been selected for analysis. By clicking on the *"Success Rate"* button, the figure presents the



Resource	Total_Executions	Total_Batch_Execution	Good_Batching	Bad_Batching
Z002NUJN	8914	6962	5864	1098
Z003MCEP	6417	5353	5347	6
Z0003PCQ	7304	4871	4458	413
Z0027ARY	5261	4329	3778	551
Z0000D8W	4799	3346	2949	397
Z0032Y7N	3807	3252	1983	1269
Z000IB3U	3508	3148	2765	383

Figure 5.11: Evaluated Working Behavior - Batching

results, providing insights into the effectiveness of size-based batching for different resources.



Resource	Total_Executions	Total_Batch_Execution	Good_Batching	Bad_Batching	Success_Rate
Z002NUJN	8914	4292	3625	667	0.8445945945945946
Z003MCEP	6417	3354	3354	0	1.0
Z0029XZJ	6387	2666	2302	364	0.8634658664666166
Z0016C3J	5818	2509	1613	896	0.6428856117975289
Z0003PCQ	7304	2295	2037	258	0.8875816993464052
Z0032Y7N	3807	1424	829	595	0.5821629213483146
Z0005X7J	3890	1382	1372	10	0.9927641099855282

Figure 5.12: Success Rate - Batching

5.1.6 Workload Pattern:

When selecting the `workload pattern` button, users are presented with a dedicated view specifically designed for this pattern. The workload pattern offers three distinct metrics: `workload per year, workload per quarter, and workload per month` as explained in the section 4.4.1. These metrics allow for a detailed analysis of resource workload at different time intervals, providing valuable insights into resource capacity and utilization. In addition to the three main metrics, the workload pattern also includes an extra button for `optimal workload`. This button provides information on success rates associated with different workload ranges, helping users identify the optimal workload range for each resource as explained in section 4.4.5. This feature offers valuable guidance in resource allocation and workload management.

Figure 5.13 illustrates the dashboard for the workload pattern. The dashboard includes the button for optimal workload, which enables users to explore the optimal workload ranges for resources. The drop-down menu allows users to select the desired metric for analysis. By simply clicking on the relevant buttons, users can access the results and gain insights into resource workload patterns. Overall, the tool's interface, functionalities, and visualization capabilities empower users to gain deep insights into resource working behavior by analyzing event logs and evaluating KPIs. The various patterns and metrics available cater to different aspects of resource performance, enabling effective decision-making and process optimization.

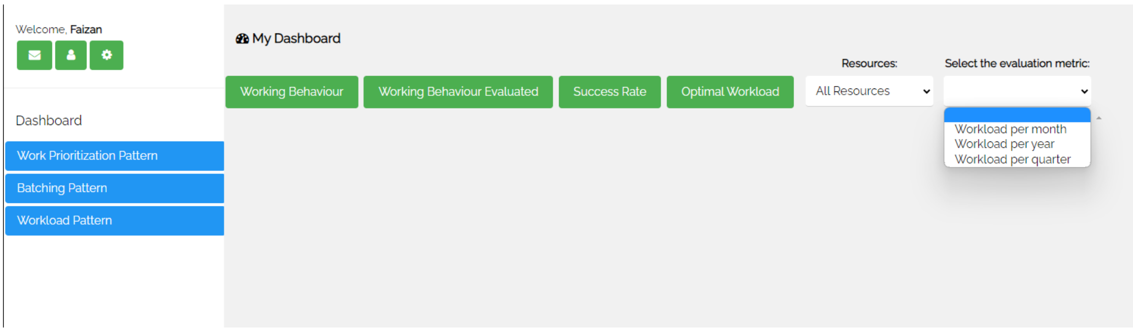
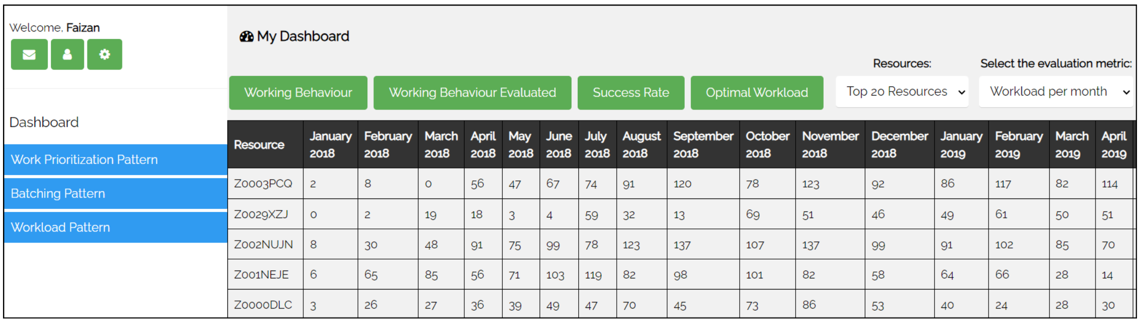


Figure 5.13: Workload Pattern Dashboard

Figure 5.14 presents an analysis of the working behavior of resources in terms of workload as explained in section 4.4.2. The metric selected for this analysis is *"Workload per month"* and the *"Top20 Resources"* are considered in the results. The table includes columns representing each month of the year, starting from the minimum timestamp recorded in the event log. This analysis provides insights into the workload assigned to each resource during each month. By examining this information, we can understand the distribution of work among the resources over time. It allows us to identify which resources handled how much work in each month, providing a comprehensive view of the workload distribution within the organization.



Resource	January 2018	February 2018	March 2018	April 2018	May 2018	June 2018	July 2018	August 2018	September 2018	October 2018	November 2018	December 2018	January 2019	February 2019	March 2019	April 2019
Z003PCQ	2	8	0	56	47	67	74	91	120	78	123	92	86	117	82	114
Z0029XZJ	0	2	19	18	3	4	59	32	13	69	51	46	49	61	50	51
Z002NUJN	8	30	48	91	75	99	78	123	137	107	137	99	91	102	85	70
Z001NEJE	6	65	85	56	71	103	119	82	98	101	82	58	64	66	28	14
Z000DLC	3	26	27	36	39	49	47	70	45	73	86	53	40	24	28	30

Figure 5.14: Working Behavior - Workload

Figure 5.15 illustrates the evaluated workload for the resources as explained in the section 4.4.3. The selected metric for evaluation is *"Workload per year"* and the analysis focuses on the *"Top20 Resources"*. Clicking on *"Working Behavior Evaluated"* provides the results. The table displays two columns: workload per year and successful workload per year. These columns represent the total workload assigned to each resource and the portion of the workload that was successfully completed, respectively. The evaluation allows us to assess the efficiency and productivity of resources in handling their assigned workload over the course of a year. By examining the results, we can identify resource-specific workload patterns, evaluate their performance, and gain insights into the resource allocation and utilization within the organization.

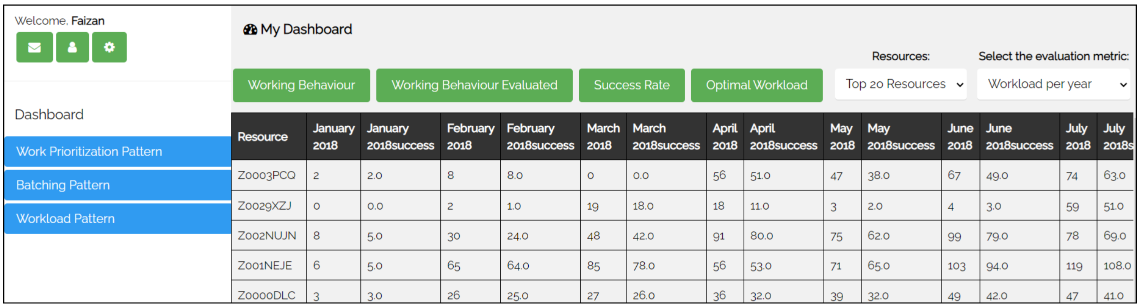


Figure 5.15: Evaluated Working Behavior - Workload

Figure 5.16 presents the success rate of the workload as explained in the section 4.4.4 for the selected metric of *"Workload per quarter"* and the *"Top20 Resources*. The success ratio is determined by calculating the percentage of successfully completed workload in each quarter. Clicking on *"Success Rate"* provides access to the corresponding results. The table displays columns for each quarter, showing the assigned workload and the corresponding success rate. This analysis enables us to evaluate the effectiveness of resource allocation and workload management in achieving successful outcomes. By examining the success rates across quarters, we can identify periods of high performance, potential bottlenecks, or areas for improvement. This information helps in optimizing resource allocation, workload distribution, and overall process efficiency to enhance organizational productivity and success.

