The benefits of automation on the purchase-to-pay process – A RPA implementation project

Riccardo Bramani

Dissertation written under the supervision of Professor Laura Wagner
[with the collaboration of industry expert Carlos Machedo]

Dissertation submitted in partial fulfilment of requirements for the MSc in Management, at the Universidade Católica Portuguesa, 01/06/2023.

Abstract

Title: The benefits of automation on the purchase-to-pay process – A RPA implementation project

Author: Riccardo Bramani

This thesis investigates how integrating robotic process automation (RPA) affects an organization's operational effectiveness. The study focuses particularly on the purchase-to-pay process.

The study takes a thorough approach, beginning with a review of the current purchase-to-pay procedure and the identification of potential areas for automation. Within the invoice validation step, the level of automation is increased to eliminate manual intervention and boost process effectiveness. The study identifies the origin and constraints of the current procedure through a thorough analysis of the complete purchase-to-pay process, with an emphasis on the invoice validation step.

The thesis also examines the processes of RPA creation, testing, and implementation, emphasizing the significance of exhaustive testing and validation to guarantee seamless integration and dependable performance.

The study looks into the benefits of RPA implementation, including enhanced productivity, cost savings, higher quality, lower risk, and compliance. It also recognizes the study's limitations as a result of the scant data available at the time of analysis and calls for more investigation using a larger dataset to provide more statistically significant findings.

The results of this thesis give us a first grasp of how RPA might affect operational effectiveness. It emphasizes the significance of ongoing assessment and further study to completely determine the long-term implications of automation. The study adds to the body of knowledge on RPA adoption and provides insightful information for businesses contemplating or implementing automation projects.

Keywords: RPA, Automation, purchase-to-pay.

Abstrato

Título: Os benefícios da automação no processo de compra-a-pagar - Um projeto de implementação de RPA

Autor: Riccardo Bramani

Esta tese investiga como a integração da automação de processos robóticos (RPA) afeta a eficácia operacional de uma organização. O estudo concentra-se particularmente no processo de compra-a-pagar.

O estudo adota uma abordagem abrangente, começando com uma revisão do procedimento atual de compra-a-pagar e a identificação de áreas potenciais para automação. No estágio de validação de faturas, o nível de automação é aumentado para eliminar a intervenção manual e aumentar a eficácia do processo. O estudo identifica a origem e as limitações do procedimento atual por meio de uma análise completa do processo de compra-a-pagar, com ênfase no estágio de validação de faturas.

A tese também examina os processos de criação, teste e implementação de RPA, enfatizando a importância de testes e validações minuciosos para garantir uma integração perfeita e desempenho confiável.

O estudo analisa os benefícios da implementação de RPA, incluindo aumento da produtividade, economia de custos, maior qualidade, menor risco e conformidade. Também reconhece as limitações do estudo devido à escassez de dados disponíveis no momento da análise e solicita mais pesquisas usando um conjunto de dados maior para obter resultados estatisticamente mais significativos.

Os resultados desta tese fornecem uma primeira compreensão de como o RPA pode afetar a eficácia operacional. Ele enfatiza a importância de avaliação contínua e estudos adicionais para determinar completamente as implicações de longo prazo da automação. O estudo contribui para o corpo de conhecimento sobre a adoção de RPA e fornece informações valiosas para empresas que estão contemplando ou implementando projetos de automação.

Table of contents

1. Introduction, repetitive tasks

In the modern corporate landscape, a variety of routine tasks exists. These tasks can often be tedious, monotonous, repetitive, and not always profitable. These activities can also be dull and uninspiring for employees. Over recent years, numerous studies have looked into how these repetitive tasks affect employee well-being. The findings are quite consistent: repetitiveness can lead to dissatisfaction and negatively impact the overall mental and emotional health of employees, as well as their productivity levels. For example, a paper titled "Experimental evidence for the effects of task repetitiveness on mental strain and objective work performance"[1] looks at how repetitive tasks can affect a person's mental well-being and performance at work. The authors of this study, Jan Häusser, Stefan Schulz-Hardt, Thomas Schultze, Anne Tomaschek, and Andreas Mojzisch, aimed to explore how the repetitiveness of tasks impacts a person's mental health and their ability to perform at work.

