Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

Trym Remme & Viktor Tingstad

The Power of RPA

Why more businesses should implement automation to enable their employees' full potential

Bachelor's thesis in Economics and Management
Supervisor: Denis Becker
April 2024

NTNU
Norwegian University of
Science and Technology

Trym Remme & Viktor Tingstad

The Power of RPA

Why more businesses should implement automation to enable their employees' full potential

Bachelor's thesis in Economics and Management
Supervisor: Denis Becker
April 2024

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of Science and Technology

I. Preface

Throughout this thesis we have had the opportunity to use our three years of education to produce a thesis which delivers value to our contributing data provider to provide data-driven arguments which highlights the benefits of implementing RPA-technology.

We would like to express our gratitude to Ren Røros Intelligent Automation, and especially Stian Berg Dyrnes (CTO), for providing us with data from their clients and facilitating our success in delivering a data-driven analysis of the impacts of robotic process automation.

We would also like to thank Sparebank 1 SMN for providing us with data and valuable insights in their RPA environment. Without their cooperation and major data-contribution this thesis would be lacking significance in light of quantitative analysis.

The content of this thesis is the responsibility of the authors.

II. Executive Summary

This thesis explores how Robotic Process Automation (RPA) enhances a business' efficiency by automating repetitive tasks which usually are done manually by human workers. By automating these tasks, the business frees up human resources, which may be used for more strategic work. The foundation for this thesis revolves around the research question "How can Robotic Process Automation-technology enhance a company's efficiency regarding resources and costs?".

The existing literature on the topic includes insights from consultant agency Deloitte and case studies from IBM. These studies have proven benefits of RPA, such as significant cost savings, compliance improvements, and productivity enhancements. To back up our hypothesis of how these benefits affect a company, economic theories including cost-benefit analysis, comparative advantage, economies of scale, and the theory of constraints, have been used.

The data was collected from multiple clients of Ren Røros Intelligent Automation, spread over different industries and degree of RPA implementation. The methodology was centered around data preprocessing, exploratory analysis, where regression and classification was included, and ending on a detailed explanation into RPA's efficiency.

Our analysis revealed the relationship between the scale of RPA implementation and the benefits entangled with it, where more complex and larger implementation showed more time- and cost saving. Further analysis showed results to support the hypothesis that an increase in volume of items would correlate with better efficiency.

The thesis' findings suggest that by implementing RPA technology, there will be an efficiency increase by reducing the time which is required to be allocated to tasks, which again leads to a significant decrease in costs. Furthermore there were key factors in predicting the influence of implementation of RPA, including the complexity of the tasks.

There is an acknowledgement regarding the limitations concerning the thesis, such as the variability in the data that was available and the difficulty of generalization across different

industries. Further research is recommended to explore the qualitative impacts RPA has on employees and businesses.

The evidence in this thesis is the basis for the conclusion that RPA is a valuable opportunity for businesses to reduce costs and enhance overall efficiency. Furthermore the evidence supports the claim that adaptation and scaling of RPA technologies would help to maximize business over performance and give competitive advantages.

III. Sammendrag

Denne avhandlingen utforsker hvordan Robotic Process Automation (RPA) forbedrer effektiviteten til en bedrift ved å automatisere gjentakende oppgaver som vanligvis utføres manuelt av menneskelige arbeidere. Ved å automatisere disse oppgavene frigjør bedriften menneskelige ressurser som kan brukes til mer strategisk arbeid. Grunnlaget for denne avhandlingen dreier seg om forskningsspørsmålet "Hvordan kan Robotic Process Automation-teknologi forbedre en bedrifts effektivitet med hensyn til ressurser og kostnader?".

Den eksisterende litteraturen om emnet inkluderer innsikt fra konsulentbyrået Deloitte og case-studier fra IBM. Disse studiene har vist fordeler med RPA, som betydelige kostnadsbesparelser, forbedringer i samsvar og økninger i produktivitet. For å støtte våre hypoteser om hvordan disse fordelene påvirker en bedrift, har økonomiske teorier, inkludert kost-nytte-analyse, komparativ fordel, stordriftsfordeler og teorien om begrensninger, blitt brukt.

Dataene ble samlet inn fra flere kunder av Ren Røros Intelligent Automation, spredt over ulike bransjer, med forskjellig grad av RPA-implementering. Metodikken dreide seg om data preprosessering, utforskende analyse, der regresjon og klassifisering ble inkludert, og endte med en detaljert forklaring på RPA-effektiviteten.

Vår analyse avslørte forholdet mellom omfanget av RPA-implementering og fordelene som fulgte med det, der mer kompleks og større implementering viste større tids- og

kostnadsbesparelser. Videre viste analysen resultater som støtter hypotesen om at økning i volum av elementer ville korrelere med bedre effektivitet.

Avhandlingens funn antyder at ved å implementere RPA-teknologi vil det være en økning i effektiviteten ved å redusere tiden som kreves for oppgaver, noe som igjen fører til betydelige kostnadsbesparelser. Videre var det nøkkel-faktorer som predikerte innflytelsen av RPA-implementering, inkludert kompleksiteten til oppgavene.

Det er en anerkjennelse av begrensningene knyttet til avhandlingen, som variabiliteten i tilgjengelige data og vanskeligheten med generalisering på tvers av ulike bransjer. Videre forskning anbefales for å utforske de kvalitative innvirkningene RPA har på ansatte og bedrifter.

Bevisene i denne avhandlingen er grunnlaget for konklusjonen om at RPA er en verdifull mulighet for bedrifter for å redusere kostnader og forbedre overordnet effektivitet. Videre støtter bevisene påstanden om at tilpasning og skalering av RPA-teknologier vil hjelpe til med å maksimere bedriftens ytelse og gi konkurransefordeler.

Table of Contents

IV. Abbreviations

- API: Application Programming Interface
- FTE: Full Time Equivalent
- GUI: Graphical User Interface
- RPA: Robotic Process Automation
- RSS: Residual Sum of Squares
- SQL: Structured Query Language
- TOC: Theory of Constraints

V. List of Tables

VI. List of Figures

1. Introduction

There are few companies today which do not have some sort of interaction with technology through their daily work. In most cases there are manual tasks taking resources like time, and therefore money, away from the company's potential. Through robotic process automation-technology (RPA) companies can free up these resources and become more efficient and innovative. In this thesis we want to answer the research question:

"How can Robotic Process Automation-technology enhance a company's efficiency regarding resources and costs?"

Our motivation for this thesis is to give the potential customers of our cooperating company data-driven selling points as to why they should implement these kinds of technologies and which opportunities emerge from freeing resources. Our interest lies both in the process of how we do it, and also the immediate effects it has on a company's bottom line and time allocation. An often overlooked issue in many companies is the deployment of highly skilled personnel for simple manual tasks. RPA offers a solution by automating these tasks, enabling employees to redirect their expertise towards more strategic areas. Our objective is to investigate the tangible effects of such implementations on a company's time utilization and overall efficiency.

To achieve this, we collected and analyzed relevant data from a spectrum of Ren Røros Intelligent Automation's clients. This diverse set of clients vary in size, industry, and degree of automation implementation. The complexity of each case ranges from straightforward archiving to intricate conditional case management.

By conducting a comprehensive analysis of data from customers which have implemented RPA technology, our study aims to uncover the technology's impact on a company's efficiency. We will explore how these implementations influence resource allocation, reduce costs, and ultimately contribute to enhanced business performance.

2. Theory and Literature

2.1. Literature Review

Since the introduction of RPA-technology in 2015 there have been researches on how RPA-technology contributes to a company's effectiveness. Most of this literature is not shared openly as the foundation of the reports is not unique for each provider of this technology, and may be generalized for any consulting agency which provides it. We have however gone through some other case studies which we thought were relevant to our thesis.

2.1.1. Deloitte: 3rd Annual Global RPA Survey Report

Deloitte is one of the major consulting companies worldwide and provides RPA-as-a-service as well as consulting in implementation and maintenance of RPA-systems. In their global RPA survey from 2017 (Watson & Wright, 2017) they received over 400 responses from a variety of companies. The results highlight five main benefits of implementing RPA-technology. The survey shows that the subjects which had implemented RPA-technology in their business experienced a **payback time of less than 12 months** on their investment. The other benefits the survey highlights is how 92% of businesses **improved their compliance**, 90% **improved the overall quality and/or accuracy** of the automated processes, 86% reported an **improvement in productivity** and 59% experienced a **reduction in costs.** This only shows how the cost of implementation is low compared to the expected time needed to experience a profit, and how most businesses experience improvement across multiple dimensions of their day-to-day business.

The survey also focuses on the cost of implementation amongst those businesses willing to share their experiences. In general, surveyed companies **underestimated the cost of implementation**, with 63% saying their expectation of time to implement were not met, and 37% saying their expectations of economic cost of implementing were not met. Yet 78% of participating companies which had already implemented RPA-technology were planning to significantly increase their investment in RPA over the next three years.

A rather surprising discovery from the survey was that only 3% of surveyed companies had started scaling their digital workforce after initial implementation, clearly showing that **scaling RPA proves to be more difficult than expected**. Another challenge reported amongst the implementers was **resistance from employees against implementing**, or piloting, RPA-technology. This resistance is reported to fade off as the implementers move to a more stable stage of implementation where employees have started to experience benefits from it. The main challenges in these stages are often meeting the expected speed of integrating more of the company's existing systems.

2.1.2. IBM: Case Studies

IBM has conducted numerous case studies on their customers which use RPA in their business. These case studies highlight the benefits of implementing RPA as well as showing the potential of implementing RPA in a variety of industries.

The Pennsylvania-based sandwich shop chain Primanti Brothers implemented a RPA robot to generate sales- and labor reports, which saved the company 2000 man hours resulting in a USD 84000 cost saving. Primanti Brothers achieved a 100% return on investment in three months. Removing the manual task of generating these reports leaves the business owners and their advisory board more time to make decisions faster and with a foundation of higher quality. (King, 2022)

In the pharmaceutical industry the 2020-startup Selta Square uses RPA-technology together with their self-developed AI algorithms to run a "first-of-its-kind, automated pharmacovigilance service". By using IBMs RPA software they have accelerated the pharmacovigilance process by a multiplier greater than 4, and reduced the average time for searches to be completed from 5 minutes down to 1 minute and 11 seconds. This allows competent analysts to work more efficiently and discover flaws in the medicines and treatments faster than before. (Spencer, 2022a)

New Mexico Mutual, a company specializing in writing workers' compensation insurance policies, have experienced a freeing of 3,5 hours every day after implementing RPA in their insurance policy renewal process. As this process normally makes up 80% of the underwriters' time, increasing efficiency in this area gives the underwriters time to handle customer care cases which delivers more value to the customer. (Spencer, 2022b)

2.2. Technical Theory

2.2.1. Introduction to Robotic Process Automation

RPA-technology is a "software technology that makes it easy to build, deploy, and manage software robots that emulate human actions interacting with digital systems and software" (UiPath, 2024). It interacts with graphical user interface (GUI) and web applications to mimic how a human uses the application to complete a given task. RPA robots often use other available tools such as application programming interface (API) to interact with applications without having to use a GUI. This approach results in significant time reduction as the robot is not required to wait for a user-friendly interface to load.

2.2.2. Advantages of Robotic Process Automation

One of the main advantages of RPA is that with the correct logic and proper monitoring, the robot processes a greater volume of cases with less errors due to the lack of manual steps. Although there is some work to be done in order to implement an automatic process flow, the return of the time invested in this implementation is more often than not, greater than what the business would have saved by not implementing it. The mobility and modifiability of RPA-technology also makes the business more scalable and flexible in its expansion towards greater implementation. After investing in the initial setup of a RPA-rig the costs of expanding the implementation is relatively low.

Another great advantage of RPA is its ability to work on most applications without any customization or invasive changes to its software. The technology introduces ways to extend a single applications utility as well as closing the gap between different applications in the business pipeline to enable them to work together. This makes it perfect for older systems which businesses rely on for their daily operations, when needing to incorporate certain parts of these systems into newer, more modern systems to increase effectiveness. The same goes for migrating data from older systems into newer systems where RPA functions as the bridge and converter between the systems.

2.2.3. Considerations

All of the aforementioned advantages contribute to boost the company's productivity, but the technology also introduces some challenges and perspectives to be evaluated and considered. Integrating RPA-technology into the existing business flow often introduces new roles to take the lead on planning, implementing and maintaining the new technology. For most companies these roles can be hard to fill from within, as the competency solemnly exists in the company's existing workforce, and the workload is not enough to recruit a new employee. One way to resolve this challenge is to subscribe to the services from a consultant firm with the appropriate competency and resources. This allows the RPA-team to work more efficiently and deliver higher quality compared to what a single new employee could do.