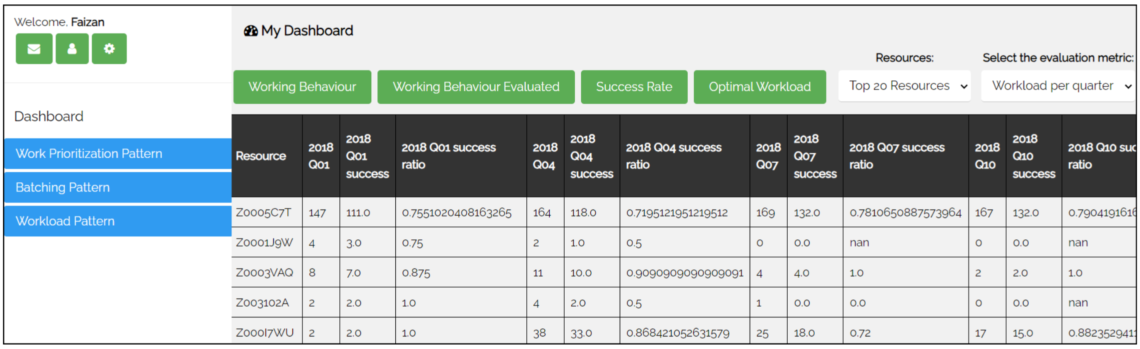
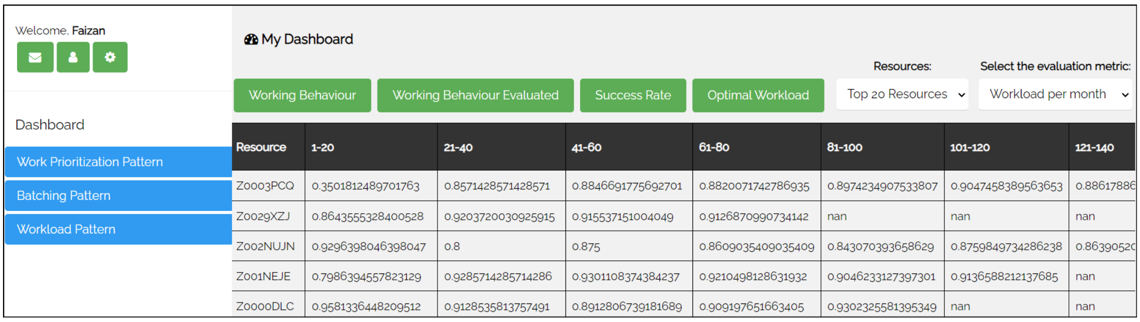


Figure 5.16: Success Rate - Workload

Figure 5.17 showcases the success rate across various workload ranges as explained in section 4.4.5, enabling the identification of optimal workload ranges for each resource. The metric chosen for analysis is *"Workload per month"* and the *"Top 20 Resources"* have been selected. By clicking on *"Optimal Workload"*, the corresponding results are obtained. The table presents success ratio values for different workload ranges, allowing us to determine the workload range that yields the highest success ratio for each resource. The optimal workload range signifies the range in which a resource demonstrates the most effective and efficient performance. This information is vital for workload allocation and resource management decisions, as it assists in optimizing resource productivity, minimizing delays, and maximizing process efficiency. By aligning resource workloads with their respective optimal ranges, organizations can achieve enhanced performance outcomes, improved resource utilization, and increased overall process success.



Resource	1-20	21-40	41-60	61-80	81-100	101-120	121-140
Z0003PCQ	0.3501812489701763	0.8571428571428571	0.8846691775692701	0.8820071742786935	0.8974234907533807	0.9047458389563653	0.88617886
Z0029XZJ	0.8643555328400528	0.9203720030925915	0.915537151004049	0.9126870990734142	nan	nan	nan
Z002NUJN	0.9296398046398047	0.8	0.875	0.8609035409035409	0.843070393658629	0.8759849734286238	0.8639052
Z001NEJE	0.7986394557823129	0.9285714285714286	0.9301108374384237	0.9210498128631932	0.9046233127397301	0.9136588212137685	nan
Z0000DLC	0.9581336448209512	0.9128535813757491	0.8912806739181689	0.909197651663405	0.9302325581395349	nan	nan

Figure 5.17: Success Rate across different Workload Ranges

Chapter 6

Evaluation

6.1 Introduction:

In the evaluation chapter, we focus on assessing the authenticity of the identified working behavior of resources, while also investigating the connection between such behavior and the success of process instances. This evaluation is a vital component in accomplishing the core aim of our research, which is to enhance process performance by identifying the optimal working behavior of resources. Through a comprehensive evaluation of resource behavior, we can acquire valuable insights into the efficiency of their working behavior and identify opportunities for improvement. This evaluation is critical for comprehending the direct link between resource behavior and the effective completion of process instances, thus facilitating more efficient and effective process management.

6.2 Summary of Methodology:

Our methodology involves a comprehensive approach to identifying resource behavior and evaluating its impact. While a detailed explanation of the methodology is provided in Chapter 4, we can summarize it as follows.

We begin by gathering an `event log` that captures relevant data about activities performed by the resources. This event log serves as the foundation for our analysis. Additionally, we select a `Key Performance Indicator (KPI)` from a set of available KPIs to measure the impact of resource behavior. To analyze resource behavior, we consider three main features: `working behavior, batching, and workload`. We assess the behavior of resources based on these factors and evaluate their impact by analyzing their working behavior in each aspect. This evaluation is conducted in relation to the chosen KPI. Furthermore, we introduce various `metrics` to evaluate resource behavior across different lines. These metrics serve as reliable indicators of resource behavior. By applying these metrics to our data set, we can thoroughly evaluate the behavior of resources and assess their influence on the overall success of the process.

6.3 Rank Calculation and Geometric Mean:

In the evaluation of the working behavior of different resources, a `rank calculation` approach is utilized to assess and compare their success rate. This approach involved analyzing various criteria associated with resource behavior, such as workload, prioritization pattern or batching behavior. The rank calculation process aimed to assign a numerical rank to each resource based on their behavior in each metric. For instance, if the metric being evaluated is workload, the resource with the highest workload would receive the top rank, indicating a higher level of activity or workload. For prioritization pattern, the resource with the highest number of tasks performed would be ranked at the top, signifying a greater degree of work in process instances. Similarly for batching, the resource with higher execution of task as batches would be ranked higher. After determining individual ranks for each metric, the ranks were consolidated using the `geometric mean`. The geometric mean is a statistical measure that accounts for the relative behavior across multiple metrics and provides a comprehensive assessment of the overall effectiveness of each resource. By calculating the geometric mean, an `aggregated rank` was obtained, representing the resource's overall standing based on the evaluated metrics. This rank calculation and geometric mean approach offered a holistic evaluation of resource working, considering their behavior across various metrics. It allowed for the identification of resources that excelled in specific metrics as well as those that demonstrated consistent high success rate across multiple metrics. Ultimately, this methodology provided valuable insights into the relative behavior and effectiveness of different resources within the context of the evaluated criteria.

6.3.1 Workload

Working Behavior - Workload: In order to evaluate the `workload` of each resource, we conducted an analysis that involved calculating the `average workload` for different time periods, including monthly, quarterly, and yearly intervals. These average values allowed us to gain valuable insights into how the workload was distributed across the resources over time. To effectively compare and evaluate the workload of each resource, we implemented a ranking system. This system involved assigning ranks to the resources based on their workload, with the highest rank given to the resource with the highest workload. This approach provided us with a comprehensive understanding of how the workload varied for each resource across different time periods. To summarize and consolidate the workload evaluation, we utilized the geometric mean of the rankings. By calculating the geometric mean, we derived a single measure that considered the relative workload of each resource. This measure, in turn, facilitated the identification of each resource's overall standing in terms of workload as we computed an aggregated rank based on the geometric mean. This aggregated rank provided a comprehensive assessment of each resource's workload. This evaluation sheds light on the distribution of workload among the resources and offers valuable insights for effective resource management and allocation. It enables us to make informed decisions regarding resource allocation and ensures optimal utilization of resources in achieving organizational goals.

The evaluated workload results are presented in Figure 6.1. The figure displays the calculated average workload values for each resource, as well as the corresponding ranks for each metric (monthly, quarterly, and yearly). Additionally, the figure showcases the geometric

Resource	Average Workload_Month	Average Workload_Quarter	Average Workload_Year	Rank_Month	Rank_Quarter	Rank_Year	Geometric Rank	Aggregated Rank Workload
admin1	23	69	171	1	1	1	1.000000	1
Resource01	18	52	122	2	2	3	2.289428	2
Resource10	17	50	125	3	3	2	2.620741	3
admin2	10	29	72	4	4	4	4.000000	4
Resource07	8	23	57	5	6	5	5.313293	5
Resource19	8	23	56	6	5	6	5.646216	6
Resource02	8	21	52	7	7	7	7.000000	7
Resource04	7	20	49	9	8	8	8.320335	8
Resource08	7	18	43	8	9	10	8.962809	9
Resource03	6	18	44	11	10	9	9.966555	10
Resource11	7	18	43	10	11	11	10.656022	11

Figure 6.1: Ranked Workload

mean rank and the aggregated rank, which provides a comprehensive assessment of each resource's workload.

Correlation between Workload Ranks and Geometric Rank: In order to assess the relationship between the rank obtained for different workload metrics and their mean geometric rank, we conducted a comprehensive `correlation` analysis. The objective was to determine the extent to which the ranks of individual workload metrics aligned with the overall ranking based on the geometric mean. To begin with, we identified specific workload metrics that represented the resource workload during different time periods. We then computed the geometric rank by combining the ranks of these workload metrics for each time period. This resulted in a composite rank that captured the overall behavior across the different metrics. To measure the relationship between each **`workload metric rank`** and the **`geometric rank`**, we utilized two commonly used correlation methods: `Pearson correlation` and `Spearman correlation`. The Pearson correlation provided insights into the strength and direction of a linear relationship between variables, while the Spearman correlation assessed the monotonic relationship regardless of linearity. The resulting correlation coefficients quantified the degree of association between each workload metric rank and the geometric rank. Higher correlation coefficients indicated a stronger alignment between the individual metric ranks and the overall ranking based on the geometric mean. Conversely, lower coefficients suggested a weaker alignment. The correlation analysis yielded valuable insights into the relationships between the ranks of different workload metrics and the geometric rank. The obtained correlation values, depicted in Figure 6.2, provided a visual representation of these relationships. Specifically, the figure displayed the correlation coefficients for both Pearson and Spearman methods, indicating the strength and direction of the correlations.