In their study, they conducted several experiments in which participants had to carry out tasks with varying degrees of repetitiveness. They measured the participants' mental stress and work performance using different methods, such as questionnaires, measurements of physiological responses like heart rate variability, and more objective measures of performance. The findings of the study showed that repetitive tasks can indeed increase mental stress and decrease performance at work. In particular, they found that participants who performed tasks that were very repetitive experienced higher levels of mental stress. This increased stress in turn had a negative effect on their ability to perform their work. Interestingly, they also found that participants who performed tasks that were less repetitive were more engaged and motivated, which resulted in better performance at work.

This research is valuable as it highlights the potential negative effects of repetitive tasks on individuals' mental well-being and job performance. It emphasizes the importance for businesses to review and revise their work processes to reduce repetitiveness where possible, and to invest in the mental health of their workforce. A happy, motivated, and mentally healthy workforce is key to the long-term success and productivity of any organization.

1.1. Context and Motivation

During my personal experience at CTT, I found out how automation, and in particular robotic process automation, can help contrasting the issue of repetitive tasks.
In this thesis I therefore talk about finding out if the use of RPA is improving CTT's practices and the well-being of its employees. The analysis of practical data will shed light on the issue and help other Portuguese and international companies making decisions about RPA transitions.
An example of a similar study, made for another company, is: "Robotic process automation at Telefónica O2"[3]. This case study focuses on the implementation of RPA at Telefónica O2, a telecommunications company. It reported that Telefónica O2 achieved significant cost savings, improved accuracy, and increased operational efficiency after implementing RPA. The company automated 15 core processes, resulting in a 75% reduction in average handling time and a 67% reduction in back-office workload. The study also highlights the importance of effective change management and governance in RPA implementations.

1.2. Methodology

The primary objective of this study is to evaluate the efficiency and effectiveness of implementing Robotic Process Automation (RPA) in the purchase-to-pay process, particularly in the validator identification, at CTT. To achieve this goal, we adopted a mixed-methods approach, which included the development and implementation of the RPA solution, the collection and of quantitative data, and the application of regression models to measure the impact of automation on the process efficiency.

First, we conducted an in-depth analysis of the purchase-to-pay process, examining the existing manual processes and identifying areas with potential for automation. This involved interviews with key employees at the company, observation of the processes in action, and the collection of relevant documentation, such as process maps and standard operating procedures. Once the invoice validation process was identified as an automation candidate, we proceeded to design the RPA solution using UiPath. This stage involved the creation of process maps for the automated workflow, the development of automation scripts, and the testing and fine-tuning of the RPA solution to ensure seamless integration with the existing purchase-to-pay process.

Following the successful implementation of the RPA solution, we initiated a data collection phase to gather quantitative information on the performance of the automated invoice validation process.
With the collected data, we conducted a comprehensive analysis to evaluate the impact of RPA on the efficiency and effectiveness of the invoice validation process. We employed descriptive statistics to gain an initial understanding of the changes brought about by the automation, followed by the development of regression models to establish the relationship between the implementation of RPA and the observed improvements in process efficiency. We synthesized the results to provide insights into the benefits and challenges of RPA implementation.

By adopting this mixed-methods approach, we were able to gain a comprehensive understanding of the impact of RPA on the purchase-to-pay process. The methodology allowed us to capture both the qualitative aspects of process improvement, such as employee satisfaction and ease of use, as well as the quantitative aspects, such as the measurable improvements in efficiency and effectiveness.