Another perspective to take into consideration is how introducing robots in the process may raise security- and data privacy-concerns. As the robot potentially will need to access systems meant for authorized personnel only, anyone with access to the robot may also get access to areas they are not meant to access. However, this might actually be a preference in certain cases where you would want the least amount of people accessing or handling the case. There have been several examples where poor access control and logging lead to serious consequences for the business, such as fines and even millions in inside theft - the latest major event being the summer intern at a norwegian banks case management which embezzled almost 75 million NOK in 2023 (Skurtveit & Sørbø, 2023). With careful considerations, proper access control and logging, such incidents may at least be detected and handled before they get out of control.

A third perspective to keep in mind is the fact that not all cases are worked perfectly by the robot, leading to different versions of exceptions. These exceptions usually require some sort of human attention and will reduce the effective time saved, as the employee receives the task in return to be performed manually. However, by following best practice, most tasks will already consist of less work because the robot has solved parts of the process, and saved the result so that the human worker can continue from where the robot failed instead of performing the whole task. So even though there will be a reduced effect of exceptioned tasks, the business still benefits from the partial job performed by the robot.

2.3. Economic Theory

We will in our discussion go through certain economic theories such as cost-benefit, the theory of comparative advantage, economies of scale and theory of constraints. They will give insight into whether or not the implementation is profitable, the benefits obtained and potential hindrance you get rid of.

2.3.1. The Cost of Employment (Cost-benefit)

Cost-benefit analysis is "the process of comparing the projected or estimated costs and benefits associated with a project decision to determine whether it makes sense from a business perspective" (Stobierski, 2019). This traditionally implies taking the total projected benefits and subtracting all the costs. When we see that the costs outnumber the benefits, we would normally conclude that it would be a bad decision to make, and the same goes the other way around, when the benefits outnumber the costs, it's generally a good decision

The general sentiment in the theory is to only make investments which give the company an upper hand. Let us say a business typically has 20 employees with an average hourly rate of 350 NOK, and they each devote 30 minutes of each of their workday to do simple manual tasks, resulting in a daily cost of 3 500 NOK. Any solution to free their time that takes less than 3 500 NOK daily will eventually pay itself back.

For instance, suppose a business invests 150 000 NOK in implementing RPA to solve a manual task, which otherwise would incur a daily cost of 2 000 NOK. After 75 days that investment will pay itself back.

2.3.2. Theory of Comparative Advantage

Comparative advantage is "an economy's ability to produce a particular good or service at a lower opportunity cost than its trading partners" (Hayes, 2023). Opportunity cost refers to the potential benefits that a business or an individual misses out on when they choose one alternative over another. Often we use this as an argument for why companies, countries and even individuals benefit from the trade industry.

We take a look at a company which can produce two products at a better rate than other companies; product A which they can produce at a much higher rate, and product B which they can produce at a slightly higher rate. In this case we would say that they have a comparative advantage when producing product A, since the opportunity cost is at its lowest. If there later would be a need for the company to get a hold of product B, they would be better off trading or buying from another company.

The theory of comparative advantages introduces and claims that there is a benefit to specializing in the field you are comparatively better than the others around you, instead of competing in fields you are equally skilled at, and rather trade with each other.

Even though there seems to be a lot of advantages to this way of setting up an economy or business, such as higher efficiency and better profits, it's not without its downsides. When developing countries trade with already developed countries, or companies in these countries trade, there is a huge risk of exploitation. Likewise over-specialization might make countries dependent on global food prices, which would be a detriment to developing countries.

In the context of RPA and automation, we might see highly skilled workers who could do manual tasks at a fast rate, perhaps even at the rate of what an automated robot could, however their knowledge and insight into a company's business dealings and decision making is far superior. It would be better for the company to put that worker at the decision table and the automated robot at the manual tasks. It would be better for a company where highly skilled workers participate in business meetings, rather than doing simple copying and pasting for 8 hours.

2.3.3. Economies of Scale

Economies of scale are commonly defined as "cost advantages reaped by companies when production becomes efficient" (Kenton, 2022). By a high production output a company experiences a lowering of the average cost of production, enabling the company to sell their products at a lower rate than their competitors and still make money.

As businesses get larger and produce more goods, the production cost per unit of a good goes down. This happens when a company is producing a larger amount of goods, and is able to spread a lot of the regular expenses, like rent or common services like HR and administration, over a bigger set of products, which in turn opens for the opportunity to sell this product at a lower rate than businesses with fewer goods produced with similar expenses.

When businesses grow and scale upwards they often have workers who specialize in certain fields in the labor process. This often happens in parallel to an increase of investment in better technological machinery. With specialization and machinery upgrades, the production time generally drops significantly which in turn will help increase the production amount.

Let us look at a bank implementing RPA across different programs and platforms in order to reduce the amount of repetitive tasks, such as invoice processing or the creation of financial reports. After the RPA implementation has been integrated into an increasing portion of the different departments, the bank will eventually start to experience economies of scale. The first and most obvious effect is the cost reduction due to decreased manual labor. Furthermore there is an efficiency increase as manual tasks no longer work as a bottleneck. The tasks are performed faster and with a higher accuracy than earlier, which leaves the bank with a higher quality of work. With the implementation across different departments the bank would also standardize the processes such that all the work is done in the same efficient way.

2.3.4. Theory of Constraints

The theory of constraints (TOC) was first introduced in the 1980s by Dr. Eliyahu M. Goldratt in his book *The Goal* (1984). In Goldratt’s book he describes how bottlenecks were identified in an industrial manufacturing process at a production plant. Goldratt’s main character receives the task to turn the operations at the plant from being unprofitable and unreliable, to successful. A physicist gets involved in the process and explains how bottlenecks impact the overall flow of materials.

The core idea of the theory is that a process is limited in achieving its goal by a few number of constraints, or bottlenecks. These constraints are usually parts of a process which have a lower throughput than the rest of the process. This could be extraction of data which has to be done using GUI or data which needs to be reviewed by an overworked third party. The theory emphasizes the importance of identifying and reducing their impact to improve overall system performance.

In the modern era a lot of these manual production processes, as described by Goldratt, have already been automated and bottlenecks minimized, but new challenges emerge as the digitization of our life rapidly increases. Technology is adopted and automation is much needed in a widespread of industries. When connecting different departments in a process the importance of smooth transitions with minimal bottlenecks is required to meet the increasing demand from consumers. TOC is used as a tool to identify bottlenecks in such complex and interconnected digital systems, and technology such as RPA is a central component in overcoming these constraints and optimizing workflows.

Even though there are a lot of advanced technologies involved in today's workflows, a lot of digital processes still involve manual tasks performed by humans. As a human has physical limitations when it comes to speed, accuracy and scalability, there is often a limit as to how efficiently a task can be performed by a human. In comparison to humans the limitations of a computer or “robot” are much less sensitive in light of the same attributes. Although some other limitations may emerge when it comes to the ability to reflect and make human-like decisions, as discussed in chapter 2.2.3.

In RPA-technology these bottlenecks could be eliminated by automating the most critical bottleneck in a process (partial implementation) or by restructuring the whole manual process to eliminate all bottlenecks, big and small, throughout the process (full implementation). These two approaches have their advantages and disadvantages depending on the specific process in which the customer wants to automate.

3. Data and Methodology

3.1. Data Gathering

In this study we have cooperated with Ren Røros Intelligent Automation, a consultant firm based in Trondheim, which delivers RPA as a service to various local and national companies. Through this cooperation we have been granted access to data regarding their customer's analytics from implementing RPA-technology in their business-flow.

In the final dataset we gathered information about each customer and their processes, as well as volume of transactions, work hours saved and robotic work time for each process. We combined these numbers with publicly accessible business data such as employee count, to generate a general estimate on the amount of resources each process saved the company.

RPA-technology in itself and the tools for implementation are accessible for whoever wants to use them, which makes the possession of it not a comparative advantage. Therefore each company needs to utilize it together with their employees' knowledge, in the most efficient and unique way possible, to get a comparative advantage to their competitors. This makes the companies we are working with somewhat hesitant when it comes to sharing a lot of information about their implementation and especially costs and business sensitive data. This has led to an inconsistent dataset where we have been able to retrieve some information from each company, but also lose a lot of comparison potential when only analyzing the common variables we have been able to retrieve from all parties. As a result we have not been able to compare key measurements such as implementation- and maintenance costs.

Another challenge we have met when gathering data for our thesis is to get in contact with the key people involved in decision-making, implementation and end-users. We wanted to conduct several qualitative interviews with these people to highlight how the freed resources from implementing RPA are utilized in a more efficient way. We also struggled with finding a good way to measure this utilization and the benefits that emerge from it, as the variety of potential interview objects was too narrow to make a generalization and the results would be highly dependent on the subjective thoughts of individuals, as well as their ability to describe their experience. This is something we would have dedicated more time and effort in the further workings of this thesis.

3.2. Data Introduction

3.2.1. Variables

Although we had several challenges in the data gathering phase, which led to a reduced dataset, the data we got was more than sufficient enough to do an analysis and achieve the objective of this thesis which is to answer the question as to why companies should implement RPA. To argue for the implementation we would need some statistical measurements, which we have calculated using a set of variables, some included in the dataset we got and some created by our calculations. Our description of the variables is found in table 3.1.

Table 3.1: Description of Variables

Variable	Description
Customer	A categorical variable that specifies the observed customer, but is disguised for confidentiality reasons. 13 customers chosen from across multiple industries, ranging from economy and real-estate to cars and renovation.
Volume	A numeric variable that measures the volume of tasks performed by a robot for each of the queues, and customers. (Per queue) **Mean**: 10 230 **Std**: 23 927 **Min**: 2 **Max**: 244 566
Queue	A categorical variable which is used to group queue items which belongs to different queues and therefore have similar characteristics and statistical foundation. **Count**: 163
Status	A categorical variable which represents which state (e.g. successful, pending or failed) the queue item was in at the time of the extraction of data.
Worked time (robot)	A numerical variable which measures the amount of time (in minutes) a robot spends on performing a certain task.
	(Table is continued on next page)

Table 3.1: Description of Variables

Variable	Description
	(Table continued from previous page)
Standard time (human)	A numerical variable which is an estimate of the amount of time (in minutes) a human would use to perform a specific task. A user's estimated time of completion multiplied with 2,5 to account for the human inefficiency. Further explained in the next chapter. **Mean**: 22.35 **Std**: 36.65 **Min**: 0 **Max**: 300
Saved time (Minutes)	Sums up all the time for successful tasks which have been performed by a robot and multiplies this with our standard time to get a realistic measurement of how much human time has been saved by this solution. **Mean**: 169k **Std**: 390k **Min**: 0 **Max**: 3 057k
Saved time (hours)	The saved time variable, but divided by 60 to get hours saved.
Saved time (FTEs)	The same saved time variable as hours, but has again been divided by 1650 to get full time equivalents (FTE) saved by this solution. The factor of 1650 is defined as one FTE, when accounting for holidays and vacations.
Company size	A categorical variable which describes the size of the company based on hours saved and reported number of employees.
Complexity	A categorical variable based on standard time, which tells us how complex the automated process is expected to be.

3.2.2. More About the Multiplier for Standard Time

The optimal flow of performing a task is often given the best circumstances for the employee performing the task, with immediate access to the resources and data needed, no distractions, coffee breaks nor barriers along the way. This is also known as the "happy"-flow. Unfortunately the actual flow almost never fits this description. One of the teams we have been in contact with said the following about their discussions with the client and their estimates for each task:

> *"When we get an estimate for 10 minutes per case we ask: 'So your employees can perform 6 tasks each hour for 8 hours throughout the workday?', and the answer is often that it is impossible and that they probably perform less than half of that volume."* [translated from Norwegian]

Based on these experiences the standard of the industry is to multiply the estimates by 2.5, meaning a task estimated to be done in 10 minutes in the happy-flow is set to 25 minutes on the average workday portion. This also accounts for factors such as inefficiency and contextual delay, i.e. the time it takes for an employee to move from one type of task to another, physically and mentally.