Measure	Correlated	Correlated With	Pearson	Spearman
Workload	Rank_Month_Workload	Geometric_Rank_Workload	0.950754	0.916305
Workload	Rank_Quarter_Workload	Geometric_Rank_Workload	0.983631	0.986105
Workload	Rank_Year_Workload	Geometric_Rank_Workload	0.964890	0.958858

Figure 6.2: Correlation : Workload VS Geometric Rank

Success Rate - Workload: To assess the `success rate` of each resource based on their workloads, we conducted an evaluation of their `success rates`. For each resource, we calculated the average success ratio across three metrics: month, quarter, and year. This comprehensive analysis allowed us to gain insights into the resource's success rate across different time frames, considering their workload as the primary evaluation criteria. To determine the success rate ranking for each resource, we assigned a rank based on the highest success ratio. Resources with a higher success rate achieved a higher rank, indicating their higher successful results. This ranking approach provided a clear understanding of how each resource performed in relation to their workload. Next, we computed the geometric mean of these rankings, which provided a consolidated measure of the overall success rate for each resource. By considering the success rate rankings across different time frames and incorporating workload as a factor, the geometric mean allowed for a comprehensive evaluation of the resource's behavior. The final aggregated ranking unveiled valuable insights into which resource demonstrated the highest success ratio relative to their workload. This ranking highlighted the resource's effectiveness and their ability to achieve successful outcomes.

Resource	Average SuccessRatio_Month	Average SuccessRatio_Quarter	Average SuccessRatio_Year	Rank_Month	Rank_Quarter	Rank_Year	Geometric Rank	Aggregated Rank SR
Resource14	0.866667	1.000000	1.000000	4	1	3	2.289428	1
Resource04	0.988889	0.984000	0.960000	1	4	6	2.884499	2
Resource01	0.929803	0.994558	0.995633	3	2	4	2.884499	3
admin2	0.969974	0.965608	0.952747	2	6	9	4.762203	4
Resource06	0.716667	0.984615	0.992647	9	3	5	5.129928	5
Resource07	0.466667	0.800000	1.000000	19	12	1	6.109115	6
Resource02	0.865568	0.915103	0.842123	5	7	19	8.728519	7
Resource22	0.511111	0.966667	0.954545	17	5	8	8.793659	8
admin1	0.837379	0.889626	0.870384	6	9	18	9.905782	9
Resource05	0.823333	0.884762	0.870879	7	10	17	10.596985	10
Resource11	0.616497	0.908889	0.939964	15	8	11	10.969613	11

Figure 6.3: Ranked Success Rate - Workload

Figure 6.3 illustrates the aggregated ranking, which was computed based on the success rate of each resource. The figure presents the average success rate for each resource across different metrics. Subsequently, the corresponding ranks determined for each metric, as well as the geometric mean, are displayed. Finally, the figure showcases the aggregated rank obtained from the evaluation process.

Correlation between Workload Success Rate Ranks and Geometric Rank: In this section, we focus on evaluating the relationship between the ranks obtained for success rate in different metrics and their mean geometric rank. Our objective is to assess the degree to which the ranks of each metric, derived from success rates, align with the overall ranking based on the geometric mean. As discussed in the previous section 6.3.1.0.3, we identified specific metrics that capture success rates across different time frames. By calculating the geometric rank, which combines the ranks of success rates for each metric, we obtained a comprehensive measure of the overall rank based on success rates. To analyze the relationship between the rank of each success rate metric and the geometric rank, we employed two common correlation methods: Pearson correlation and Spearman correlation as we did in section 6.3.1.0.2. A stronger alignment between the ranks of the measures and the overall ranking is shown by higher correlation coefficients, whilst a weaker alignment is indicated by lower values.The obtained correlation values are depicted in Figure 6.4. We examined the correlation between the ranks of different metrics, namely *"Rank month"*, *"Rank quarter"*, *"Rank year"*, based on success rates, and the overall *"Geometric rank"*. The correlation coefficients for both Pearson and Spearman methods are displayed in the figure. By correlating the ranks of each success rate metric with the geometric rank, we gained insights into the strength and direction of their relationships. The correlation values provide an indication of how closely the ranks of individual metrics, based on success rates, align with the overall ranking determined by the geometric mean.

Measure	Correlated	Correlated With	Pearson	Spearman
Success Ratio	Rank_Month_SuccessRatio	Geometric_Rank_SuccessRatio	0.928774	0.924363
Success Ratio	Rank_Quarter_SuccessRatio	Geometric_Rank_SuccessRatio	0.958633	0.957039
Success Ratio	Rank_Year_SuccessRatio	Geometric_Rank_SuccessRatio	0.909229	0.905474

Figure 6.4: Correlation: Success Rate Workload VS Geometric Rank

Correlation between Workload and Success Rate: To evaluate the correlation between workload and success rate, we followed a systematic approach. First, we merged the two datasets, combining the workload data with the success rate data, ensuring that each record had corresponding ranks based on workload and ranks based on success rates. With the merged dataset and the calculated geometric ranks, we proceeded to perform a correlation analysis. Specifically, we correlated each column associated with workload *"Rank month workload," "Rank quarter workload," "Rank year workload," "Geometric Rank Workload,"* and *"Aggregated Rank Workload"* with the corresponding column associated with success rate, namely *"Rank month success rate," "Rank quarter success rate," "Rank year success rate," "Geometric Rank Success Rate,"* and *"Aggregated Rank Success Rate."* We employed both Pearson correlation and Spearman correlation to capture different aspects of the relationship, as discussed in Section 6.3.1.0.2. By examining the correlation coefficients for each pair of workload and success rate metrics, we obtained valuable insights into the strength and direction of their association. Higher correlation coefficients indicated a stronger alignment between the ranks, suggesting a consistent and impactful relationship between workload and success rate. Conversely, lower correlation coefficients implied a weaker or less consistent association. This thorough analysis enabled us to uncover the intricate interplay between workload and success rate. The correlation coefficients served as quantitative measures, allowing us to assess the degree of association and providing valuable insights into the impact of workload on success rate and vice versa. By following this comprehensive evaluation approach, we gained a deeper understanding of how workload and success rate interact and influence each other in the context of our study.

Figure 6.5 illustrates the correlation values obtained through our analysis, which aimed to examine the potential relationship between resource workload and process instance success. By correlating the ranks derived from the workload data with the ranks obtained from the success rate data, we sought to determine whether there is any correlation between the workload of a resource and the success of process instances.

Optimal Workload: To determine the `optimal workload` range for each resource, a similar evaluation was conducted based on the success rate identified against different

Measure	Correlated	Correlated With	Pearson	Spearman
Workload vs SuccessRatio	Rank_Month_Workload	Rank_Month_SuccessRatio	0.748263	0.748263
Workload vs SuccessRatio	Rank_Quarter_Workload	Rank_Quarter_SuccessRatio	0.810682	0.810682
Workload vs SuccessRatio	Rank_Year_Workload	Rank_Year_SuccessRatio	0.693769	0.693769
Workload vs SuccessRatio	Geometric_Rank_Workload	Geometric_Rank_SuccessRatio	0.813832	0.811138
Workload vs SuccessRatio	Aggregated_Rank_Workload	Aggregated_Rank_SuccessRatio	0.810790	0.810790

Figure 6.5: Correlation: Workload VS Success Rate

`workload ranges`. This evaluation aimed to identify the workload range that exhibited the highest success ratio for each resource, indicating the range where the resource performed most effectively. The evaluation process involved calculating the average success ratio for each resource against different workload ranges. This was achieved by consolidating the relevant data from three metrics: workload per month, workload per quarter, workload per year. These metrics provided information on the success rate associated with specific workload ranges for each resource. To obtain the average success ratio against each workload range, the data for these metrics was merged, the mean success ratio for each resource was computed by grouping the merged data by *"Resource"* and calculating the `mean value`. This step provided a consolidated measure of the average success ratio for each workload range, specific to each resource. Finally, the optimal workload range for each resource was determined by identifying the workload range with the `highest success ratio`. Overall, this evaluation provided valuable insights into identifying the optimal workload range for each resource, based on their respective success ratios. The findings of this analysis can contribute to effective resource management and allocation, ensuring that resources are assigned workload ranges where they demonstrate the highest level of success.

Resource	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-inf	optimal_workload
Resource01	1.000000	1.000000	NaN	0.990476	0.987805	NaN	NaN	0.986394	1.000000	NaN	0.991266	11-20
Resource02	0.833333	0.857472	NaN	0.810811	NaN	NaN	NaN	NaN	0.919540	NaN	NaN	81-90
Resource03	1.000000	1.000000	1.000000	0.948718	NaN	NaN	0.971014	NaN	0.977011	NaN	NaN	11-20
Resource04	1.000000	0.966667	0.920000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	71-80
Resource05	0.889286	0.833333	NaN	NaN	NaN	0.884615	NaN	NaN	NaN	NaN	NaN	51-60
Resource06	0.991667	0.974359	NaN	NaN	NaN	NaN	0.985294	NaN	NaN	NaN	NaN	61-70
Resource07	1.000000	1.000000	1.000000	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	11-20
Resource08	0.975000	1.000000	0.892857	1.000000	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	11-20
Resource09	0.449735	0.847059	NaN	NaN	0.857143	NaN	NaN	NaN	NaN	NaN	NaN	41-50
Resource10	0.730357	1.000000	0.680000	NaN	NaN	NaN	0.893939	0.932432	NaN	NaN	0.912725	11-20

Figure 6.6: Evaluated Optimal Workload Range

Figure 6.6 showcases the evaluation of optimal workload ranges for each resource. The figure presents the average success ratio calculated for different workload ranges. By computing the average success ratio across various workload ranges, we gained insights into the successful tasks of each resource in relation to different workload levels. The workload range with the highest success ratio was considered to be the optimal workload range for the respective resource.