2. Automation

The main advantages that the use of automation brings are:

- Efficiency/cost reduction: One of the key advantages of implementing robotic process automation (RPA) is the potential for increased efficiency and cost reduction. Unlike human workers who are limited to a certain number of hours per week, a robot can operate 24/7, 365 days a year. This means that a robot can work up to 168 hours per week, significantly surpassing the hours a person can dedicate to tasks. Even in cases where exceptions allow individuals to work longer hours, the total number of hours worked by a robot remains significantly higher.
  Furthermore, robots offer cost advantages compared to human workers. Itrex[5], a technology consulting firm, estimates that the total cost of developing an automation robot typically ranges from $5,000 to $15,000 USD. Once the initial development costs are covered, the ongoing operational costs of maintaining and running the robot

are comparatively lower than employing a human worker. Robots do not require salaries, benefits, or other overhead costs associated with human resources.

- Increased quality and risk reduction: Implementing automation leads to increased quality and reduced risks associated with human errors. With minimal manual intervention, the probability of human errors is significantly decreased. Studies have shown that automation can reduce human errors by as much as 80% to 99%[2]. Robots, programmed with precise instructions and algorithms, consistently perform tasks with a high level of accuracy and attention to detail. Automation ensures tasks are executed consistently, mitigating risks associated with human errors, such as financial losses, compliance violations, and compromised data security. By minimizing human intervention, organizations gain greater control and reliability in their operations.
- Trasparency: everything that a robot does can be recorded, from execution times to eventual errors. This allows the company that controls it to explore a great amount of analytical data, that it can use to improve the process. Another added benefit of data analysis is that errors are spotted more quickly and are therefore not only easier to understand and solve, but also less likely to cause a chain reaction of other errors in related departments.
- Worker's motivation: sometimes, automation is seen as a tool that will replace humans, basically leaving them out of jobs. 8 years ago, a video was shared on YouTube, explaining how the horse was replaced by cars in the last century, and claiming that humans are bound to receive the same treatment by robots in this century. The video became hugely popular and created a lot of discussions on social networks. In reality, until now, automation has only been helpful to humans. The main tasks that are being done by robots are repetitive ones, which we've already briefly argued to be damaging for the human body and mind. Human resources can in this way be realigned to what they can do best, and employees can develop new skills.

2.1. Activities that can be automated

The potential of smart machines and RPA robots is expanding rapidly, the article: "Just How Smart Are Smart Machines?"[4], for example, points out that RPA have substantial potential to transform businesses and industries, even if the authors caution that these technologies are not a panacea and that their limitations and ethical considerations should be carefully managed.

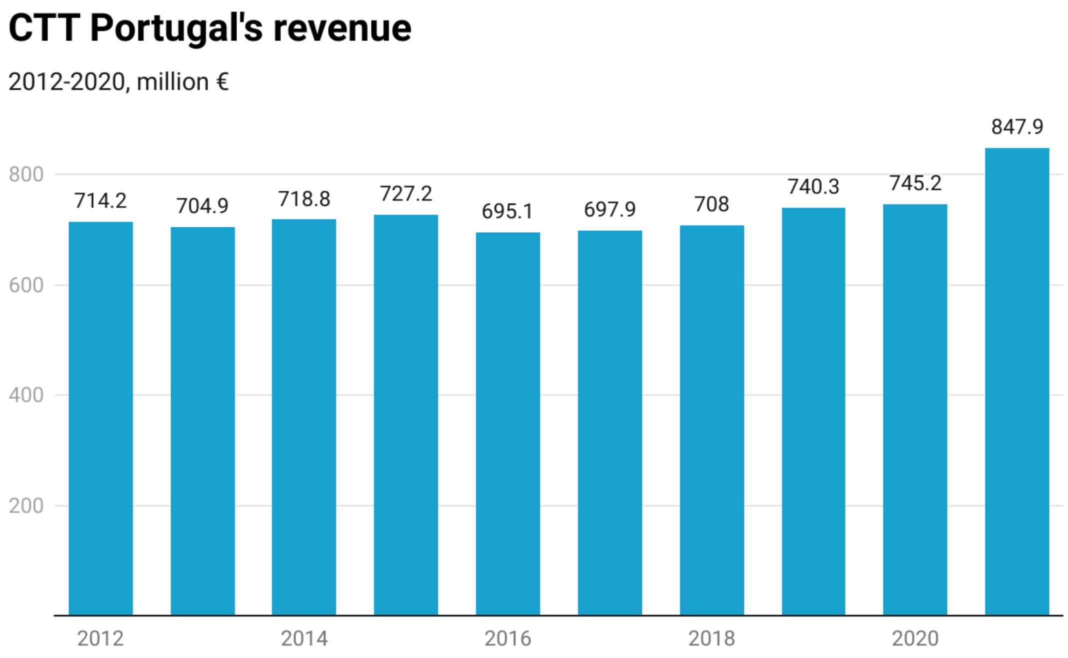
The array of activities that are suitable for automation is vast and expanding, some examples are:

- Send and open emails, downloading and handling eventual attachments.
- Perform login to various applications.
- Fill out forms, on and offline.
- Perform transactions in applications, granted they have permission to do it.
- Copy and paste contents between different files or applications.
- Register and read data in and from databases.
- Collect metrics (ex. Execution times) and produce reports.
- Perform calculations and other MS Office operations.
- Connect with APIs (application programming interfaces).
- Extract data from applications and online services.

This list is already relatively long, but it's non-exclusive, moreover, the number of activities that can be automated keeps growing as the technology evolves. It's important to notice one more time that all the activities in this list are repetitive tasks, that bring little to no value to human well-being.

2.2. Automation at CTT

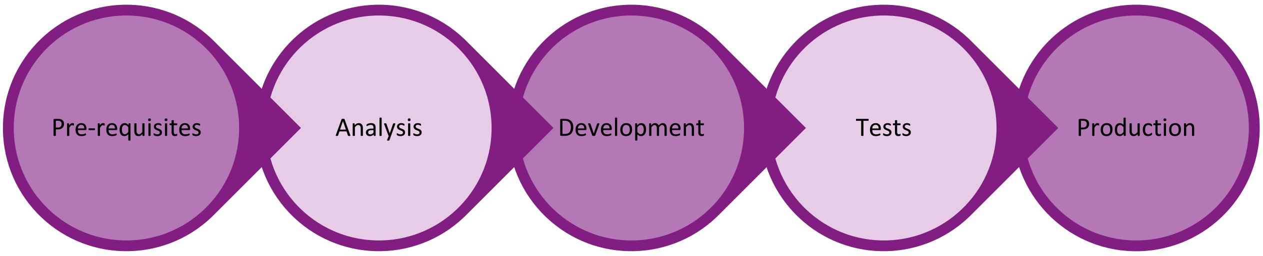
CTT – Correios de Portugal is a Portuguese company that operates as both the national postal service of Portugal and a commercial group with subsidiaries operating in banking, e-commerce, and other postal services. It was founded in 1520 by King Manuel I of Portugal, during the Portuguese Renaissance. CTT is the oldest company still in operation in Portugal to this day. The company' revenue in 2021 was 847.9 million, signaling a continued growth over the past few years.[2]



The process of automating tasks at CTT started recently, in 2020. In 2021, the process became more consistent with: a new branding, increased quality of the deliverables, increased training for team development, a shift in focus (from quantity to net profit), improved communication, and expansion of the automation team and a skill espansion (Agilepoints and chatbots).
In total there were 101 robots in production and over €773.000 in potential savings.
2022 was another important year for the company's automation process, with better identification of automation-capable processes, increased communication of this process results and a higher concentration on the quality of the single automations.

2.2.1. Automation development

The development of automation programs follows specific steps:



- Pre-requisites: Before initiating an automation procedure at CTT, certain prerequisites need to be fulfilled. The first step involves identifying the opportunities that automation can bring and assessing its viability. To accomplish this, a mandatory meeting takes place with the Center of Excellence (CoE), a dedicated team responsible for driving automation initiatives within the organization.