Other theories such as lean- and agile methodologies support these assumptions and divide it into resource- and flow efficiency. Resource efficiency describes how many tasks an employee manages to get done in a maximum utilization of the resources available. Flow efficiency, on the other hand, describes how the organization as a whole (one or more employees combined) is able to complete a concrete task so that the flow is continuous and effective, with minimal breaks and delays.

The need for factorization is often more visually detectable in environments and workspaces where the average workday is affected by delays, overproduction, wrongful production, bad management and/or a variable workload and with lack of standardization. However, the theory also applies to the digital workspace of most offices where the employees experience delays in digital systems, slow response from different contributors and a culture for coffee breaks.

3.3. Data Preprocessing

The RPA-software BluePrism, like most RPA-technologies, are based on working a queue of work items, so the extraction of our data is done by exporting these work queues as csv-files from a SQL-database (Appendix A). Based on our scope we removed several items to minimize the size of the dataset. Our filtering included extracting only data from 2023 as our scope was to look at one year of production and we chose to look at the production in the latest full year (i.e. items finished after 01.01.2023 and before 31.12.2023). The exported raw data came in a standard format with 15 columns, but only 6 were of interest for our analysis[1].

The next steps were to clean the data and remove potential duplicates. When working with queues there is a column indicating which attempt each case was performed at. This form of control comes in handy when maintaining a RPA queue, but it also means the queue includes failed attempts on the same cases. This would not be a problem in general, but as we use worked time as a variable in our analyses and the queue adds time to the total each attempt, resulting in a deviation from the actual time worked for cases with more than one attempt.

Some of the queues from one of our participating data providers were combined between two of their customers, which meant we had to split these and assign each customer their share of the common queues. This was done in cooperation with the data provider to find the most correct split ratio. We ended up with a 40/60 split because related data indicated that one of the customers accounted for 60% of the total amount of work items on the most significant queues.

Although the standard export format of BluePrism queues include a lot of information on each worked case, the businesses also have their own system to estimate the standard time for each queue. Therefore we were required to merge these lists into our dataset by matching the queue name and customer. To minimize the workload on our machines, we chose to perform a grouping of the work items by their respective queues, before merging this information with the standard time from the queue definitions.

With all the queues linked to the corresponding customer and merging in the standard time we could calculate how much time each queue saved the customer in worked minutes, hours and FTEs. We then proceeded to rename some columns in our dataset for better readability, before the dataset was ready for our exploratory analysis.

[1] DISCLAIMER: The original dataset (before preprocessing) contained sensitive data and will therefore not be included in this thesis.

3.4. Exploratory Analysis

3.4.1. Classification

In order to better understand how different segments of our study performed in different analyses, we decided to classify them into unifying groups depending on different variables. The most obvious classifications for our particular dataset was to group the customers by size, to see if size matters when it comes to perceived effect, and to group the queues by their complexity, to see what kind of processes gave the best benefit.

Company Size

To collect information about the participating customers' size we used their reported employee count and revenue for 2022, as most numbers from 2023 were not available at the time of this thesis. We were unfortunately not able to get the reported revenue for each customer as some of them reported as part of a bigger concern. This made the classification somewhat simple and one-sided, as a count of employees in a company is rarely a good measurement alone. However, we decided to continue the classification process to see where it led.

By comparing the reported employee count and the customer's saved time in FTEs we generated a measurement for the company's alternative workforce increase. This measurement is the ratio of saved FTEs over reported employee count, and shows how the company's would have needed to increase their workforce if they were to match today's productivity with a hybrid workforce of both human and digital workers. The results of the calculation showed that there was a lot of variety among the participating customers regardless of their reported employee count.

There were three distinct groups of customers when we analyzed the results (table 3.2):

Table 3.2: Classification by Alternative Workforce Increase

Classification	**Increase**
Low increase	Less than 1%
Medium increase	1-20%
High increase	More than 20%

This gave us the following classification for each customer (table 3.3):

Table 3.3: Customers Classified by Alternative Workforce Increase

Customer (anonymized)	Employees	Saved time (FTEs)	Alternative Workforce Increase	Classification
FINANCE 1	180	83.9660	0.4665	High increase
FINANCE 2	450	169.6412	0.3769	High increase
FINANCE 3	51	10.8306	0.2124	High increase
TRANSPORT 1	19	1.0613	0.0559	Medium increase
FINANCE 4	20	0.9633	0.0482	Medium increase
REAL-ESTATE 1	260	8.4666	0.0326	Medium increase
RENOVATION 1	120	1.6836	0.0140	Medium increase
CONSTRUCTION 1	81	0.4796	0.0059	Low increase
RENOVATION 2	103	0.3135	0.0030	Low increase
CAR 1	132	0.2272	0.0017	Low increase
CAR 2	121	0.0371	0.0003	Low increase
CAR 3	56	0.0089	0.0002	Low increase
CONSTRUCTION 2	259	0.0071	0.0000	Low increase

The majority of the participating companies did not have a significant increase. In conversations with one of the data providers we discovered a correlation between this measurement and the perceived degree of implementation for each company. The degree of implementation was a variable we would like to use in our analyses as we have hypotheses that we would be able to explain why some customers experience more effect from implementing RPA than others. However, there were unfortunately no existing good measurements of degree of implementation. Although we saw a correlation between our "Alternative Workforce Increase"-index and the data provider's perceived degree of implementation, the foundation for the index was too narrow and the degree of implementation not defined well enough. Therefore we decided not to use this as a substitute for degree of implementation in our further analyses.

Complexity

When automating different processes, you must first come to the realization that not all the processes are equally as simple. We have therefore defined three different classifications for the degree of complexity for each queue. These classes are based on the amount of time a human would use to perform such a task, i.e. the “standard time”-variable (table 3.4).

Table 3.4: Classification by Standard Time

Classification	Timeframe (standard time)[2]
Simple	Less or equal to 5 minutes
Medium Complexity	Between 5 minutes and 20 minutes (including 20 minutes)
Very Complex	More than 20 minutes

This distinction is made because we believe that there would be a correlation between how complex a process is and how much time is saved. Our hypothesis was that the companies which were in the beginning of their RPA-journey would benefit more from automating the less complex tasks, as the required effort and knowledge is lower than for more complex tasks.

We also believed that there is a correlation between complexity and volume, where simple tasks have a higher volume, which could be linked to higher value for customers and employees. However, our analysis shows that in our dataset with this split, the simple tasks did not have the highest volume. The overall volume was mainly represented in the “medium complexity”-group where more than half of the total volume was located (54%), accounting for 38% of saved time. Whereas simple tasks with 27% of the total volume accounted for 7% of saved time, and the highest complexity had 19% of the volume, but did account for almost 55% of saved time. The fact that complex tasks stood for most of the saved time came as no surprise, but these processes also require a lot more development and resources.

3.4.2. Linear Regression

Regression is a statistical method that is used to determine the strength between a dependent variable and one or more other independent variables. The objective of linear regression is to estimate the coefficients β_0 and β_1 which minimizes the sum of squared differences between

[2] As mentioned before these numbers are multiplied by 2.5 to account for human inefficiency

the observed and predicted values by the linear equation $y = \beta_0 + \beta_1 x + \epsilon$, where x is the independent variable, y is the dependent variable and epsilon (ϵ) is the error term.

Assumptions

There are some assumptions that are made before doing linear regression that should be noted:

1. Linearity - The relationship between the independent variable, denoted by X, and the mean of the dependent variable, denoted by Y, is linear.
2. Homoscedasticity - The variability of the residuals, the difference between the observed and predicted values of the dependable variable, is constant across all values of the independent variable. This means that the spread of the residuals along the regression line is unchanged as we move along the spread of the independent variable. (This will be pointed out visually further down)
3. Independence - the observations are not influenced by each other. Each observation is distinct and thus not affected by the values of the other observations, meaning each value of the dependent variable is not directly influenced by another value of the dependent variable.
4. Normality - The residuals from the model should be normally distributed.

We accept these assumptions as they are supported by various universities and statistical institutions (JMP, 2024).

Regression Analysis

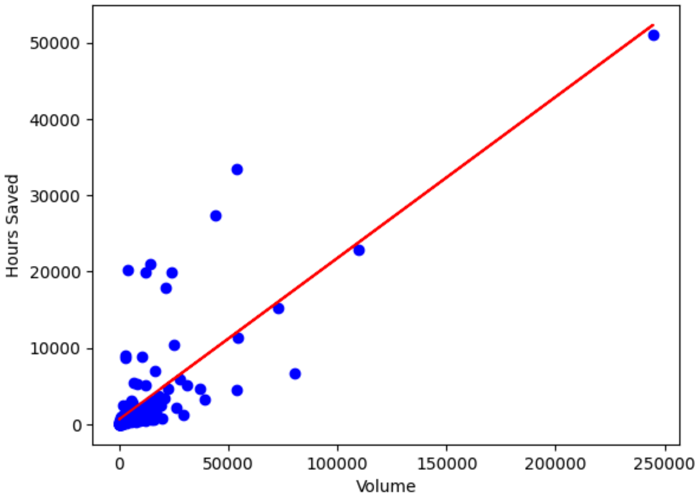
We want to take a look at the relationship between the volume of queue items and the amount of hours saved by RPA implementation. To begin with we created a correlation matrix between the two variables which resulted in a correlation of 78% (table 3.5). Correlation is simply a statistical measurement that explains the linear relationship between the two variables, but it does not explain the cause and effect of the relationship.

Table 3.5: Correlation Matrix for Volume of Queue and Saved Time (hours)

Correlation	**Volume**	**Saved Time (hours)**
Volume	1	0.78
Saved Time (hours)	0.78	1

In figure 3.1 we see a simple scatter plot and a linear regression line that predicts the hours saved based on the volume of completed queue items. We see that a lot of the data is centered around the left side of the diagram. This indicates that we are dealing with a non-linear relationship between the two variables, which means that we will have to transform the data to get a better understanding of their correlation.

Figure 3.1: Regression Line and Scatter Plot per Queue



Residuals

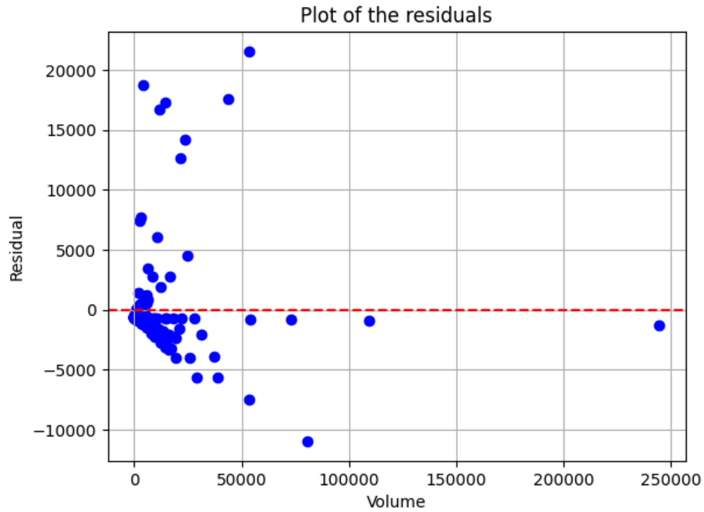
When looking at fitness of a model, how accurate the model is at predicting our data, a great way is to take a look into the residuals of the model. Residuals is an error term that is used to measure the variance in a dataset, which is not explained by the regression model. Residuals are the sum of the distances between the data points and the regression line.

One way of measuring the variance in the data is to calculate the Residual Sum of Squares (RSS). The lower the RSS is the better the model is fit to the data. RSS is calculated by taking the predicted value y_i, subtracting the actual value, $f(x_i)$, and then squaring that, such that the sum of these numbers provides a reliable measure of the distance of data points from the expected value, regardless of their positive or negative values.

$$RSS = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

For our first model we get a relatively high RSS, which is as expected when looking at the visual representation of the data in figure 3.2:

Figure 3.2: Residuals for Original Model



The residual plot in figure 3.2 shows the difference between the observed values in the dataset and the predicted values. The model predicts that the values should be along the red line above and the reality is that they are skewed to the right. For a perfect linear regression model and a normally distributed dataset, we would see the residuals randomly scattered around the red line, with no distinct pattern and along the entire range of the predicted values.

Transformation

As mentioned earlier, when the data is skewed this way we will have to transform the data to better represent the relationship between our variables. We chose to look at three types of transformation:

- **Log transformation** - A commonly used transformation for normalizing right-skewed distribution. A more normalized distribution will make our data analysis more reliable. Log transformation takes the logarithmic value of each data point, and can help linearize the relationship between variables.
- **Square root transformation** - Generally used to reduce skewness and further normalize the distribution and establish a better linear relationship. The transformation consists of taking the square root of the values.
- **Box-cox transformation** - A power transformation used to stabilize the variance and make the data more normalized. Its primary purpose is to normalize non-normally distributed data.