To visualize the numerosity of resources with optimal workload levels, we create a graph that illustrates the numerosity of resources within each range. This graph provides a clear representation of how resources are distributed across optimal workload levels. Figure 6.7 showcases the graph depicting the numerosity of resources with optimal workload within the defined ranges. The x-axis represents the workload ranges, while the y-axis represents the number of resources falling within each range. The height of each bar on the graph indicates the number of resources in that workload range. By analyzing the graph, we can observe the optimal distribution pattern for resources. Th workload evaluation, coupled with this graph visualization, provides valuable insights into resource allocation efficiency and workload optimization. It enables us to make data-driven decisions to improve resource utilization, redistribute workloads, and ensure an optimal balance across the system.

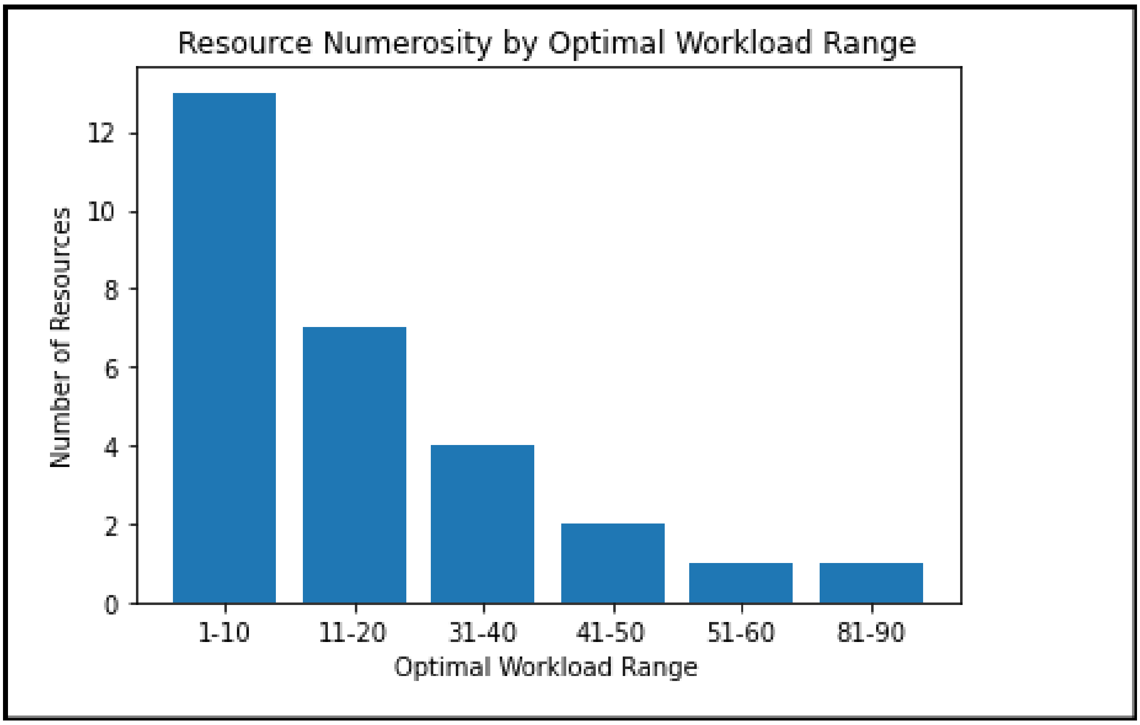


Figure 6.7: Optimal Workload Chart

6.3.2 Batching

Working Behavior: To assess the `batching` behavior exhibited by resources, a ranking approach was employed. The resources were ranked based on their implementation of batching behavior, considering three metrics: activity-based batch, time-based batch, and size-based batch. This evaluation aimed to identify the resources that displayed a higher degree of batching behavior. The ranking process involved examining the total number of batch executions performed by each resource. Resources with a greater number of `total batch executions` were assigned higher ranks, indicating a higher level of batching behavior. This ranking procedure was applied to each of the three metrics, resulting in separate rankings for activity-based batch, time-based batch, and size-based batch. To consolidate these rankings and obtain an overall assessment of batching behavior, the geometric mean of the ranks was calculated. The geometric mean accounted for the relative rankings across the three metrics, providing a comprehensive measure of the resources' batching behavior. By identifying the geometric mean rank, an aggregated rank was computed, indicating the resource's overall batching behavior. This evaluation process enabled the determination of which resource prominently opted for the batching behavior. The resource with the highest aggregated rank exhibited the highest level of batching behavior across the three metrics. By employing this ranking methodology, valuable insights were gained into the extent of batching behavior demonstrated by the resources. These findings contribute to a better understanding of resource behavior, facilitating improved resource management and decision-making processes.

Resource	Total_Batch_Execution_AB	Total_Batch_Execution_TB	Total_Batch_Execution_SB	Rank_AB	Rank_TB	Rank_SB	Geometric Mean	Aggregated Rank
Resource10	261	222	72	2.0	15.0	1.0	3.107233	1
Resource01	40	1156	22	8.0	1.0	4.0	3.174802	2
admin1	332	242	22	1.0	14.0	4.0	3.825862	3
Resource04	15	461	32	14.0	4.0	2.0	4.820285	4
Resource07	54	370	22	5.0	7.0	4.0	5.192494	5
Resource03	2	541	22	24.0	2.0	4.0	5.768998	6
Resource05	2	422	32	24.0	5.0	2.0	6.214465	7
Resource02	19	502	0	11.0	3.0	13.0	7.541987	8
Resource19	101	86	11	4.0	21.0	9.0	9.109767	9
Resource11	29	277	11	9.0	12.0	9.0	9.905782	10

Figure 6.8: Ranked Batching

6.8 focused on assessing the batching behavior exhibited by the resources. To determine the ranking for each resource based on their batching behavior, the total batch executions were taken into consideration. Resources with a higher number of batch executions were assigned a higher rank, indicating a greater degree of batching behavior. This ranking approach allowed for the comparison and assessment of the resources' batching behavior across various metrics, such as activity-based batch, time-based batch, and size-based batch. After assigning ranks for each metric, a geometric mean was calculated. The geometric mean provides a comprehensive measure that considers the relative ranking across different metrics. By calculating the geometric mean, an aggregated rank was obtained, which identified the overall batching behavior ranking for each resource.

Rank Correlation between Batching Metrics Rank and Geometric Rank: To evaluate the relationship between the ranks obtained for different batching metrics and the overall geometric rank, we conducted a comprehensive correlation analysis. Our focus was to assess the alignment between the ranks of activity-based batching (AB), time-based batching (TB), and size-based batching (SB) with the calculated geometric rank. By combining the ranks of AB, TB, and SB, we derived a single geometric rank that represented the overall batching behavior. This composite rank provided a consolidated measure of the effectiveness of the different batching metrics in achieving successful outcomes. To measure the relationship between each batching metric rank (rank AB, rank TB, and rank SB) and the geometric rank, we employed both Pearson correlation and Spearman correlation as discussed in Section 6.3.1.0.2. Higher correlation coefficients indicated a stronger alignment between the ranks of the individual batching metrics and the overall geometric rank. Conversely, lower coefficients suggested a weaker alignment. Thus, the correlation analysis allowed us to understand the extent to which the ranks of AB, TB, and SB aligned with the overall batching behavior as indicated by the geometric rank. We calculated both the Pearson and Spearman correlation coefficients to capture different aspects of the relationship between the batching metric ranks and the geometric rank. These correlation coefficients provided quantitative measures of the strength and direction of the associations. Figure 6.9 visually presents the correlation values for rank AB, rank TB, and rank SB with the geometric rank, further illustrating their relationships.

Measure	Correlated	Correlated With	Pearson	Spearman
Batching	Rank_AB_Batching	Geometric_Rank_Batching	0.809559	0.824864
Batching	Rank_TB_Batching	Geometric_Rank_Batching	0.868753	0.878341
Batching	Rank_SB_Batching	Geometric_Rank_Batching	0.740122	0.671425

Figure 6.9: Correlation: Batching VS Geometric Rank

Success Rate : To assess the relationship between batching behavior and the success of process instances, we conducted an evaluation of the `success rates` for each resource based on their batching practices. The objective was to determine the impact of batching behavior on the overall success of process instances. For each resource, a ranking was assigned based on their success rate against the specific batching behavior implemented. The resource with the highest success ratio against the batching behavior received the top rank, and subsequent ranks were assigned accordingly. This ranking process was performed for each type of batching behavior, including activity-based, time-based, and size-based batching. To consolidate these rankings and gain an overall understanding of the resource's successful tasks in relation to batching behavior, the geometric mean of the ranks was calculated. The geometric mean took into account the relative rankings across the different batching metrics, providing a comprehensive measure of the resource's success rate in relation to their batching behavior. By identifying the geometric mean rank, an aggregated rank was derived, indicating the overall rank of the resource based on their batching practices. This evaluation methodology allowed us to identify which resources demonstrated the most successful tasks in relation to their batching behavior. Resources with higher aggregated ranks exhibited a strong correlation between their batching practices and the success of process instances. By employing this ranking approach, we gained valuable insights into the impact of batching behavior on the success of process instances. These findings contribute to a deeper understanding of resource behavior, enabling improved decision-making and resource management strategies.