  During this meeting, the CoE evaluates the potential benefits and impact of automation in the specific process under consideration. They analyze various aspects such as process complexity, volume of transactions, potential error reduction, and time savings. This assessment helps determine whether implementing automation is a suitable and beneficial choice.

  In addition to evaluating the operational advantages, the CoE collaborates with the process owner to calculate and establish the financial aspects of automation. This includes preparing a comprehensive business case that outlines the projected automation earnings and associated cost savings. The business case takes into account factors such as reduced labor costs, increased productivity, minimized errors, and improved efficiency.

  Once the business case is prepared, it undergoes a thorough review and approval process. The process owner, who has a deep understanding of the process and its financial implications, plays a crucial role in validating the calculations and ensuring the financial viability of automation. This step ensures that the proposed automation initiative aligns with the company's strategic goals and provides tangible financial benefits.

- Analysis: this step consists of a series of meetings to examine the requirements details, some of the key elements that are analyzed are:
    - Critical analysis of the process and suggest reengineering in the situations where it justifies.
    - Check their reporting needs.
    - Assess who the stakeholders are– involved in the process, persons authorized to communicate with robot (i.e. send emails, etc.)
    - Installation of systems/applications/API's involved and Access credentials.
    - Request access to the Output and Input mailbox and its applicable rules (if applicable).

- Development: The development phase of an automation project begins with the design of the process architecture. This involves the development team creating a process diagram that outlines the step-by-step workflow of the automated process. The process diagram serves as a visual guide that helps the team understand the flow of activities and identify areas where automation can be applied effectively.
  During the design phase, the development team also looks for opportunities to reuse common activities from existing automations. This approach saves time and effort by leveraging pre-existing automation components that have already been developed and tested. By reusing these components, the team can accelerate the development process and ensure consistency across different automation projects.
  Once the process architecture is defined, the development team proceeds with configuring the automation file. They follow a template structure that guides them in organizing the various components of the automation, such as input data, rules, decision points, and output actions. This template-based approach ensures consistency in the structure and organization of automation files, making it easier for developers to collaborate and maintain the automations over time.
  Throughout the development process, the team also handles any errors that may arise. They diligently test the automation, both during development and in dedicated testing environments, to identify and rectify any issues or bugs. Error handling involves implementing appropriate exception handling mechanisms, error logging, and recovery procedures to ensure the automation operates smoothly and reliably.
  The development phase requires close collaboration between the development team, subject matter experts, and stakeholders. Feedback and input from these stakeholders

are valuable in refining the automation design and ensuring that it aligns with the desired goals and requirements. Regular communication and iterative development cycles help to enhance the automation and address any challenges or changes that arise during the development process.

- Tests: After the initiation of the robotic process automation (RPA) implementation, thorough testing becomes a crucial step. These tests are conducted in dedicated quality environments to ensure that the RPA solution functions correctly and meets the desired requirements. The testing phase plays a vital role in identifying and resolving any potential issues or bugs before the RPA solution is deployed in the live production environment.
  During testing, the RPA development team executes various test scenarios to validate the functionality and performance of the automation. These tests cover different aspects, including data input, process execution, error handling, and output verification.
  Following the testing phase, the code is reviewed by the Center of Excellence (CoE) to ensure adherence to best practices and industry standards. The CoE, composed of automation experts, carefully examines the code to ensure it is well-structured, efficient, and maintainable.
  To validate the final code, every stakeholder involved in the RPA project receives an RPA demo. This demonstration showcases the capabilities and functionalities of the RPA solution, providing stakeholders with a firsthand experience of how the automation performs. Through the demo, stakeholders can provide feedback, validate that the RPA solution meets their requirements, and identify any additional improvements or adjustments that may be necessary.
  The involvement of stakeholders in the validation process is essential as it ensures that the RPA solution aligns with their expectations and needs. Their feedback and validation play a crucial role in fine-tuning the automation solution and ensuring its effectiveness in addressing the targeted business objectives.
  This comprehensive testing and validation process contribute to the successful implementation and deployment of the RPA solution, providing stakeholders with confidence in the automation's performance and its potential to drive efficiency and productivity gains.