We performed a test for each of the new models after transformation. The results are shown in table 3.6:

Table 3.6: Test Results for Each Transformed Model

Type	**RSS**	**Adj. R-squared**	**P-value**	**Residual Standard Error**
Original model	2.78*10^9	65.8%	< 0.001	4 145.93
Log transformation	187.7	97.3%	< 0.001	1.08
Square root transformation	89 497.2	80.3%	< 0.001	23.50
Box-cox transformation	790.1	95.7%	< 0.001	2.21

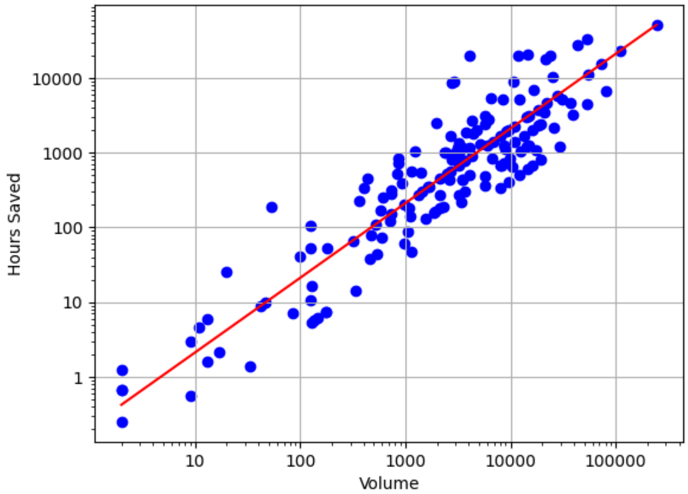
Explanation for each measurement:

- RSS - how big the difference between prediction and observed values are.
- Adjusted R-squared - a measure of how well fit the model is, the higher the number the better.
- P-value - a low p-value suggests that the findings are statistically significant.
- Residual Std Error - measuring the standard deviation of the residuals, low values indicates that predictions are better

From table 3.6, we see that the best results are achieved when using logarithmic transformation, and will therefore further explore the model obtained by this transformation.

Figure 3.3 shows our new model after log transforming both our variables, as this transformation has the overall best score on our tests.

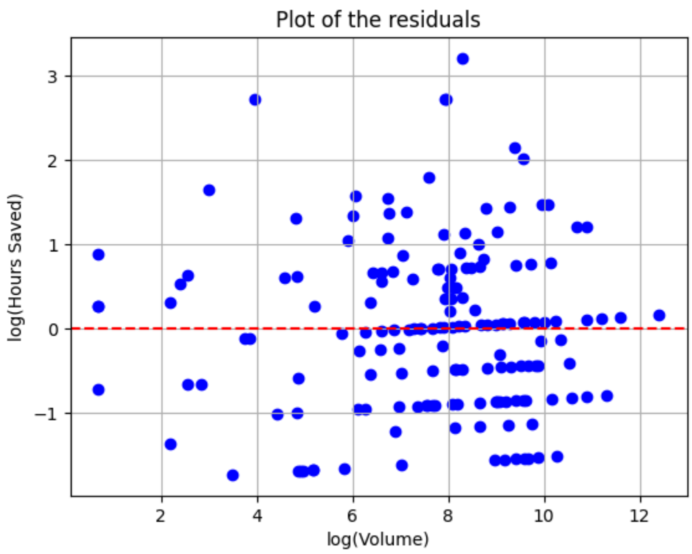
Figure 3.3: Regression Line for Logarithmic Transformed Model



We see that the linear relationship between our two variables is explained a lot better after log transformation, than the other two alternative transformations. It is important to note that we are now dealing with logarithmic transformed axes. This method of transformation is one of the most common transformations and widely supported as a means to better represent non-linear relations when dealing with skewed data (Feng et.al., 2014).

When we plot the residuals for our new model (figure 3.4), we observe that the residuals are better than for the original model. They are still a bit skewed, but closer to normally distributed, which in turn improves our model assumptions.

Figure 3.4: Residuals for Logarithmic Transformed Model



Linear regression serves as a valuable tool to understand the relationship between the volume of queue items and time saved by RPA implementation. However, before employing this method, it is crucial to ensure certain assumptions are met. These include linearity, homoscedasticity, independence, and normality of residuals.

3.4.3. Efficiency

Saved time as a measurement of the freed resources a company gets from automation of their processes are always a good measurement to look at, but should be accompanied by whether or not there is an increased efficiency by doing the automation. If we see a slower completion time, this might slow the overall workflow of the business down and in turn make them less efficient.

In our analysis we have defined efficiency as the reduced time it takes to complete a task. This is measured by taking the standard time, the time it would take a human worker to complete a certain task, and subtracting the work time for the robot, the time it actually took to complete the task.

In table 3.7 the efficiency per customer is calculated and measured. The table consists of the volume of tasks performed, the total hours of improvement and the average improvement per task.

Table 3.7: Efficiency Improvement by Customer

Customer (anonymized)	Volume of Tasks	Total Hours of Improvement	Average Improvement per Task, Measured in Minutes
FINANCE 2	982 885	241 003	14.71
FINANCE 1	388 208	117 369	18.14
FINANCE 3	185 655	12 923	4.18
REAL-ESTATE 1	43 587	12 805	17.63
RENOVATION 1	20 594	2 769	8.07
TRANSPORT 1	25 330	1 745	4.13
FINANCE 4	1 728	1 375	47.76
CONSTRUCTION 1	11 471	791	4.14
RENOVATION 2	2 483	516	12.49

CAR 1	5 358	374	4.19
CAR 2	980	60	3.72
CAR 3	352	14	2.50
CONSTRUCTION 2	282	11	2.50

We see that the average improvement per task for a company is above 2,5 minutes for all customers, and more often than not it is above 4 minutes. This means that on average, an increase in volume will improve efficiency by at least 2,5 minutes per task. We also have to point out an outlier in the customer anonymized as finance 4: while having the fourth lowest volume, they rank in the middle of the table and have an abnormally high average improvement. After further investigation we identified the cause for this abnormality to be that the customer has only three automated processes which have an average standard time above 60 minutes. Together with their low volume and the robot's way more efficient task-performance they end up saving 47.76 minutes per task on average.

This concludes our analysis on efficiency and in conclusion there seems to be a lot of evidence which point in the direction that automation of processes improves the amount of processes a company can handle and that these processes are done at a faster rate than a human worker would be able to do.

4. Results and Discussion

4.1. Overview of Data Findings

The analysis of our data revealed several key insights into the implementation and impact of Robotic Process Automation (RPA) in organizational settings. Our initial data analysis showed a positive skew indicating that the relationship between our variables was non-linear. This forced us to explore different variants of data transformation, where logarithmic transformation gave the most accurate depiction of the relationship between variables.

Through calculation and classification by alternative workforce increase, we began to enhance our understanding of the impact of RPA, revealing three distinct segments of customers with varying degrees of productivity increase. Furthermore, our classification by process complexity gave us some unexpected results around which type of process gave the best ratio between required development and resources, and the saved time. Our linear regression analysis unveiled a strong relationship between queue item volume and time saved, confirming our general hypothesis that higher volume gives higher reward.

These findings offer valuable insights for businesses navigating the terrain of RPA adoption, emphasizing the need for strategic alignment with organizational dynamics and careful consideration of process complexities.

4.2. Analysis of Results

4.2.1. Regression

Regression analysis was conducted to establish a model which predicts the dependent variable, time saved, based on the independent variable, volume of queue items. Our initial assumption was that this was going to be a linear relationship. After a look into the original model and residuals for said model, we realized that this was not the case. The goal was now to transform the variables in order to better explain the relationship between the dependent variable and the independent variable.

Through different transformations of our variables, we were able to establish a better model for our data. Ending up with a logarithmic transformed model, as it presents a more accurate insight into what values to expect from having a certain volume. The model we obtained has an overall better fit, had a lower RSS-value, and a low residual standard error.

4.2.2. Classification

Our classification analysis aimed to identify certain predictive variables within our own variables that influence our output. Our first attempt of classification was based on a new introduced variable, employee count, which was gathered from public registers. From these values we calculated our "alternative workforce increase"-index to classify the customers into three distinct groups. These productivity increases were not easy to base upon any of our quantifiable variables, but we did see a correlation between this index and the data provider's perceived degree of implementation for several customers. This correlation would have opened opportunities to explore several other analyses, but we did not see our correlation as strong enough to work with further. We will discuss why this is together with other limiting factors in chapter 4.6.

In our second attempt of classification we wanted to explore whether process complexity had an impact on saved time per task. Building our classification on the estimates for human workers to perform the task (standard time), we were able to split the processes into three distinct levels of complexity from simple to high complexity. Although we had our hypothesis that simple tasks would have the best ratio standard-saved time, as they would have the highest volume of cases, we observed a better ratio for the processes with medium complexity.

These findings could be impacted by several factors which could cause an erroneous conclusion. If the customers in our study had an overrepresentation of medium complex tasks, we could end up having an imbalanced dataset to begin with, and therefore draw a wrongful conclusion. We also see an abnormal spread of our base variable "standard time" which is likely caused by human perception of time. We will further discuss this in chapter 4.6.

4.3. Exceptions

Despite the overall consistency of our data, we observed some exceptions which we had to deal with. We had to remove certain values which can not be included when doing specific transformations - logarithmic transformation is one of these. In order to be able to perform this transformation, we had to exclude both volumes and standard time of 0. Volumes of 0 will not have any implications as these instances only emerged because the queue definitions provided by the data providers included processes which were inactive in the time period of our scope. A standard time of 0 has generally come from processes which have not been done by humans before. These processes are rather created as a step between different queues or to facilitate other processes that the human workers performed earlier, and would not result in a saved amount of work time.

4.4. Comparison with Theoretical Expectations

Although we did not have access to economic data regarding the companies' costs of RPA implementation, we have tried comparing our empirical findings with the theoretical framework established in the literature review as much as possible and used generalized examples where this is required. This limitation is mentioned in chapter 4.6. where we discuss its impact and how we solved it.

The Cost of Employment

A company's payroll is often seen as one of their highest expenditures, so when a company is making the decision of whether to implement RPA or not, there should be discussions about the amount of hours saved, which again should be calculated in terms of salary. From a purely isolated economic standpoint, the salary saved should be high enough that it covers the maintenance costs, and pays back the initial implementation cost over a desired period of time.

Theory of Comparative Advantage

Deciding how to allocate different resources is a decision which will impact a business' economic future to a high degree. When deciding this, companies and their decision makers often take a look at the opportunity costs. We found through our analysis that by letting human workers do these repetitive manual tasks, the companies lose out on workers who are fit to do more cognitive strenuous tasks, which robots are not yet fit to do, at the same time as they also lose out on efficiency. Structuring the company's workforce with both human and digital workers will in turn give them an advantage over other companies, who do not structure their business in the same way.

Economies of Scale

When implementing RPA into a company's workflow, it is important to notice that the implementation cost is going to be the same no matter the volume of items processed during its lifetime. Whether a process runs 10 times or 10 000 times, the implementation costs end up being the same. Looking at our regression analysis, we see that we have a process that runs close to 250 000 times. The implementation costs would have been the same if it only ran a tenth of the queue items, but the difference is that the company has been able to reduce their average costs for this task significantly by having a higher volume, thus achieving the advantages of economies of scale.

Theory of Constraints

Companies are always working with constraints and within bottlenecks in their day to day operations. When observing the efficiency of the implementation, the result was that on average these processes are becoming more efficient than they used to be, by being completed at a faster rate. Future or present bottlenecks might have been eliminated through automation and the general workflow has been optimized, leading to new bottlenecks emerging and giving companies the chance to further optimize their business.

A potential bottleneck that could emerge as a result of mass implementation is that there are too few robots to work the processes. The solution to this is however quite simple, compared to higher demand in the human workforce. Increasing the robotic workforce is more or less plug and play, where robots only perform tasks from a universal instruction, whereas increasing the human workforce requires more training.

4.5. Implications for RPA Adoption

The implications of our findings for the adoption and implementation of RPA are significant. Specifically, our classification by alternative workforce increase provides a perspective on the impact of RPA on workforce productivity. By grouping the customers into three distinct groups with varying degrees of productivity enhancement, our study underscores the importance of considering factors beyond employee count when assessing the economic benefits of implementing RPA-technology.