6.10 presents the success rate of each resource across different metrics. The evaluation aimed to assess the resources based on their success rates in relation to batching activities. To determine the ranking for each resource, their success rates were considered across various metrics. Resources with higher success rates were assigned higher ranks, indicating higher successful tasks in relation to batching activities. This ranking approach allowed for a comprehensive comparison and evaluation of resource behavior across different metrics, including activity-based batch, time-based batch, and size-based batch. After assigning ranks for each metric, a geometric mean was calculated. The geometric mean provided a consolidated measure that considered the relative rankings across different metrics. By computing the geometric mean, an aggregated rank was obtained, which

Resource	Success_Rate_AB	Success_Rate_TB	Success_Rate_SB	Rank_AB	Rank_TB	Rank_SB	Geometric Mean	Aggregated Rank SR
Resource07	1.000000	1.000000	1.000000	1.0	1.0	1.0	1.000000	1
Resource14	1.000000	1.000000	0.000000	1.0	1.0	9.0	2.080084	2
admin3	1.000000	1.000000	0.000000	1.0	1.0	9.0	2.080084	3
Resource01	1.000000	0.992215	1.000000	1.0	12.0	1.0	2.289428	4
Resource03	1.000000	0.970425	1.000000	1.0	18.0	1.0	2.620741	5
Resource18	1.000000	0.986395	0.000000	1.0	14.0	9.0	5.013298	6
Resource06	1.000000	0.984925	0.000000	1.0	15.0	9.0	5.129928	7
Resource17	1.000000	0.982759	0.000000	1.0	16.0	9.0	5.241483	8
Resource04	0.866667	0.997831	1.000000	22.0	11.0	1.0	6.231680	9
Resource11	1.000000	0.823105	0.000000	1.0	28.0	9.0	6.316360	10

Figure 6.10: Ranked Success Rate - Batching

identified the overall ranking for each resource based on their success rates in relation to batching activities.

Rank correlation between Success Rate Metrics Rank and Geometric Rank: In this section, our focus shifts to evaluating the relationship between the ranks obtained for success rates in different batching metrics and their mean geometric rank. Our objective is to assess the extent to which the ranks of each batching metric, derived from success rates, align with the overall ranking based on the geometric mean. As discussed in the previous section 6.3.2.0.3, we consider three specific batching metrics: *"Rank AB Success Rate"*, *"Rank TB Success Rate"*, and *"Rank SB Success Rate"*. These metrics capture success rates across different batching strategies. By calculating the geometric rank, which combines the ranks of success rates for each batching metric, we obtain a comprehensive measure of overall batching behavior based on success rates. To analyze the relationship between the rank of each success rate metric (*"Rank AB Success Rate"*, *"Rank TB Success Rate"*, *"Rank SB Success Rate"*) and the geometric rank, we employ both Pearson correlation and Spearman correlation, similar to the approach discussed in section 6.3.1.0.2. Higher correlation coefficients indicate a stronger alignment between the ranks of the success rate metrics and the overall ranking, while lower coefficients suggest a weaker alignment. Figure 6.11 displays the obtained correlation values for *"Rank AB Success Rate"*, *"Rank TB Success Rate"*, and *"Rank SB Success Rate"* with the geometric rank. By correlating the ranks of each success rate metric with the geometric rank, we gain insights into the strength and direction of their relationships. These correlation values provide a quantitative assessment of how closely the ranks of individual success rate metrics, based on different batching strategies, align with the overall ranking determined by the geometric mean. In summary, this analysis enables us to evaluate the impact and effectiveness of activity-based, time-based, and size-based batching strategies in achieving successful outcomes. By assessing the correlation between the ranks of success rate metrics and the geometric rank, we gain a deeper understanding of how these batching strategies relate to the overall batching performance based on success rates.

Correlation between Batching and Success Rate: To assess the correlation between batching strategies and success rates, we followed a systematic approach. Initially, we merged the datasets containing batching data and success rate data, ensuring that each record had corresponding ranks based on batching metrics and ranks based on success rates. After merging the datasets and calculating the geometric ranks, we conducted

Measure	Correlated	Correlated With	Pearson	Spearman
Success Ratio	Rank_AB_SuccessRatio	Geometric_Rank_Success_Ratio	0.667931	0.601355
Success Ratio	Rank_TB_SuccessRatio	Geometric_Rank_Success_Ratio	0.744835	0.827883
Success Ratio	Rank_SB_SuccessRatio	Geometric_Rank_Success_Ratio	0.426001	0.365151

Figure 6.11: Correlation: Success Rate Batching VS Geometric Rank

a correlation analysis. Specifically, we correlated each column associated with batching metrics, including *"Rank AB Success Rate," "Rank TB Success Rate," "Rank SB Success Rate," "Geometric Rank Batching,"* and *"Aggregated Rank Batching,"* with the corresponding column associated with success rate, such as *"Rank month success rate," "Rank quarter success rate," "Rank year success rate," "Geometric Rank Success Rate,"* and *"Aggregated Rank Success Rate"*. To capture various aspects of the relationship between batching and success rate, we employed both Pearson correlation and Spearman correlation, as discussed in Section 6.3.1.0.2. By examining the correlation coefficients for each pair of batching and success rate metrics, we gained insights into the strength and direction of their association. Higher correlation coefficients indicated a stronger alignment between the ranks, suggesting a consistent and influential relationship between batching strategies and success rates. Conversely, lower correlation coefficients indicated a weaker or less consistent association. This comprehensive evaluation approach allowed us to uncover the intricate interplay between batching strategies and success rates. The correlation coefficients served as quantitative measures, enabling us to assess the degree of association and providing valuable insights into the impact of batching on success rates and vice versa. By following this approach, we gained a deeper understanding of how batching strategies and success rates interact and influence each other within the context of our study.

Figure 6.12 depicts the correlation values resulting from our analysis, which aimed to explore the potential connection between resource batching behavior and process instance success. We conducted the analysis by correlating the ranks derived from various batching metrics with the ranks obtained from the success rate data. Our objective was to determine whether a correlation exists between a resource's batching behavior and the success of process instances.

Measure	Correlated	Correlated With	Pearson	Spearman
Batching vs Success Ratio	Rank_AB_Batching	Rank_AB_SuccessRatio	0.459295	0.705572
Batching vs Success Ratio	Rank_TB_Batching	Rank_TB_SuccessRatio	0.284126	0.401323
Batching vs Success Ratio	Rank_SB_Batching	Rank_SB_SuccessRatio	0.681923	0.801339
Batching vs Success Ratio	Geometric_Rank_Batching	Geometric_Rank_SuccessRatio	0.261055	0.404245
Batching vs Success Ratio	Aggregated_Rank_Batching	Aggregated_Rank_SuccessRatio	0.403387	0.403387

Figure 6.12: Correlation: Batching VS Success Rate

6.4 KPI Evaluation

To evaluate the key performance indicators (KPIs) and their correlation in the context of resource behavior analysis, an event log was utilized. Each KPI was assigned a specific column in the event log, with corresponding values of either 1 or -1 based on the success or failure of the KPI for each case. This approach allowed for a comprehensive evaluation of various KPIs to understand their impact on the process's performance. One important KPI considered was rework, which measures the occurrence of repetitive activities within a case. By checking for repetitions of activities within a case, a value of -1 was assigned if rework was present, indicating the need for additional effort and potential inefficiencies. Conversely, a value of 1 was assigned if no repetitions were observed, reflecting smoother and more efficient process execution. Another significant KPI was lead time, which measures the total time taken to complete a case. To calculate lead time, the total number of days for each case was determined. Additionally, the mean number of days across all cases was computed. If the total days for a specific case exceeded the mean days, a value of -1 was assigned, indicating a longer processing time compared to the average. Conversely, if the total days were less than or equal to the mean days, a value of 1 was assigned, signifying a faster processing time. Furthermore, due date compliance was considered a KPI, reflecting the ability of a case to be completed within the planned time frame. For each case, it was determined whether the case ended before the planned date. If the case was completed on or before the planned date, a value of 1 was assigned, indicating successful compliance. Conversely, if the case exceeded the planned date, a value of -1 was assigned, highlighting a delay or non-compliance. By following this approach, a table was generated with assigned values of 1 or -1 for each KPI against their respective case IDs. This table provided a clear representation of the performance of each case in terms of the considered KPIs. The correlation between the different KPIs was then analyzed to identify any relationships or dependencies.

Figure 6.13 illustrates the obtained KPI correlation values, providing a visual representation of the relationships and dependencies between the different KPIs. This analysis aids in understanding how resource behavior influences process performance and helps identify potential areas for improvement.

Correalted	Correlated with	Pearson	Spearman
Lead Time	Due Date Compliance	0.464706	0.464706
Lead Time	Rework	0.153916	0.153916
Due Date Compliance	Rework	0.160915	0.160915

Figure 6.13: KPI Evaluation

Chapter 7

Discussion

In this chapter, we delve into the analysis and interpretation of the obtained results, aiming to gain deeper insights into the relationship between resource behavior and process performance. The discussion revolves around the key findings, addressing the research objectives and research questions posed earlier. We have conducted an examination of the correlations between various factors, including working behavior, success ratio, and the relationship between working behavior and success ratio. Additionally, we explore the theoretical expectations, relevant literature correlations, and the implications of the observed correlations. The discussion provides a comprehensive understanding of the relationship between resource behavior and process performance, shedding light on the potential factors influencing efficiency, effectiveness, and optimization. Through this analysis, we aim to contribute valuable knowledge that can guide future process improvement initiatives and resource management strategies. Let us now delve into the detailed discussion and interpretation of the results.

7.1 Working Behavior - Workload VS Geometric Rank

The results presented in 6.2 focus on the analysis of workload across different time periods and the correlation between individual `workload ranks` and the `geometric ranks`. The purpose of this evaluation is to ensure the correctness of the rank calculation and validate its alignment by comparing geometric rank with the individual ranks based on monthly, quarterly, and yearly workloads. Let's discuss these results in more detail.

7.1.1 Rank Month Workload VS Geometric Rank Workload

Pearson correlation: 0.950754
Spearman correlation: 0.916305

The high Pearson and Spearman correlation coefficients suggest a `strong positive` relationship between the rank based on the workload per month and the geometric rank. This indicates that the monthly workload ranks align closely with the overall workload ranking derived from the geometric mean. The consistency between these measures provides confidence in the correctness of the geometric rank calculation.