- Production: The production or implementation phase of the automation involves deploying the finalized RPA solution into the live operational environment. During this phase, the development team works closely with the IT department to ensure a seamless integration of the automation solution with the existing systems and infrastructure. The RPA solution is carefully deployed, and any necessary configuration or setup steps are completed to enable its operation. Throughout the implementation phase, close monitoring and support are provided to address any potential issues or challenges that may arise. This phase marks the transition from development to active usage of the RPA solution, enabling the organization to leverage the benefits of automation in real-time operations.

2.2.2. General objectives

At CTT some general objectives to look for in processes that can be automated are:

- Rule-Based:
  Automation is particularly beneficial for rule-based processes that follow consistent patterns and guidelines. These processes rely on predefined rules and logic, making them suitable candidates for automation. CTT can therefore ensure that these processes are executed accurately and consistently, minimizing the potential for errors or deviations.
- Large Volume of Data:
  Automation is highly advantageous for tasks that involve the systematic processing of a large volume of data. Manual processing of such data can be time-consuming, prone to errors, and inefficient. Automation can swiftly handle the processing and analysis of vast amounts of data, improving efficiency, reducing processing time, and enhancing accuracy.
- Digital Trigger:
  Processes that involve digital reception or sending of documents, coupled with intermittent manual steps, can greatly benefit from automation. For example, when a process begins with the electronic receipt of documents containing important data, automation can streamline subsequent manual steps, ensuring a seamless and error-free workflow.

- High manual error rates:

  Tasks that involve paper data entry or tasks that are highly interdependent are susceptible to manual errors. Automation can significantly reduce manual error rates by eliminating the need for manual data entry and by automating the flow of interdependent tasks. This helps maintain data integrity and reduces the risk of errors propagating throughout the process.
- Manual Calculations

  Laborious tasks that require manual calculations, where one error can lead to subsequent errors, are ideal candidates for automation. Automating these calculations eliminates the risk of manual calculation errors and ensures accurate and consistent results, saving time and improving overall efficiency.
- Work off-hours

  Processes that involve seasonal work overloads or tasks that require 24-hour availability can benefit from automation. Automating these processes allows for uninterrupted operation, even during off-hours, and enables timely resolution of complaints, requests, and other time-sensitive activities.
- High Compliance

  Processes that require strict adherence to regulatory compliance can benefit from automation. Automation can ensure consistent adherence to compliance guidelines, generate audit evidence, and provide a traceable record of actions, thereby enhancing compliance management and reducing the risk of non-compliance.
- Validations

  Tasks that involve multiple systems and require validations at each step can be effectively automated. Automation can streamline the validation process by integrating multiple systems, performing the necessary validations automatically, and ensuring data accuracy and integrity throughout the process.
- Large number of features

  Processes that involve numerous resources or contain multiple steps can be complex to manage manually. Automation can simplify and streamline these processes by orchestrating the various features and resources involved. This reduces the likelihood of errors, improves efficiency, and ensures a smooth and coordinated workflow.

2.3. UI path

CTT uses the program UIPath to design, develop, and deploy its software robots. UiPath is a Robotic Process Automation (RPA) platform that allows businesses to automate repetitive tasks by creating bots. It provides a visual interface for designing automation workflows, which can interact with various applications and systems just like a human would. It can automate interactions with desktop and web applications, extract structured data from documents, websites, or applications, use AI to identify and interact with elements in virtual desktop environments, process and extract data from documents and connect with various applications and services, such as Excel, databases, email, or APIs.