Our analysis of process complexity highlights how different approaches of implementing RPA have varying effects on time savings, emphasizing the potential for significant productivity gains in complex processes despite them being resource-intensive.

The exploratory analysis of different data transformation methods underscores the importance of paying attention to the misrepresentation from unprocessed data. By using logarithmic transformation we effectively showed improvement in our model fit, which provides practical guidance for researchers and practitioners in analyzing data related to RPA. These insights hold significant implications for organizational decision-making.

By understanding these economic and technical aspects, organizations can make more informed decisions regarding investment in RPA technology and resource allocation. The identification of process complexities and the potential for productivity gains in automating processes can inform strategic priorities and resource allocation within organizations.

4.6. Limitations

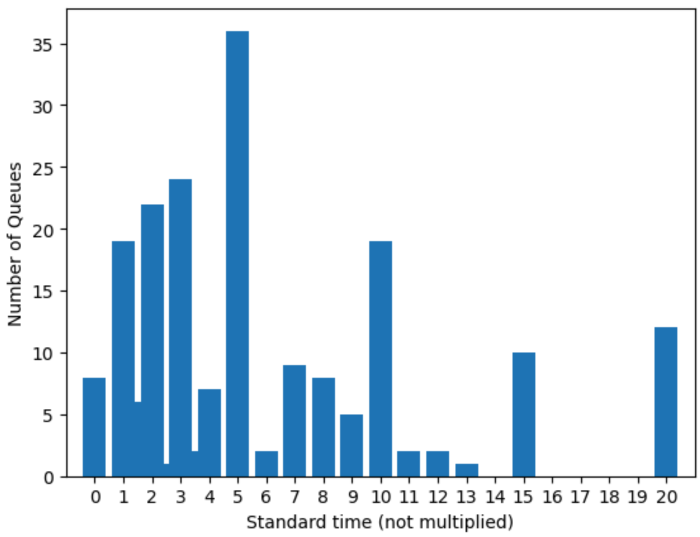
It is important to acknowledge the limitations of our study. When starting with this thesis we believed it would be easier to extract data from our data providers and that the data would be more balanced than the data we received. Some of the data was not in our hands before late March, which gave us only one month to do our data processing and analysis.

The variety and amount of the data has also come short to our expectations. We had hoped for a pool of up to 50 customers across multiple industries, but ended up with only 13 in a more narrow spectrum of industries. This has introduced problems regarding our dataset's validity and decreased our opportunities to generalize our findings to a greater population. If we were to find the same correlation in regards to degree of implementation and our "alternative workforce increase"-index in a greater population, this correlation could have been used as a substitute for degree of implementation. A company's degree of implementation is often hard to measure without extensive analysis, whereas our "alternative workforce increase"-index is easily measured and scaled to fit a wider range of companies implementing RPA.

We also wanted to analyze more of the economic aspects of RPA implementation, but as mentioned earlier there has been a reluctance to share data which could be viewed as business sensitive by the administrations. It is also difficult to measure some of these economic costs and benefits as the stages of exploration, development and maintenance often happen at the same time for different processes in a single team of RPA developers. We have however chosen to keep our economic theory in order to discuss these conditions as accurately as we can, without sharing business sensitive data, by using generalized examples instead of real data.

A fourth limiting factor is how the standard times (before multiplication) often are "guesstimated" of the actual time human workers use to perform the task. This may be because accurate measurement of the task could lead to the worker performing the task faster than they would do in a realistic situation. It is also human nature to round off estimates to "nice" numbers such as 1, 2, 3, 5, 10, and so on, instead of 4 or 9 (Jain, 2020). This rounding contributes to a dataset which is less accurate and could generate invalid conclusions if not taken into account. For example, in our dataset we see an overrepresentation of the numbers 1, 2, 3, 5, 10, 15 and 20, compared to the numbers in between (figure 4.1). The expected similar results would be a normally distributed curve around 5 minutes.

Figure 4.1: Distribution of "Guesstimated" Standard Time (not multiplied) [0-20]



4.7. Further Research

Future research could build upon our findings by dedicating more time to qualitative data in order to better explain the greater spectrum of benefits from implementing RPA technology. If we had dedicated more time to dive deeper into this topic, we would have conducted several structured interviews where we analyzed the effects employees and management experienced before, during and after the implementation. This would give insight into how the time saved by RPA would influence other work related areas, as well as which possibilities have emerged after the implementation. When RPA is implemented, a lot of human competency is released from performing repetitive manual tasks, giving them more time to participate in more complex decisions requiring human discussion and understanding. It is difficult to measure the effect of having more brain power as generalizing human competency proves difficult to scale up to a greater population. The alternative use of this competency varies a lot for each case in each business, making it even more difficult to measure it without extensive analysis.

Another perspective we would like to research further is the value added as customer satisfaction increases because their cases get processed faster by robots. By automating the processes towards the customer, they often get a more elaborate and correct answer as to what a single case manager would have been able to deliver. To further research this topic, a customer center which had implemented RPA solutions to answer the most common cases, would be a fit candidate for a case study. Another good candidate would be any provider of digital services which requires some kind of validation before goods are delivered to the customer. One example could be a bank which gives out loans to customers based on certain criterias where the case manager would have to perform lookups in several systems.

Our hypothesis that “alternative workforce increase” may be used to measure degree of implementation is something we would like to perform further analyses on to confirm our findings in a wider dataset. If further research shows that defining the degree of implementation with this quantifiable variable is valid to some degree, it would open up for deeper research into the business strategic topics of RPA implementation on a much greater scale.

5. Conclusion

The purpose of this bachelor thesis was to get a better foundation to answer the research question “How can Robotic Process Automation-technology enhance a company’s efficiency regarding resources and costs?”.

Through our analysis we have highlighted several aspects of how RPA-technology has enhanced the participating companies’ efficiency and which impacts it can have in organizational settings. By uncovering insights in the relationship between RPA adoption, workforce productivity, and process complexity, we have elevated our understanding of the dynamics surrounding RPA deployment.

Our analysis showed that about 23% of the observed customers experienced that they would have to increase their current workforce by 20-50% if they were not to automate part of their processes. On average the customer saves 21.36 full time equivalents, however, there is a huge spread in effectiveness and size amongst the observed customers. 30% of our participating customers save more than 8 full time equivalents each year.

In conclusion, we see a lot of benefits with implementing RPA in a variety of industries. As projected by our literature review, our study shows that the enhancement of a company’s efficiency is likely to be significant, both financially and resource-wise. However, it is important to take several factors into consideration when implementing RPA. By learning from the findings in this thesis an implementation of RPA should result in a significant increase in efficiency from today's levels.

References

- Feng, C. et.al. (2014) *Log-Transformation and Their Implications for Data Analysis*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120293/ (Accessed: 4 April 2024)
- Goldratt, E.M. (1984) *The Goal*. Great Barrington, MA: North River Press.
- Hayes, A. (2023) *What is Comparative Advantage?* Available at: https://www.investopedia.com/terms/c/comparativeadvantage.asp (Accessed: 28 January 2024)
- Jain, G. et.al. (2020) *Revisiting attribute framing: The impact of number roundedness on framing*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0749597820303617 (Accessed: 17 april 2024)
- JMP (2024) *Regression Model Assumptions*. Available at: https://www.jmp.com/en_us/statistics-knowledge-portal/what-is-regression/simple-linear-regression-assumptions.html (Accessed: 26 February 2024)
- Kenton, W (2022) *Economies of Scale: What Are They and How Are They Used?* Available at: https://www.investopedia.com/terms/e/economiesofscale.asp (Accessed: 28 January 2024)
- King, B. (2022) *AI clocks in: how bots are streamlining restaurant operations*. Available at: https://www.ibm.com/downloads/cas/MWXG4VOL (Accessed: 28 January 2024)
- Kirkwood, G. (2019) *Impact of RPA on Employee Engagement*. Available at: https://www.uipath.com/blog/rpa/impact-of-rpa-on-employee-engagement-forrester (Accessed: 22 January 2024).
- Skurtveit, S. & Sørbø, K. (2023) *Let som han skulle teste systemet – svindla banken for 75 millionar kroner*. Available at: https://www.nrk.no/trondelag/bankvikar-tiltalt-for-underslag-av-75-millionar-kroner-forklarar-seg-i-retten-1.16657675 (Accessed: 4 February 2024)
- Spencer, R. (2022a) *A game changer in the pharmaceutical industry*. Available at: https://www.ibm.com/downloads/cas/B8XYAXML (Accessed: 28 January 2024)
- Spencer, R. (2022b) *Getting 3 ½ hours back every day*. Available at: https://www.ibm.com/downloads/cas/NLY0BRDZ (Accessed: 28 January 2024)

- Stobierski, T. (2019) *How to do a cost-benefit analysis & why it's important.* Available at: https://online.hbs.edu/blog/post/cost-benefit-analysis (Accessed: 22 January 2024).
- UiPath (2024) *What is robotic process automation?*. Available at: https://www.uipath.com/rpa/robotic-process-automation (Accessed: 22 January 2024).
- Watson, J. & Wright, D. (2017) *The 3rd Annual Global RPA Survey Report*. Available at: https://www2.deloitte.com/us/en/pages/operations/articles/global-robotic-process-automation-report.html (Accessed: 27 January 2024)

Appendix

Appendix A - Sample SQL-query to Extract Data from BluePrism

```
SELECT items.[id]
      ,[name] AS 'queue'
      ,[attempt]
      ,[status]
      ,[loaded]
      ,[completed]
      ,[worktime]
      ,[finished]
  FROM [Blueprism_prod].[dbo].[BPAWorkQueueItem] AS items
    LEFT JOIN [Blueprism_prod].[dbo].[BPAWorkQueue] AS queues
      ON (items.queueid = queues.id)
  WHERE [finished] > '2023-01-01'
  AND [finished] < '2024-01-01'
```

Appendix B - Python Notebook for Verifiability

Imports

In [1]:

```
# Data formatting
import numpy as np
import pandas as pd
import math
import json

# Plotting diagrams
import matplotlib.pyplot as plt
from matplotlib.ticker import ScalarFormatter

# Regression and data transformation
from sklearn.linear_model import LinearRegression
from scipy.stats import linregress
from scipy.stats import boxcox

# Models and evaluation
import statsmodels.api as sm
```

In [2]:

```
# Connect to shared file storage in Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount
("/content/drive", force_remount=True).
```

In [3]:

```
# Settings for Pandas
pd.options.display.float_format = '{:.2f}'.format
```

Statics

In [4]:

```
# Define filepaths
data_folder = 'drive/MyDrive/V2024/AF3035/Data/'

customer_anonymization_filename = 'customer_anonymization.json'

data_smn_filename = 'QueueData - All.csv'
queues_smn_filename = 'Queue Definitions.xlsx'

data_rria_filename = 'rria_customers_data.xlsx'
queues_rria_filename = 'rria_customers_queues.xlsx'

customer_info_file = 'Customer_info.xlsx'
```

In [5]:

```
# Define headers for headerless file
smn_headers = ['ID', 'Queue', 'Attempt', 'Status', 'Loaded', 'Completed', 'Worktime', 'F
inished']
```

In [6]:

```
# Define man-year standard
man_year_hours = 1650
```

```
# Define standard time multiplier
st_multiplier = 2.5
```

In [7]:

```
# Function: Classification based on alternative workforce increase
def customer_classification_by_size(alternative_workforce_increase: float) -> str:
  '''Input: alternative workforce increase'''
  if alternative_workforce_increase < 0.01: return 'Low increase'
  elif alternative_workforce_increase < 0.20: return 'Medium increase'
  else: return 'High increase'
```

In [8]:

```
# Function: anonymize customers
with open(data_folder+customer_anonymization_filename, 'r') as f:
  rename_dict = json.load(f)

def anonymize_customers(df: pd.DataFrame) -> pd.DataFrame:
  '''Note: The input DataFrame must contain column 'Customer'.'''
  for index, row in df.iterrows():
    new_name = rename_dict[row['Customer']]
    df.loc[index,'Customer'] = new_name

  return df
```

Load and Format Data

In [9]:

```
# Import data from csv and excel files
df_smn = pd.read_csv(data_folder+data_smn_filename, sep=';', names=smn_headers)
queues_smn = pd.read_excel(data_folder+queues_smn_filename)

df_rria = pd.read_excel(data_folder+data_rria_filename)
queues_rria = pd.read_excel(data_folder+queues_rria_filename)
```