7.1.2 Rank Quarter Workload VS Geometric Rank Workload

Pearson correlation: 0.983631
Spearman correlation: 0.986105

The `exceptionally high` Pearson and Spearman correlation coefficients demonstrate a robust correlation between the rank based on the workload per quarter and the geometric rank. The quarterly workload ranks align very closely with the overall workload ranking derived from the geometric mean. These strong correlations provide further evidence of the accuracy and consistency of the geometric rank calculation.

7.1.3 Rank Year Workload VS Geometric Rank Workload

Pearson correlation: 0.964890
Spearman correlation: 0.958858

The high Pearson and Spearman correlation coefficients indicate a `strong positive` relationship between the rank based on the workload per year and the geometric rank. The yearly workload ranks align well with the overall workload ranking derived from the geometric mean. This consistency between the measures reinforces the confidence in the accuracy of the geometric rank calculation.

The observed high correlations between the individual workload ranks (monthly, quarterly, and yearly) and the geometric rank provide assurance that the geometric rank calculation accurately reflects the overall behavior. The alignment between these measures indicates that the geometric rank appropriately captures the variations and trends in workload across different time periods. This evaluation of the workload ranks and their correlations with the geometric rank reinforces the robustness and reliability of the methodology employed in the identification and calculation of the ranks. It ensures that the geometric rank provides a meaningful representation of the workload across different time periods, validating its use in further analyses and interpretations.

7.2 Success Rate - Workload VS Geometric Rank

In the figure 6.4, we focus on the correlation analysis between the `success ratio rank` across different time periods (month, quarter, and year) and the `geometric rank`. The purpose of this evaluation is to validate the accuracy of the calculated ranks and their alignment with the geometric rank. Let's discuss these results:

7.2.1 Rank Month SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.928774
Spearman correlation: 0.924363

The high Pearson and Spearman correlation coefficients suggest a `strong positive` relationship between the rank based on the success ratio per month and the geometric rank. This indicates that the monthly success ratio ranks align closely with the overall success ratio ranking derived from the geometric mean. The consistency between these measures provides confidence in the correctness of the geometric rank calculation for the success ratio.

7.2.2 Rank Quarter SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.958633
Spearman correlation: 0.957039

The `high` Pearson and Spearman correlation coefficients demonstrate a robust correlation between the rank based on the success ratio per quarter and the geometric rank. The quarterly success ratio ranks align very closely with the overall success ratio ranking derived from the geometric mean. These strong correlations provide further evidence of the accuracy and consistency of the geometric rank calculation for the success ratio.

7.2.3 Rank Year SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.909229
Spearman correlation: 0.905474

The high Pearson and Spearman correlation coefficients indicate a `strong positive` relationship between the rank based on the success ratio per year and the geometric rank. The yearly success ratio ranks align well with the overall success ratio ranking derived from the geometric mean. This consistency between the measures reinforces the confidence in the accuracy of the geometric rank calculation for the success ratio.

The observed high correlations between the individual success ratio ranks (monthly, quarterly, and yearly) and the geometric rank provide assurance that the geometric rank calculation accurately reflects the overall success ratio. The alignment between these measures indicates that the geometric rank appropriately captures the variations and trends in the success ratio across different time periods. It is important to note that these discussions focus solely on the correctness and alignment of the ranks for the success ratio, rather than the performance outcomes or implications. These results serve as a crucial step in validating the methodology and establishing a foundation for the subsequent analyses and discussions on resource behavior and its impact on process performance.

7.3 Correlation : Workload VS Success Rate

In figure 6.5, when examining the results of the correlation analysis between resource behavior measured by `workload` and the `success` of the process, several key observations can be made:

Theoretical Expectation: The correlation between workload and success ratio can have both `positive` and `negative` implications. On one hand, a positive correlation is expected, suggesting that as workload increases, the success ratio improves. This expectation is grounded in the notion that higher workloads can drive increased productivity, efficiency, task prioritization, and decision-making, leading to higher success rates. On the other hand, a negative correlation is also plausible, indicating that increasing workload may result in more errors, reduced attention to detail, and decreased success rates. The specific nature of the relationship between workload and success ratio can vary depending on the context, the capacity to handle the workload, and the effectiveness of resource management and allocation.

Analysis of Obtained Results: The obtained results reveal varying correlation strengths between workload and success ratio across different measures. These correlation strengths provide insights into the relationship between the variables.

7.3.1 Rank Month Workload VS Rank Month SuccessRatio

The `strong positive` correlation coefficient of **0.748263** indicates a robust association between monthly workload ranks and monthly success ratio ranks. This suggests that as monthly workloads increase, there is a higher likelihood of achieving higher success ratios. In other words, when a resource is assigned more tasks during a particular month, they tend to produce more successful results. This correlation implies that factors such as increased engagement, improved focus, or better task management may contribute to higher success ratios when workloads are higher on a monthly basis.

7.3.2 Rank Quarter Workload VS Rank Quarter SuccessRatio

The correlation coefficient of **0.810682** indicates a `strong positive` correlation between quarterly workload ranks and quarterly success ratio ranks. This implies that quarterly workloads have a more pronounced impact on quarterly success ratios compared to monthly workloads. When a resource is assigned more tasks throughout a quarter, there is a higher likelihood of achieving higher success ratios during that period. The stronger correlation suggests that the resource's ability to produce successful outcomes are influenced to a greater extent by the workload assigned over a longer time span. This could be attributed to factors such as sustained focus, adaptability, and efficient task allocation over the course of a quarter.

7.3.3 Rank Year Workload VS Rank Year SuccessRatio

The correlation coefficient of **0.693769** suggests a `positive` correlation between yearly workload ranks and yearly success ratio ranks, albeit slightly weaker than the monthly and quarterly correlations. This implies that yearly workloads contribute to yearly success ratios, but to a lesser extent than shorter time periods. The positive correlation indicates that when a resource is assigned a higher workload over the course of a year, there is a tendency for them to achieve higher success ratios. However, the slightly weaker correlation suggests that other factors beyond workload, such as individual capabilities, external circumstances, or varying project demands, may also impact the yearly success ratio. It is important to consider a holistic view that takes into account various factors when analyzing the relationship between workload and success ratio over longer time periods.

7.3.4 Geometric Rank Workload VS Geometric Rank SuccessRatio

The high Pearson correlation coefficient **0.813832** and Spearman correlation coefficient **0.811138** indicate a `strong positive` relationship between workload geometric ranks and success ratio geometric ranks. The high correlation between the geometric ranks of workload and success ratio implies that there is a strong positive relationship between these two variables. When individuals have higher workload ranks, they also tend to have higher success ratio ranks. From an interpretive perspective, this correlation suggests that there is a potential connection between workload and success. It is likely that individuals who are assigned higher workloads have a greater opportunity to showcase their skills and capabilities, leading to more successful outcomes. Additionally, the higher workload may require individuals to be more focused, diligent, and efficient in their work, resulting in higher success ratios.

7.3.5 Aggregated Rank Workload VS Aggregated Rank SuccessRatio

The `strong positive` correlation coefficient of **0.810790** between the aggregated workload ranks and the aggregated success ratio ranks indicates a robust positive correlation. This suggests that higher aggregated workloads are associated with higher aggregated success ratios. One possible interpretation of this correlation is that individuals assigned with higher workloads tend to have higher success ratios. This could be due to several factors. For instance, individuals with higher workloads may be more engaged and motivated to complete their tasks efficiently, leading to higher success ratios. Additionally, increased workload may provide more opportunities for skill development and experience, which can positively impact performance and success. However, it is important to consider that correlation does not imply causation. While the correlation suggests a relationship between aggregated workload and aggregated success ratio, other factors may also contribute to success. Individual capabilities, resources, and external factors can all play a role in determining success ratios.

The varying strengths of correlation emphasize the influence of different time periods and aggregation methods on the connection between workload and success ratio. The stronger correlations observed for quarterly measures, geometric ranks, and aggregated ranks suggest that examining longer time periods and taking into account overall trends can lead to a more reliable understanding of the relationship. By considering quarterly measures, we can capture the broader patterns and fluctuations in workload and success ratio over time. This enables a more comprehensive analysis that accounts for any seasonal or long-term trends that may affect the relationship between these variables. Similarly, the use of geometric ranks and aggregated ranks helps to minimize the impact of outliers and individual variations. These methods provide a more holistic perspective by focusing on the overall performance trends rather than specific instances or specific individuals. This allows for a more robust assessment of the relationship between workload and success ratio. Overall, adopting longer time periods, geometric ranks, and aggregated ranks offers a more comprehensive and reliable understanding of how workload and success ratio are related. It allows for the identification of broader trends and patterns, providing valuable insights into the dynamics between these variables.

7.4 Working Behavior - Batching VS Geometric Rank

In Figure 6.9, we examine the correlation analysis between batching metrics (activity-based, size-based, and time-based) ranks against the geometric rank. The purpose of this evaluation is to validate the accuracy and alignment of the calculated ranks in relation to the geometric rank. Let's discuss these results:

7.4.1 Rank AB Batching VS Geometric Rank Batching

Pearson correlation: 0.809559
Spearman correlation: 0.824864

The `moderate to strong positive` Pearson and Spearman correlation coefficients indicate a significant relationship between the rank based on activity-based batching and the geometric rank. This suggests that the ranking of activity-based batching aligns closely with the overall batching ranking derived from the geometric mean. The consistency between these measures provides confidence in the accuracy of the geometric rank calculation for activity-based batching.

7.4.2 Rank TB Batching VS Geometric Rank Batching

Pearson correlation: 0.868753
Spearman correlation: 0.878341

The `high` Pearson and Spearman correlation coefficients demonstrate a robust correlation between the rank based on time-based batching and the geometric rank. The ranking of time-based batching exhibits a strong alignment with the overall batching ranking derived from the geometric mean. These strong correlations further validate the accuracy and consistency of the geometric rank calculation for time-based batching.