The main components of the program are variables, arguments, and activities:

- Variables: Variables in UiPath are used to store and manipulate data within a workflow. They can hold different types of data, such as text, numbers, or more complex data structures like arrays and data tables. Variables enable you to store data temporarily and pass it between activities.
- Arguments: Arguments are similar to variables, but they are used to pass data in and out of reusable components called workflows. By using arguments, you can create modular and reusable automation processes that can be invoked with different input data or return data to the calling workflow.
- Activities: Activities are the fundamental building blocks in UiPath. They represent individual actions or operations that can be combined to create automation workflows. UiPath provides a wide range of activities for various purposes, such as working with files, automating web browsers, interacting with APIs, or performing calculations.

UIPath also offers a debugging functionality that allow you to set breakpoints, step through your workflow, and inspect the values of variables and arguments at runtime. This helps you to diagnose and resolve any problems in your automation.

3. The current Process

The existing purchase-to-pay process already incorporates some elements of RPA in the invoice validation step. However, the company is looking to expand the scope of automation in this step to increase the number of processes managed directly by the system and to reduce the need for manual intervention. Initially, an analysis of the entire purchase-to-pay process will be conducted, followed by a more focused examination of the invoice validation stage. This approach will give a comprehensive understanding of the project's starting point and also help to identify potential limitations of the current process.

The first process outlines the progression from purchasing to payment, the second one focuses on the invoice processing procedure, particularly in cases where automated invoice scanning and payment are not executed, and the third one covers the validator's identification, which will be the step interested by the automation explained in this paper. In the report, each step in the flowchart is labeled with a number. This will make it easier to understand how things move along in the purchase-to-pay process and show which areas they want to automate. The goal is to make the whole process more efficient by using RPA and reducing the need for manual work.

3.1. As-Is purchase-to-pay process

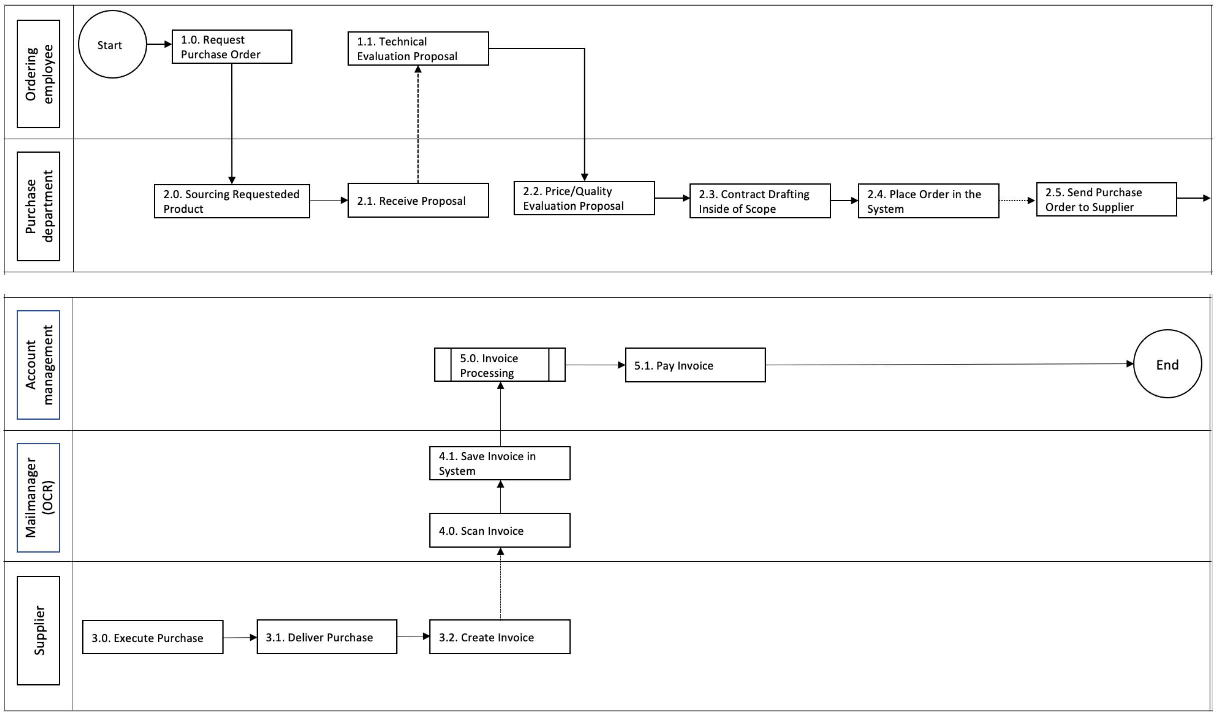
The Purchase to Pay Process starts with placing an order and ends in its settlement. A CTT employee initiates the process by submitting a request for an order to the purchasing department (1.0). The receiving department then seeks the requested goods or services from approved preferred suppliers (2.0). Upon receiving suitable proposals (2.1), the order department performs a technical evaluation (1.1) before sending the reviewed bids back to the procurement department.