In [10]:

```
# Set common columns for data
df_smn.drop(columns=['Finished', 'Completed'], inplace=True)

df_rria['Attempt'] = 1 # To account for differences in RPA-systems
df_rria.drop(columns=['Customer'], inplace=True)
df_rria['Worktime'] = df_rria['Worktime']*60 # To convert from minutes to seconds
```

In [11]:

```
# Set common columns for queues
queues_rria.rename(columns={'QueueName': 'Queue'}, inplace=True)

queues_rria.drop(columns='Unfactorized time', inplace=True)
```

In [12]:

```
# Merge dataframes for common processing
df = pd.concat([df_smn, df_rria], ignore_index = True)

all_queues = pd.concat([queues_smn, queues_rria], ignore_index = True)
```

In [13]:

```
all_queues = anonymize_customers(all_queues)
```

Preprocessing

In [14]:

In [14]:

```
# Sort DataFrame by ID and filter out failed attempts
df_sorted = df.sort_values(by="ID")
try:
  max_attempt_indices = pd.read_pickle(data_folder+'max_attempt_indices.pkl')
except:
  max_attempt_indices = df_sorted.groupby('ID')['Attempt'].idxmax()

  pd.to_pickle(max_attempt_indices, data_folder+'max_attempt_indices.pkl')

# Use these indices to filter the DataFrame
data = df_sorted.loc[max_attempt_indices]
```

In [15]:

```
# Join merge DataFrames data and queues
queue_count = data.groupby('Queue').agg({'ID':'count', 'Worktime':'sum'}).reset_index()

queue_count = queue_count.rename(columns={'ID': 'Volume'})

queues = all_queues.merge(queue_count, on='Queue')
```

In [16]:

```
#Split Finance 5 to 40% Finance 1 and 60% Finance 2 as instructed by data provider
new_queues = pd.DataFrame()
for index, q in queues.iterrows():
  if q['Customer'] == 'FINANCE 5':
    volume = q['Volume']
    work = q['Worktime']

    #Creating customer Finance 1
    Fin_one = q.copy()
    Fin_one['Volume'] = math.floor(volume * 0.4)
    Fin_one['Worktime'] = math.floor(work * 0.4)
    Fin_one['Customer'] = 'FINANCE 1'
    new_queues = new_queues._append(Fin_one, ignore_index = True)

    # Creating customer Finance 2
    Fin_two = q.copy()
    Fin_two['Volume'] = math.ceil(volume * 0.6)
    Fin_two['Worktime'] = math.ceil(work * 0.6)
    Fin_two['Customer'] = 'FINANCE 2'
    new_queues = new_queues._append(Fin_two, ignore_index = True)

  else:
    new_queues = new_queues._append(q, ignore_index = True)

queues = new_queues.copy()
```

In [17]:

```
# Calculate saved time
queues['Saved time (minutes)'] = queues['Standard time'] * queues['Volume']
queues['Saved time (hours)'] = queues['Saved time (minutes)'] / 60
queues['Saved time (man-years)'] = queues['Saved time (hours)'] / man_year_hours
```

In [18]:

```
# Group queues by customer
customer_volume = queues.groupby('Customer').agg({'Volume': 'sum', 'Saved time (man-year
s)': 'sum'}).reset_index()
```

In [19]:

```
customer_volume
```

Out[19]:

	Customer	Volume	Saved time (man-years)

	Customer	Volume	Saved time (man-years)
0	CAR 1	5358	0.23
1	CAR 2	980	0.04
2	CAR 3	352	0.01
3	CONSTRUCTION 1	11471	0.48
4	CONSTRUCTION 2	282	0.01
5	FINANCE 1	388201	83.97
6	FINANCE 2	982892	169.64
7	FINANCE 3	185655	10.83
8	FINANCE 4	1728	0.96
9	REAL-ESTATE 1	43587	8.47
10	RENOVATION 1	20594	1.68
11	RENOVATION 2	2483	0.31
12	TRANSPORT 1	25330	1.06

Exploratory Analysis

Classification

Company Size

In [20]:

```
# Import customer information
customer_info = pd.read_excel(data_folder+customer_info_file)
```

In [21]:

```
# Anonymize Customer Info
customer_info = anonymize_customers(customer_info)
```

In [22]:

```
# Merge in queue_count and item_volume
customer_df = customer_info.merge(customer_volume, on='Customer')
```

In [23]:

```
# Create implementation index
customer_df['Alternative Workforce Increase'] = customer_df['Saved time (man-years)']/cu
stomer_df['Employees']
```

In [24]:

```
# Classify alternative workforce increase
customer_df['Classification'] = customer_df['Alternative Workforce Increase'].apply(cust
omer_classification_by_size)
```

In [25]:

```
customer_df
```

Out[25]:

	Customer	Employees	Volume	Saved time (man-years)	Alternative Workforce Increase	Classification
0	CONSTRUCTION 1	81	11471	0.48	0.01	Low increase
1	REAL-ESTATE 1	260	43587	8.47	0.03	Medium increase

	Customer	Employees	Volume	Saved time (man-years)	Alternative Workforce Increase	Classification
2	FINANCE 3	51	185655	10.83	0.21	High increase
3	RENOVATION 2	103	2483	0.31	0.00	Low increase
4	FINANCE 1	180	388201	83.97	0.47	High increase
5	CAR 1	132	5358	0.23	0.00	Low increase
6	TRANSPORT 1	19	25330	1.06	0.06	Medium increase
7	CONSTRUCTION 2	259	282	0.01	0.00	Low increase
8	FINANCE 2	450	982892	169.64	0.38	High increase
9	CAR 2	121	980	0.04	0.00	Low increase
10	RENOVATION 1	120	20594	1.68	0.01	Medium increase
11	CAR 3	56	352	0.01	0.00	Low increase
12	FINANCE 4	20	1728	0.96	0.05	Medium increase

Complexity

In [26]:

```
queues = queues[queues['Standard time'] != 0]
queues = queues[queues['Volume'] != 0]


# Adding the complexity variable
cond = [
    queues['Standard time'] <=5,
    (queues['Standard time'] >5) & (queues['Standard time'] <= 20),
    queues['Standard time'] > 20
]
comp = ['Simple', 'Medium Complexity', 'Complex']

queues['Complexity'] = np.select(cond, comp)
```

In [27]:

```
# Calculate saved time
queues['Saved time (minutes)'] = queues['Standard time'] * queues['Volume']
queues['Saved time (hours)'] = queues['Saved time (minutes)'] / 60
queues['Saved time (man-years)'] = queues['Saved time (hours)'] / man_year_hours
```

Regression

In [28]:

```
# Correlation matrix
corr = queues[['Volume','Saved time (hours)']]

corr_matrix = corr.corr()
corr_matrix
```

Out[28]:

	Volume	Saved time (hours)
Volume	1.00	0.78
Saved time (hours)	0.78	1.00

In [29]:

```
# Initialize result-table for different models
model_results = pd.DataFrame(columns=['Type','RSS','Adj. R-squared','P-value','Residual S
tandard Error'])
```

Linear Regression

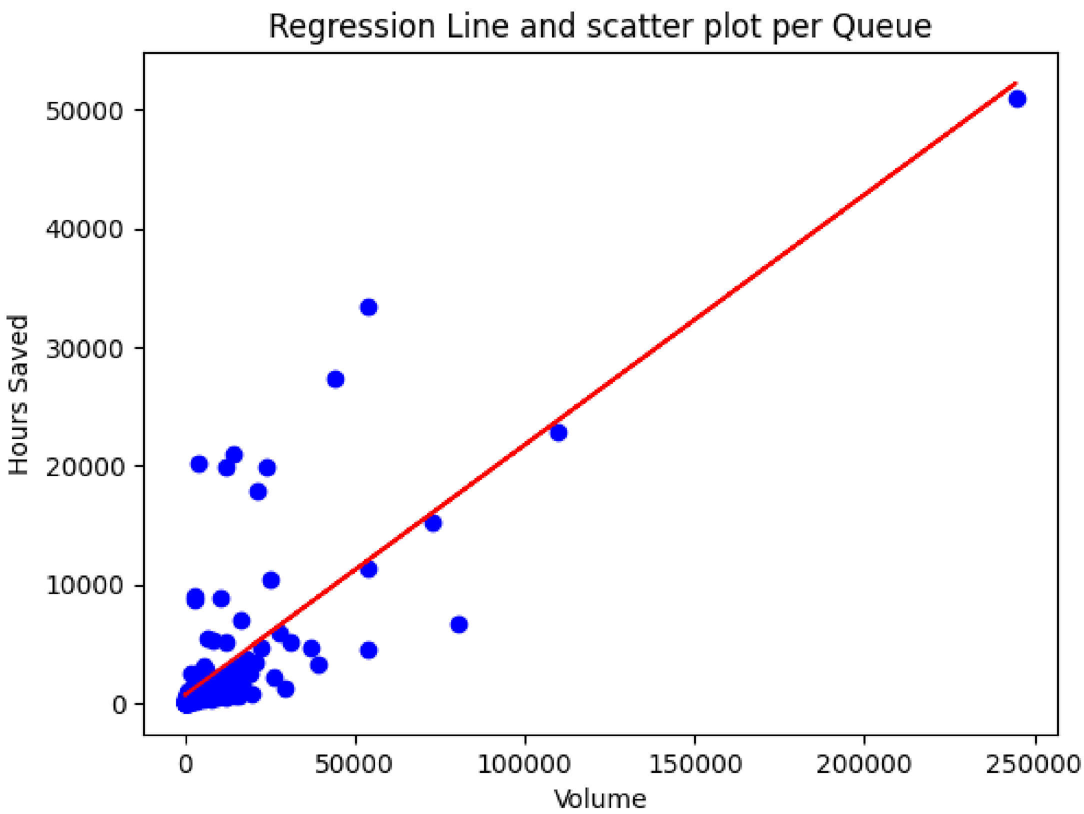
In [30]:

```
# Copy queues and filter for values = 0
reg = queues.copy()
reg = reg[reg['Standard time'] != 0]
reg = reg[reg['Volume'] != 0]
```

Simple regression and redisuals for unprocessed data

In [31]:

```
# Plot for simple regression
X1 = reg[['Volume']]
y1 = reg['Saved time (hours)']

original_model = LinearRegression()

original_model.fit(X1,y1)

original_pred = original_model.predict(X1)

fig, ax = plt.subplots()

plt.scatter(X1, y1, color = 'b', label = 'Scatter plot')
plt.plot(X1, original_pred, color = 'r', label = 'RegLine')

plt.title('Regression Line and scatter plot per Queue')
plt.xlabel('Volume')
plt.ylabel('Hours Saved')

plt.show()
```



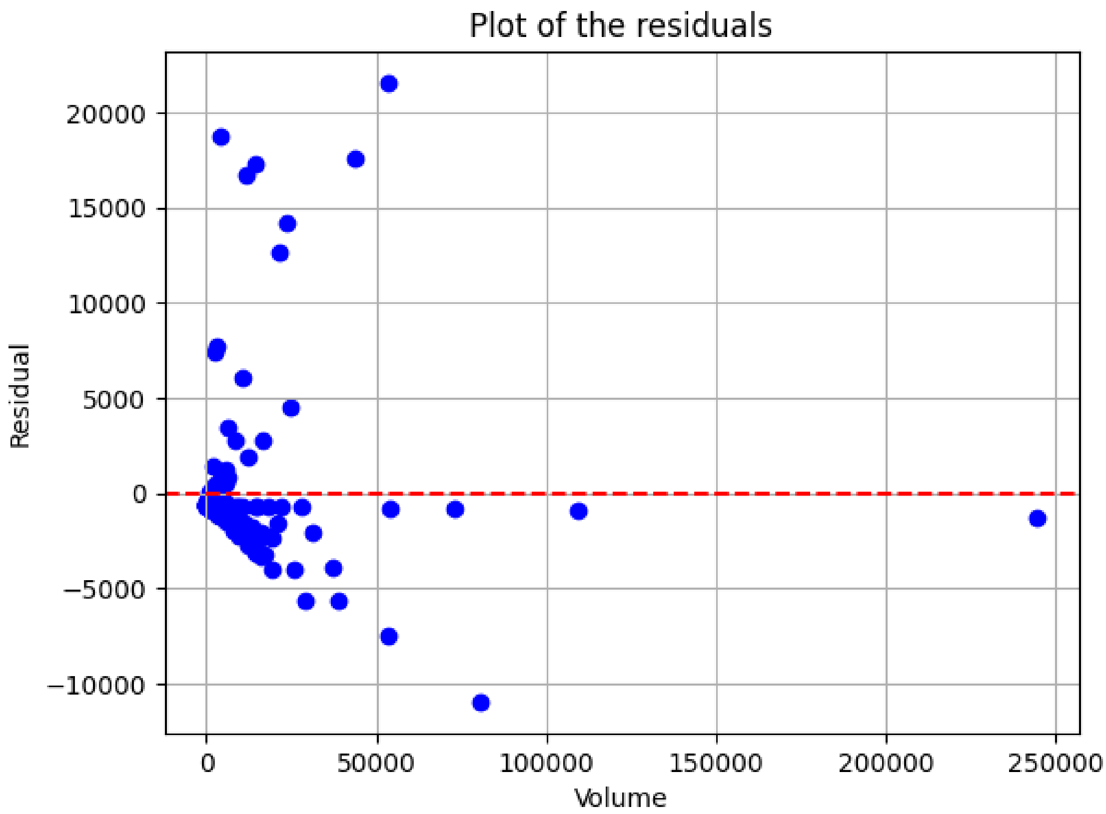
In [32]:

```
# Append result to model_results
original_model = sm.OLS(y1, X1).fit()

model_results = model_results._append({
    'Type':'Original model',
    'RSS':(original_model.resid**2).sum(),
```

```
    'Adj. R-squared':original_model.rsquared_adj,
    'P-value':original_model.f_pvalue,
    'Residual Standard Error':np.sqrt(original_model.mse_resid)
}, ignore_index=True)
```