7.4.3 Rank SB Batching VS Geometric Rank Batching

Pearson correlation: 0.740122
Spearman correlation: 0.671425

The moderate Pearson and Spearman correlation coefficients suggest a `positive` relationship between the rank based on size-based batching and the geometric rank. The ranking of size-based batching shows a reasonable alignment with the overall batching ranking derived from the geometric mean. Although the correlations are not as strong as the other batching metrics, they still indicate a meaningful relationship.

The observed correlations between the individual batching metrics (activity-based, size-based, and time-based) and the geometric rank provide assurance that the geometric rank calculation accurately reflects the overall batching behavior. The alignment between these measures suggests that the geometric rank appropriately captures the variations and trends in the batching strategies employed.

7.5 Success Rate : Batching VS Geometric Rank

In Figure 6.11, we examine the correlation analysis between the `success rate` ranks for different batching metrics (activity-based, time-based, and size-based) and the corresponding `geometric rank`. The purpose of this evaluation is to assess the alignment and accuracy of the success ratio ranks in relation to the overall geometric rank. Let's discuss these results:

7.5.1 Rank AB SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.667931
Spearman correlation: 0.601355

The moderate Pearson and Spearman correlation coefficients indicate a `moderate positive` relationship between the rank based on activity-based batching and the geometric rank. This suggests that the ranking of activity-based batching in terms of success ratio exhibits some alignment with the overall success ratio ranking derived from the geometric mean. However, the correlations are not as strong as desired, indicating potential variations or discrepancies between the activity-based batching and the overall success ratio ranking.

7.5.2 Rank TB SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.744835
Spearman correlation: 0.827883

The moderate to strong Pearson and Spearman correlation coefficients indicate a significant `positive` relationship between the rank based on time-based batching and the geometric rank. This suggests that the ranking of time-based batching in terms of success ratio aligns closely with the overall success ratio ranking derived from the geometric mean. The consistency between these measures provides confidence in the accuracy of the geometric rank calculation for time-based batching.

7.5.3 Rank SB SuccessRatio VS Geometric Rank SuccessRatio

Pearson correlation: 0.426001
Spearman correlation: 0.365151

The low to moderate Pearson and Spearman correlation coefficients suggest a `weaker positive` relationship between the rank based on size-based batching and the geometric rank. This indicates that the ranking of size-based batching in terms of success ratio may have limited alignment with the overall success ratio ranking derived from the geometric mean. The correlations suggest potential discrepancies or variations between the size-based batching and the overall success ratio ranking.

The observed correlations between the success ratio ranks for different batching metrics and the geometric rank provide insights into the alignment and accuracy of these rankings. While time-based batching demonstrates a strong alignment with the overall success ratio ranking, activity-based and size-based batching show varying degrees of alignment.

7.6 Correlation : Batching VS Success Rate

In Figure 6.12, we examine the correlation analysis between different `batching` metrics and `success rate` ranks. Let's discuss these results:

7.6.1 Rank AB Batching VS Rank AB SuccessRatio

The correlation coefficient of **0.459295** indicates a `moderate positive` correlation between activity-based batching ranks and success ratio ranks. This suggests that higher ranks in activity-based batching are associated with higher success ratio ranks. In other words, there is a tendency for resources that engage in more batching behavior to achieve higher success ratios. However, it is important to note that the correlation strength is not very strong, implying that other factors may also influence success ratios.

7.6.2 Rank TB Batching VS Rank TB SuccessRatio

The correlation coefficient **0.284126** indicates a `weak positive` correlation between time-based batching ranks and success ratio ranks. This implies that there is a slight relationship between higher time-based batching ranks and higher success ratio ranks, but the correlation strength is relatively low.

7.6.3 Rank SB Batching VS Rank SB SuccessRatio

The correlation coefficient of **0.681923** reveals a `moderate positive` correlation between size-based batching ranks and success ratio ranks. This suggests that there is a tendency for higher ranks in size-based batching to be associated with higher success ratio ranks. Consequently, the findings indicate that applying batching techniques has the potential to yield positive results, as evidenced by the correlation between size-based batching and process success. . However, it is important to consider other influencing factors.

7.6.4 Geometric Rank Batching VS Geometric Rank SuccessRatio

The correlation coefficient of **0.261055** indicates a `weak positive` correlation between the geometric ranks of batching metrics and the geometric ranks of success ratio. This suggests that there is a slight relationship between the geometric ranking of batching metrics and the overall ranking of success ratio. However, it is important to note that the correlation strength is relatively low. Based on these findings, it is not necessary that the person who has used the most batching is the most successful. Other factors and variables may also play a role in determining the success of a process. However, the presence of a positive correlation suggests that applying batching techniques could have a positive impact on the performance of the process. It indicates that higher geometric ranks in batching metrics tend to be associated with higher geometric ranks in the success ratio.

7.6.5 Aggregated Rank Batching VS Aggregated Rank SuccessRatio

The correlation coefficient of **0.403387** suggests a `moderate positive` correlation between the aggregated ranks of batching metrics and the aggregated success ratio ranks. This indicates that higher aggregated ranks in batching metrics are associated with higher aggregated success ratio ranks. In other words, there is a tendency for resources with higher aggregated ranks in batching metrics to also have higher aggregated ranks in the success ratio. It is important to note that correlation does not imply causation. While this correlation suggests a relationship between batching metrics and success ratio, other factors may also contribute to the overall success of the process.

The observed correlations between the different batching metrics and success ratio provide insights into their relationship. The moderate correlations for activity-based and size-based batching suggest a relatively stronger association with success ratio ranks. On the other hand, the weak correlations for time-based batching and geometric ranks suggest a less pronounced relationship.

7.7 Correlation between KPIs

In figure 6.13 The correlation analysis between `Due Date Compliance (DDC)`, `Rework`, and `Lead Time` yields the correlation values that range from *0.1539 to 0.4647*. These correlation values indicate the strength and direction of the relationships between the variables.

7.7.1 Correlation: Lead Time VS Due Date Compliance

For the correlation between lead time and due date compliance, a coefficient of **0.4647** suggests a `moderate positive` correlation. This indicates that there is a tendency for cases with shorter lead times to have a higher likelihood of fulfilling the due date compliance requirement. However, it's important to note that the correlation is moderate, meaning that other factors may also influence the relationship between lead time and due date compliance.

7.7.2 Correlation: Lead Time VS Rework

Regarding the correlation between lead time and rework, the coefficient of **0.1539** indicates a `weak positive` correlation. This suggests a slight tendency for cases with shorter lead times to have a slightly lower likelihood of experiencing rework. However, as with the correlation between lead time and due date compliance, the correlation is weak, indicating that other factors may also contribute to the occurrence of rework.

7.7.3 Correlation: Due Date Compliance VS Rework

The correlation between due date compliance and rework yields a coefficient of **0.1609**, indicating a `weak positive` correlation. This suggests that there may be a slight association between fulfilling the due date compliance requirement and a lower likelihood of rework. However, as with the other correlations, the relationship is weak, and other factors may also influence the occurrence of rework.

It's important to note that correlation coefficients in the range of *0.1 to 0.3* (as observed in these results) generally indicate weak correlations. This means that while there may be some degree of association between the variables, the relationships are not strong or consistent. In conclusion, the correlation analysis suggests that there may be some weak positive associations between lead time, due date compliance, and rework.

7.8 Optimal Workload

The provided results in figure 6.7 show the `numerosity of resources` distributed across various `optimal workload ranges`. When considering the optimal workload ranges as a whole, it is observed that there is a relatively high cumulative numerosity of resources in the lower workload ranges, specifically in the *"1-10"* and *"11-20"* ranges. This suggests that a significant number of resources are capable of effectively managing workloads within these lower ranges. On the other hand, the cumulative numerosity of resources decreases as the workload ranges increase. The *"31-40"* and *"41-50"* ranges exhibit a relatively lower cumulative numerosity of resources, indicating that fewer individuals are equipped to handle workloads falling within these higher ranges. Overall, these results suggest that the organization has a greater availability of resources capable of managing lower workloads, while the availability decreases as the workload ranges become higher. This information can be valuable for workload allocation and resource management decisions within the organization.

Chapter 8

Conclusion

In conclusion, our research has employed a comprehensive methodology to identify and evaluate resource behavior and its impact on process success. This robust framework encompasses various elements and techniques that have allowed us to gain valuable insights into resource utilization and its relationship with overall process performance. Through the gathering of an event log capturing relevant data about resource activities, we have established a solid foundation for our analysis. By selecting key performance indicators (KPIs) from a range of available metrics, we have quantified the impact of resource behavior on process success. Our evaluation of resource behavior has focused on three key aspects: working behavior, batching, and workload. These dimensions have provided a comprehensive understanding of how resources contribute to the process and have allowed us to assess their individual and collective impact on success. Within our methodology, we have introduced a set of tailored metrics to effectively evaluate resource behavior across these dimensions. These metrics have served as reliable indicators and objective measures of resource performance. By applying these metrics to our dataset, we have conducted a thorough evaluation of resource behavior and its influence on process success. The findings from our analysis highlight areas of strength and improvement in resource behavior, contributing to more efficient process management. By understanding the optimal workload ranges, prioritization patterns, and the benefits of working in batches, organizations can optimize resource allocation, enhance task management strategies, and improve overall process efficiency. Our research has emphasized the importance of evaluating resource behavior and its impact on process management. The implications of this research extend to various industries and sectors where process management and resource allocation play crucial roles. By understanding and optimizing resource behavior, organizations can achieve higher process efficiency, cost savings, and improved customer satisfaction.

In conclusion, our comprehensive evaluation of resource behavior, based on a robust methodology, has provided valuable insights into resource utilization and its direct influence on process success. The findings serve as a foundation for making informed decisions regarding resource allocation, development, and process improvement initiatives. By leveraging these insights, organizations can strive for more efficient and effective process management.