In [33]:

```
# Plot for residuals
original_res = y1 - original_pred

plt.scatter(X1, original_res, color = 'b')
plt.axhline(y=0, color = 'r', linestyle = '--')
plt.title('Plot of the residuals')
plt.xlabel('Volume')
plt.ylabel('Residual')
plt.grid()
plt.show()
```



Data Transformation

Log transformation

In [34]:

```
# Create log transformed values for volume and hours saved
reg['log_Vol'] = np.log(reg['Volume'])
reg['log_hours'] = np.log(reg['Saved time (hours)'])

X2 = reg[['log_Vol']]
y2 = reg['log_hours']

model_log = LinearRegression()

model_log.fit(X2,y2)

pred_log = model_log.predict(X2)
```

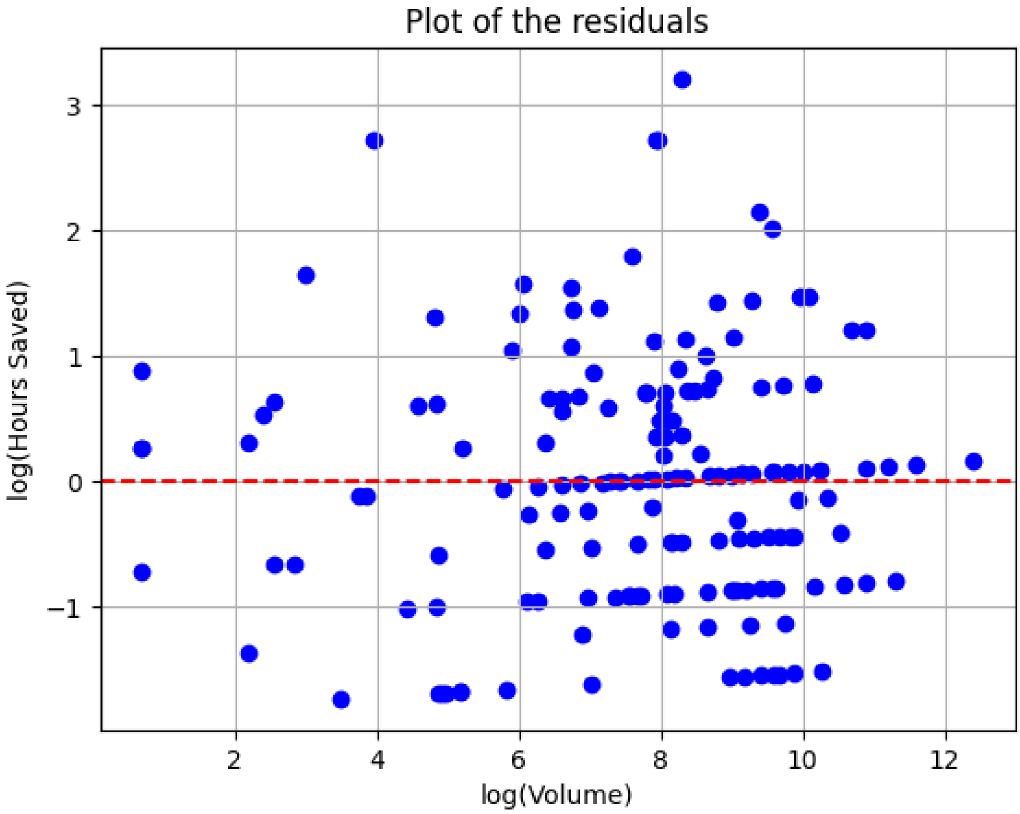
In [35]:

```
# Append result to model_results
model_log = sm.OLS(y2, X2).fit()
```

```
model_results = model_results._append({
    'Type':'Logarithmic Transformation',
    'RSS':(model_log.resid**2).sum(),
    'Adj. R-squared':model_log.rsquared_adj,
    'P-value':model_log.f_pvalue,
    'Residual Standard Error':np.sqrt(model_log.mse_resid)
}, ignore_index=True)
```

In [36]:

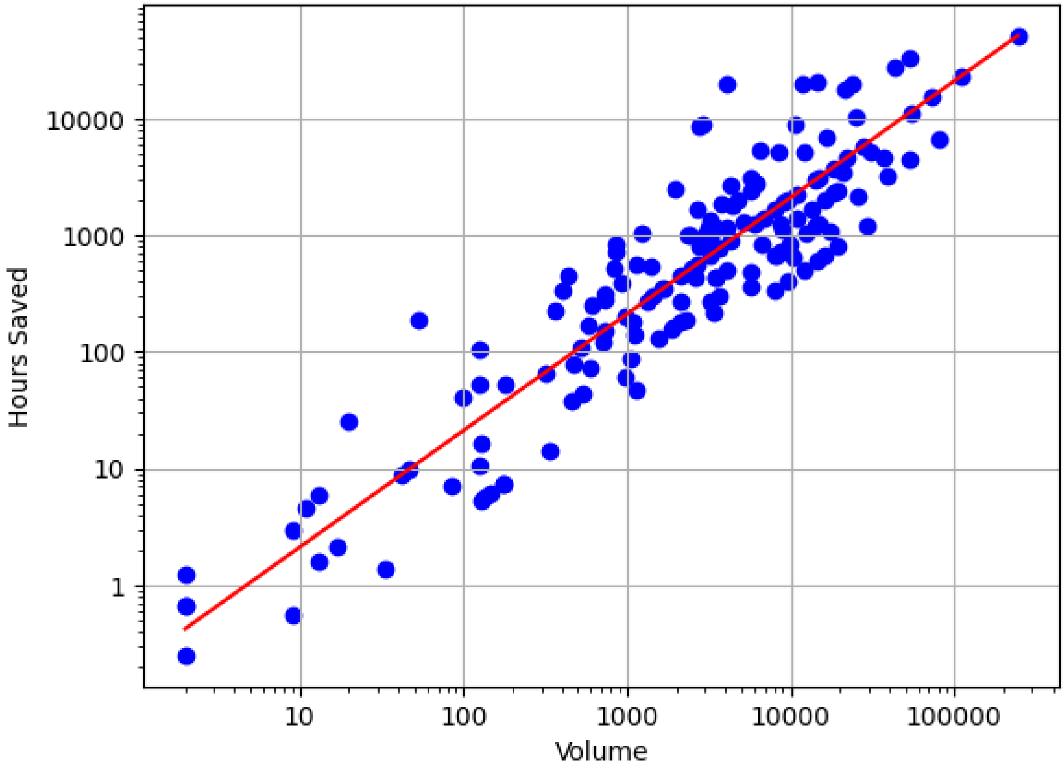
```
# Plot of residuals for log transformed model
res_log = y2 - pred_log

plt.scatter(X2, res_log, color = 'b')
plt.axhline(y=0, color = 'r', linestyle = '--')
plt.title('Plot of the residuals')
plt.xlabel('log(Volume)')
plt.ylabel('log(Hours Saved)')
plt.grid()
plt.show()
```



In [37]:

```
# Plot for showing logarithmic relationship
X3 = reg['Volume']
y3 = reg['Saved time (hours)']


fig, ax = plt.subplots()

plt.scatter(X3, y3, color = 'b', label = 'Scatter plot')


slope, intercept, r_value, p_value, std_err = linregress(X3, y3)

plotrange = range(reg['Volume'].min(),reg['Volume'].max())

# Calculate regression line
regression_line = slope * plotrange

# Plot regression line
plt.plot(plotrange, regression_line, color='red', label='Regression Line')
```

```
plt.xscale('log') # Changes the increase on the graph
plt.yscale('log') # Changes the increase on the graph


for axis in [ax.xaxis, ax.yaxis]:
    axis.set_major_formatter(ScalarFormatter())

plt.title('Regression Line Showing Logarithmic Releationship')
plt.xlabel('Volume')
plt.ylabel('Hours Saved')
plt.grid()
plt.show()
```

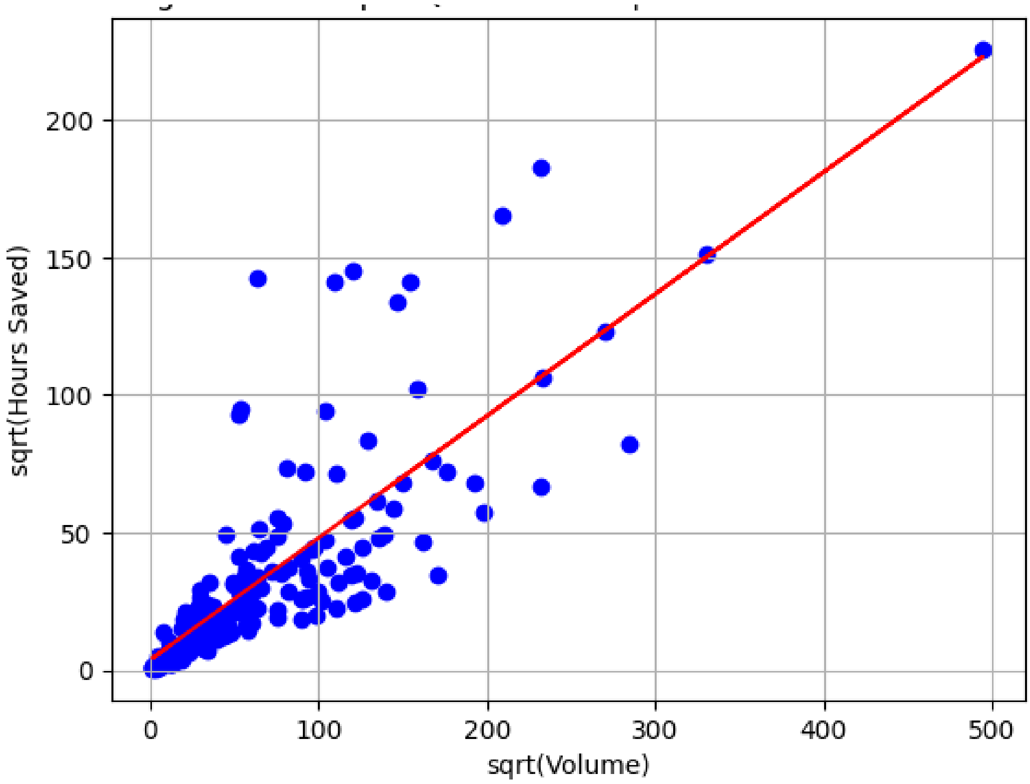


Square root transformation

In [38]:

```
reg['sqrt_Vol'] = np.sqrt(reg['Volume'])
reg['sqrt_hours'] = np.sqrt(reg['Saved time (hours)'])

X4 = reg[['sqrt_Vol']]
y4 = reg['sqrt_hours']

model_sqrt = LinearRegression()

model_sqrt.fit(X4, y4)

sqrt_pred = model_sqrt.predict(X4)

plt.scatter(X4, y4, color = 'b', label = 'Scatter plot')

plt.plot(X4, sqrt_pred, color = 'r', label = 'RegLine')

plt.title('Regression line per Queue with square root transformation')
plt.xlabel('sqrt(Volume)')
plt.ylabel('sqrt(Hours Saved)')
plt.grid()
plt.show()
```

Regression line per Queue with square root transformation



In [39]:

```
# Append result to model_results
model_sqrt = sm.OLS(y4, X4).fit()

model_results = model_results._append({
    'Type':'Square Root Transformation',
    'RSS':(model_sqrt.resid**2).sum(),
    'Adj. R-squared':model_sqrt.rsquared_adj,
    'P-value':model_sqrt.f_pvalue,
    'Residual Standard Error':np.sqrt(model_sqrt.mse_resid)
}, ignore_index=True)
```

Box-Cox Transformation

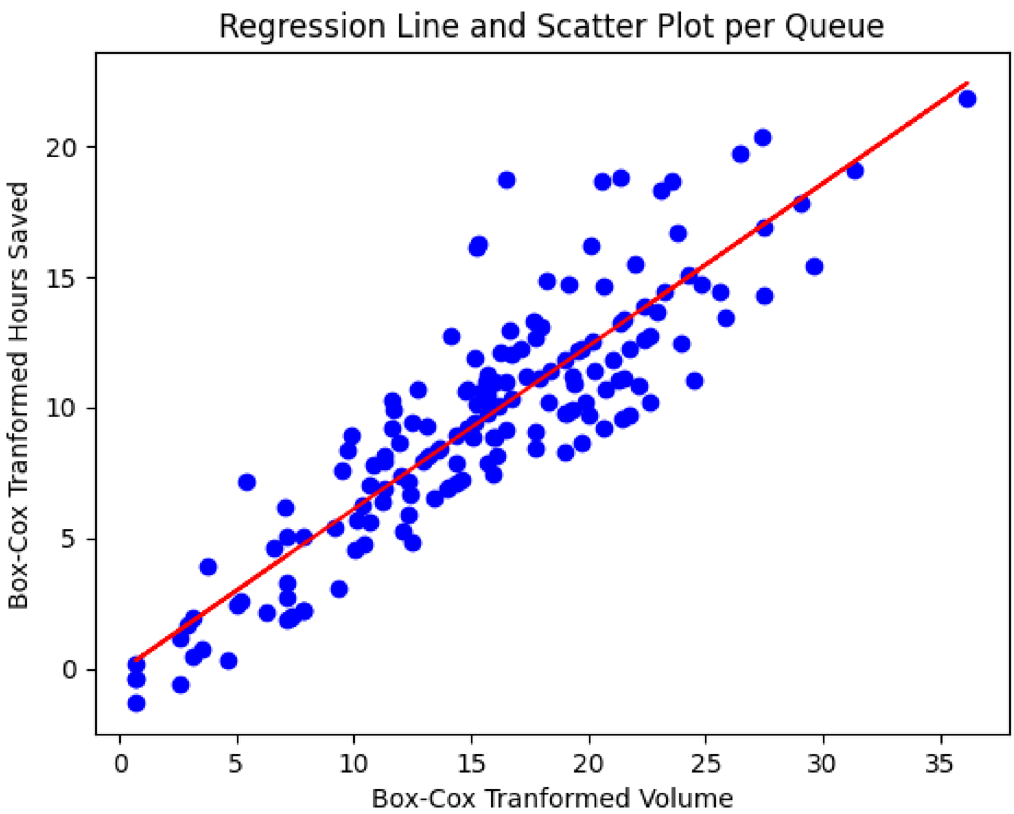
In [40]:

```
# Doing box-cox transformation on our variables
transformed_x, lambda_x = boxcox(reg['Volume'])

transformed_y, lambda_y = boxcox(reg['Saved time (hours)'])

transdf = pd.DataFrame({'TransVol': transformed_x, 'TransHours': transformed_y})
```

In [41]:

```
# Model and plot for box-cox
X5 = transdf[['TransVol']]
y5 = transdf['TransHours']

model_box = LinearRegression()

model_box.fit(X5,y5)

box_cox_pred = model_box.predict(X5)

fig, ax = plt.subplots()

plt.scatter(X5, y5, color = 'b', label = 'Scatter plot')
plt.plot(X5, box_cox_pred, color = 'r', label = 'RegLine')

plt.title('Regression Line and Scatter Plot per Queue')
plt.xlabel('Box-Cox Tranformed Volume')
plt.ylabel('Box-Cox Tranformed Hours Saved')
```

```
plt.show()
```



In [42]:

```python
# Append result to model_results
model_box_cox = sm.OLS(y5, X5).fit()

model_results = model_results._append({
    'Type':'Box-Cox Transformation',
    'RSS':(model_box_cox.resid**2).sum(),
    'Adj. R-squared':model_box_cox.rsquared_adj,
    'P-value':model_box_cox.f_pvalue,
    'Residual Standard Error':np.sqrt(model_box_cox.mse_resid)
}, ignore_index=True)
```

Comparison

In [43]:

```python
# Format results
formatted_model_results = model_results.copy()
formatted_model_results['RSS'] = formatted_model_results['RSS'].map('{:.1f}'.format)
formatted_model_results['Adj. R-squared'] = formatted_model_results['Adj. R-squared'].map
('{:.3f}'.format)
formatted_model_results['P-value'] = formatted_model_results['P-value'].map('{:.3f}'.form
at)
```

In [44]:

```python
# Print results
formatted_model_results
```

Out[44]:

	Type	RSS	Adj. R-squared	P-value	Residual Standard Error
0	Original model	2784576651.0	0.656	0.000	4145.93
1	Logarithmic Transformation	187.7	0.973	0.000	1.08
2	Square Root Transformation	89497.2	0.803	0.000	23.50

3	Box-Cox Transformation	790.1	0.957	0.000	2.21
	Type	RSS	Adj. R-squared	P-value	Residual Standard Error

Efficiency

In [45]:

```
eff_df = queues.copy()
```

In [46]:

```
eff_df = eff_df.groupby('Customer').agg({'Saved time (minutes)': sum, 'Volume': sum, 'Wo
rktime': sum}).reset_index()
```

In [47]:

```
# Defining Hours faster a robot is for each Customer, and the average time efficiency for
each customer
eff_df['Worktime (minutes)'] = eff_df['Worktime']/60
eff_df['Efficency'] = eff_df['Saved time (minutes)'] - eff_df['Worktime (minutes)']
eff_df['Efficency_hours'] = eff_df['Efficency']/60
eff_df['Average Time Faster in min'] = eff_df['Efficency']/eff_df['Volume']
```

In [48]:

```
eff_df
```

Out[48]:

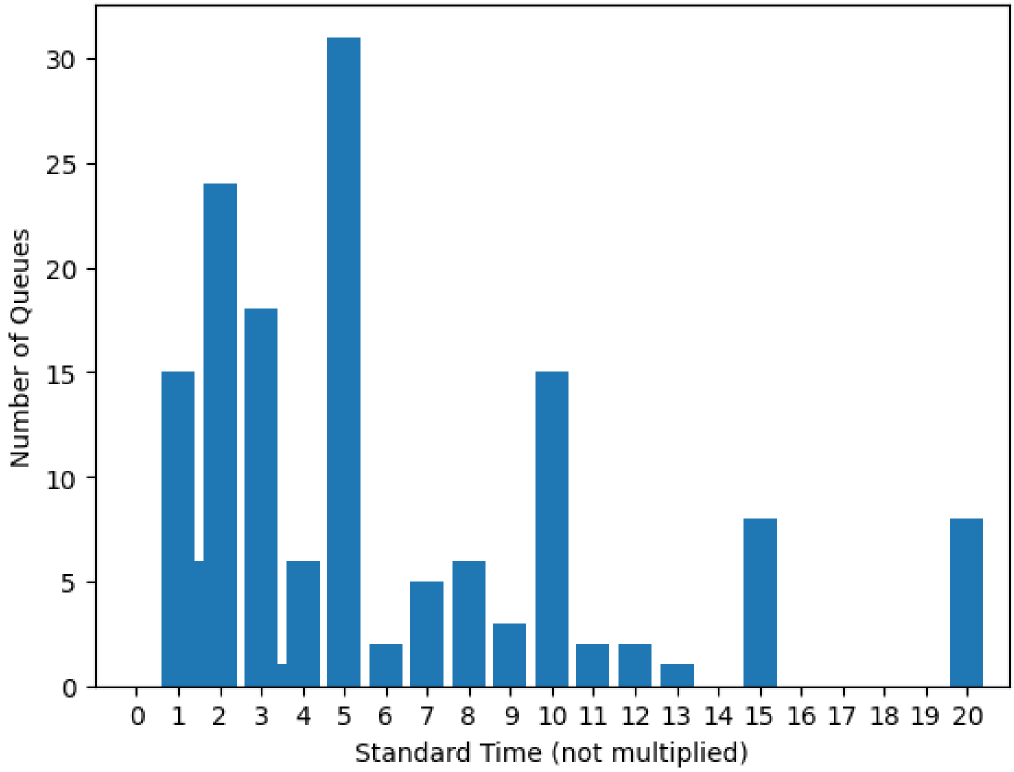
	Customer	Saved time (minutes)	Volume	Worktime	Worktime (minutes)	Efficency	Efficency_hours	Average Time Faster in min
0	CAR 1	22496.25	5358	3319.80	55.33	22440.92	374.02	4.19
1	CAR 2	3675.00	980	1473.60	24.56	3650.44	60.84	3.72
2	CAR 3	880.00	352	10.80	0.18	879.82	14.66	2.50
3	CONSTRUCTION 1	47482.50	11471	865.20	14.42	47468.08	791.13	4.14
4	CONSTRUCTION 2	705.00	282	19.20	0.32	704.68	11.74	2.50
5	FINANCE 1	8312637.50	388201	76230572.00	1270509.53	7042127.97	117368.80	18.14
6	FINANCE 2	16794487.50	982891	140047108.00	2334118.47	14460369.03	241006.15	14.71
7	FINANCE 3	1072226.25	185655	17807852.00	296797.53	775428.72	12923.81	4.18
8	FINANCE 4	95370.00	1728	770643.00	12844.05	82525.95	1375.43	47.76
9	REAL-ESTATE 1	838192.50	42140	4108956.00	68482.60	769709.90	12828.50	18.27
10	RENOVATION 1	166677.50	20594	28974.00	482.90	166194.60	2769.91	8.07
11	RENOVATION 2	31037.50	2483	1989.60	33.16	31004.34	516.74	12.49
12	TRANSPORT 1	105067.50	25330	21512.40	358.54	104708.96	1745.15	4.13

Additional Calculations

In [49]:

```
# Distribution of "Guesstimated" standard time
distribution = queues.groupby('Standard time').agg({'Queue':'count'}).reset_index()

distribution['Guesstimate'] = distribution['Standard time']/st_multiplier

plt.bar(distribution['Guesstimate'],distribution['Queue'])
plt.xlim(-1,21)
plt.xticks(range(0,21,1))
plt.xlabel('Standard Time (not multiplied)')
```

```
plt.ylabel('Number of Queues')
plt.show()
```



Describe

Volume, Queue, Standard time, saved time

In [50]:

```
queues['Volume'].describe()
```

Out[50]:

```
count       163.00
mean      10229.85
std       23927.36
min           2.00
25%         738.50
50%        3204.00
75%       10566.00
max      244566.00
Name: Volume, dtype: float64
```

In [51]:

```
queues['Queue'].count()
```

Out[51]:

```
163
```

In [52]:

```
queues['Worktime'].describe()
```

Out[52]:

```
count         163.00
mean      1466400.59
std       3929695.68
min             0.00
25%         16942.20
50%        166508.00
75%        866416.00
```

```
max      32207136.00
Name: Worktime, dtype: float64
```

In [53]:

```
queues['Standard time'].describe()
```

Out[53]:

```
count   163.00
mean     22.35
std      36.65
min       2.50
25%       5.00
50%      12.50
75%      25.00
max     300.00
Name: Standard time, dtype: float64
```

In [54]:

```
queues['Saved time (minutes)'].describe()
```

Out[54]:

```
count       163.00
mean     168656.04
std      390340.88
min          15.00
25%        9480.00
50%       40320.00
75%      119578.75
max     3057075.00
Name: Saved time (minutes), dtype: float64
```

NTNU

Norwegian University of
Science and Technology