Future Work: In the future, there are several potential areas of research and development that can further enhance the analysis of resource behavior and its correlation with process performance.

We can explore `additional KPIs` that can provide more comprehensive insights into resource behavior and process performance. Consider different metrics and KPIs such as `resource utilization`, `task completion time`, or `customer satisfaction`, which may further illuminate the relationship between resource behavior and process outcomes. We can use `machine learning` algorithms to identify patterns and predict resource behavior based on historical data. By leveraging advanced analytic and predictive modeling, it may be possible to optimize resource allocation and workload assignment in real-time, leading to more adaptive and efficient process management. We can develop `optimization algorithms` or `decision support systems` that can recommend optimal resource behavior based on specific process requirements and objectives. These tools can assist in finding the most effective working patterns, task prioritization strategies, or workloads to maximize process efficiency and achieve desired performance outcomes.

Bibliography

[1] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo Alexander Reijers. Fundamentals of business process management. In *Springer Berlin Heidelberg*, 2018.

[2] Hajo Alexander Reijers and Sabeur Mansar. Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega-international Journal of Management Science*, 33:283–306, 2005.

[3] Walid Gaaloul, Sadek Alaoui, Karim Baïna, and Claude Godart. Mining workflow patterns through event-data analysis. *2005 Symposium on Applications and the Internet Workshops (SAINT 2005 Workshops)*, pages 226–229, 2005.

[4] Massimiliano de Leoni and Wil M.P. van der Aalst. Data-aware process mining: discovering decisions in processes using alignments. In *ACM Symposium on Applied Computing*, 2013.

[5] Wil Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, volume 136. 01 2011. ISBN 978-3-642-19344-6. doi: 10.1007/978-3-642-19345-3.

[6] Wil M.P. van der Aalst, Helen Schonenberg, and Minseok Song. Time prediction based on process mining. *Inf. Syst.*, 36:450–475, 2011.

[7] Wil M.P. van der Aalst, Hajo Alexander Reijers, and Minseok Song. Discovering social networks from event logs. *Computer Supported Cooperative Work (CSCW)*, 14:549–593, 2005.

[8] Minseok Song and Wil M.P. van der Aalst. Towards comprehensive support for organizational mining. *Decis. Support Syst.*, 46:300–317, 2008.

[9] Joyce Nakatumba and Wil M.P. van der Aalst. Analyzing resource behavior using process mining. In *Business Process Management Workshops*, 2009.

[10] Zhengxing Huang, Xudong Lu, and Huilong Duan. Resource behavior measure and application in business process management. *Expert Syst. Appl.*, 39:6458–6468, 2012.

[11] Manuel Camargo, Marlon Dumas, and Oscar González Rojas. Automated discovery of business process simulation models from event logs. *ArXiv*, abs/1910.05404, 2019.

[12] Alessandro Berti. Mining weighted leaders and peripheral workers in organizational social networks based on event logs. 2016.

[13] Aekyung Kim, Josué Obregon, and Jae-Yoon Jung. Constructing decision trees from

process logs for performer recommendation. In *Business Process Management Workshops*, 2013.

[14] Anastasiia Pika, Wil M. P. van der Aalst, Colin J. Fidge, Arthur H. M. ter Hofstede, and Moe T. Wynn. Profiling event logs to configure risk indicators for process delays. In Camille Salinesi, Moira C. Norrie, and Óscar Pastor, editors, *Advanced Information Systems Engineering*, pages 465–481, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-38709-8.

[15] Arik Senderovich, Matthias Weidlich, Avigdor Gal, and Avishai Mandelbaum. Mining resource scheduling protocols. In *International Conference on Business Process Management*, 2014.

[16] Suriadi Suriadi, Chun Ouyang, Wil M.P. van der Aalst, and Arthur H. M. ter Hofstede. Event interval analysis: Why do processes take time? *Decis. Support Syst.*, 79: 77–98, 2015.

[17] Cristina Cabanillas, Lars Ackermann, Stefan Schönig, Christian Sturm, and Jan Mendling. The ralph miner for automated discovery and verification of resource-aware process models. *Software and Systems Modeling*, 19:1415 – 1441, 2020.

[18] Reza Akhavian and Amir H. Behzadan. Evaluation of queuing systems for knowledge-based simulation of construction processes. *Automation in Construction*, 47:37–49, 2014.

[19] Arik Senderovich, Matthias Weidlich, Avigdor Gal, and Avishai Mandelbaum. Queue mining - predicting delays in service processes. In *International Conference on Advanced Information Systems Engineering*, 2014.

[20] Arik Senderovich, Sander J. J. Leemans, Shahar Harel, Avigdor Gal, Avishai Mandelbaum, and Wil M.P. van der Aalst. Discovering queues from event logs with varying levels of information. In *Business Process Management Workshops*, 2016.

[21] Arik Senderovich, Matthias Weidlich, Avigdor Gal, and Avishai Mandelbaum. Queue mining for delay prediction in multi-class service processes. *Inf. Syst.*, 53:278–295, 2015.

[22] Suriadi Suriadi, Moe Thandar Wynn, Jingxin Xu, Wil M.P. van der Aalst, and Arthur H. M. ter Hofstede. Discovering work prioritisation patterns from event logs. *Decis. Support Syst.*, 100:77–92, 2017.

[23] Stefan Creemers and Marc R. Lambrecht. Modeling a hospital queueing network. 2011.

[24] S Saghafian, G. Austin, and S.J Traub (M.D.). Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Transactions on Healthcare Systems Engineering*, 5(2):101–123, 2015.

[25] Linda Green. *Queueing Analysis in Health Care*, volume 91, pages 281–307. 10 2006. ISBN 978-0-387-33635-0. doi: 10.1007/978-0-387-33636-7_10.

[26] Esaignani Selvarajah and George Steiner. Approximation algorithms for the supplier's supply chain scheduling problem to minimize delivery and inventory holding costs. *Oper. Res.*, 57:426–438, 2009.

[27] Sebastian Henn, Sören Koch, and Gerhard Wäscher. Order batching in order picking warehouses: A survey of solution approaches. 2012.

[28] G. Cachon and Christian Terwiesch. *Matching Supply with Demand*. 01 2006.

[29] Luise Pufahl and Mathias Weske. Requirements framework for batch processing in business processes. In *BPMDS/EMMSAD@CAiSE*, 2017.

[30] Luise Pufahl, Andreas Meyer, and Mathias Weske. Batch regions: Process instance synchronization based on data. *2014 IEEE 18th International Enterprise Distributed Object Computing Conference*, pages 150–159, 2014.

[31] Luise Pufahl, Ekaterina Bazhenova, and Mathias Weske. Evaluating the performance of a batch activity in process models. In *Business Process Management Workshops*, 2014.

[32] Luise Pufahl and Mathias Weske. Batch processing across multiple business processes based on object life cycles. *EMISA Forum*, 37:36–37, 2016.

[33] Yiping Wen, Zhigang Chen, Jianxun Liu, and Jinjun Chen. Mining batch processing workflow models from event logs. *Concurrency and Computation: Practice and Experience*, 25, 2013.

[34] Joyce Nakatumba. Resource-aware business process management : analysis and support. 2013.

[35] Niels Martin, Marijke Swennen, Benoît Depaire, Mieke Jans, An Caris, and Koen Vanhoof. Retrieving batch organisation of work insights from event logs. *Decis. Support Syst.*, 100:119–128, 2017.

[36] Kan Wu. Taxonomy of batch queueing models in manufacturing systems. *Eur. J. Oper. Res.*, 237:129–135, 2014.

[37] Jennifer Horkoff, Fatma Başak Aydemir, Evellin C. S. Cardoso, Tong Li, Alejandro Maté, Elda Paja, Mattia Salnitri, Luca Piras, John Mylopoulos, and Paolo Giorgini. Goal-oriented requirements engineering: an extended systematic mapping study. *Requirements Engineering*, 24:133 – 160, 2017.

[38] Axel van Lamsweerde. Goal-oriented requirements engineering: A guided tour. In *IEEE International Requirements Engineering Conference*, 2001.

[39] Lin Liu and Eric S. K. Yu. Designing information systems in social context: a goal and scenario modelling approach. *Inf. Syst.*, 29:187–203, 2004.

[40] Geert Poels, Ken Decreus, Ben Roelens, and Monique Snoeck. Investigating goal-oriented requirements engineering for business processes. *J. Database Manag.*, 24: 35–71, 2013.

[41] Daniel Amyot and Gunter Mussbacher. User requirements notation: The first ten years, the next ten years (invited paper). *J. Softw.*, 6:747–768, 2011.

Acknowledgments

In the name of Allah, the most merciful and the most beneficent. First and foremost, I express my sincere gratitude to my wife, Sara Naz Haider, for her unwavering support, guidance, and encouragement throughout this academic journey. In addition, I am genuinely grateful for her patience and faith, which have been instrumental in my success. I also extend my most profound appreciation to my supervisor, Alessandro Berti, for his invaluable guidance, expertise, and unwavering support. Without his mentorship, this research would not have been possible. I am immensely grateful for the opportunity to work on such an exciting project and for the constructive feedback that has significantly improved the quality of this work. I must also acknowledge the contributions of my dear friends, Asad Tariq and Muhammad Abdullah, for their invaluable assistance and support throughout this journey. Their editing help, feedback sessions, and discussions were instrumental in shaping the ideas and arguments presented in this research. Lastly, I express my heartfelt gratitude to my parents for allowing me to pursue my academic aspirations at a prestigious university. Despite the challenges, their unwavering support and encouragement have been an immense strength and motivation. Their unwavering commitment to my education is a debt I can never fully repay.