Taking into account both the technical review and additional pricing factors, a purchasing decision is made (2.2) and subsequently transformed into a purchase agreement within the defined parameters (2.3).

The order is then entered into the purchase system and sent to the selected vendor (2.4 & 2.5). The supplier fulfills the order (3.0) and sends the products along with an invoice (3.1). The technical unit automatically scans the invoice (4.0), and the procurement system registers it using optical character recognition (OCR) technology (4.1).

Subsequently, the invoice is automatically forwarded to the finance department (5.0), which proceeds with processing and settling the invoice if it meets the necessary criteria (5.1).

The process reaches completion upon approval of the invoice and receipt of payment.



3.2. As-Is invoice validation process

In this analysis, our focus shifts to the second process mentioned, which is critical for understanding the current system and designing an automated solution. This process deals with the handling and validation of supplier invoices, which play a significant role in the overall purchase-to-pay workflow.

The initial step in this analysis involves the arrival of the supplier's invoice in the designated inbox, where it is processed using optical character recognition (OCR) technology (1.0). Everything goes smoothly, and the so called “happy path” starts if the invoice has a purchase number that matches a purchase order in the system.

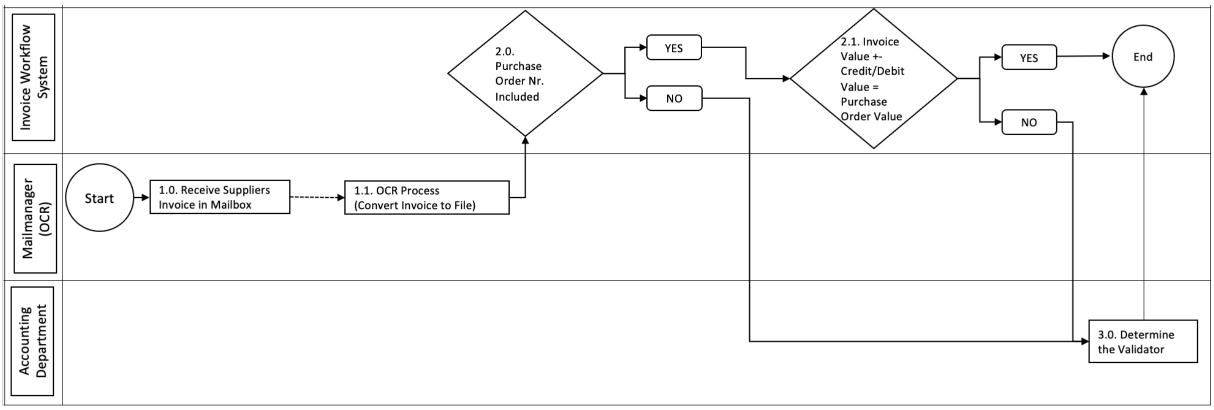
Then, the system checks if the invoice amount matches the purchase order amount (2.1). If the invoice value, including any previous credit/debit adjustments with the supplier, is equal to the purchase order value, the happy path is done.

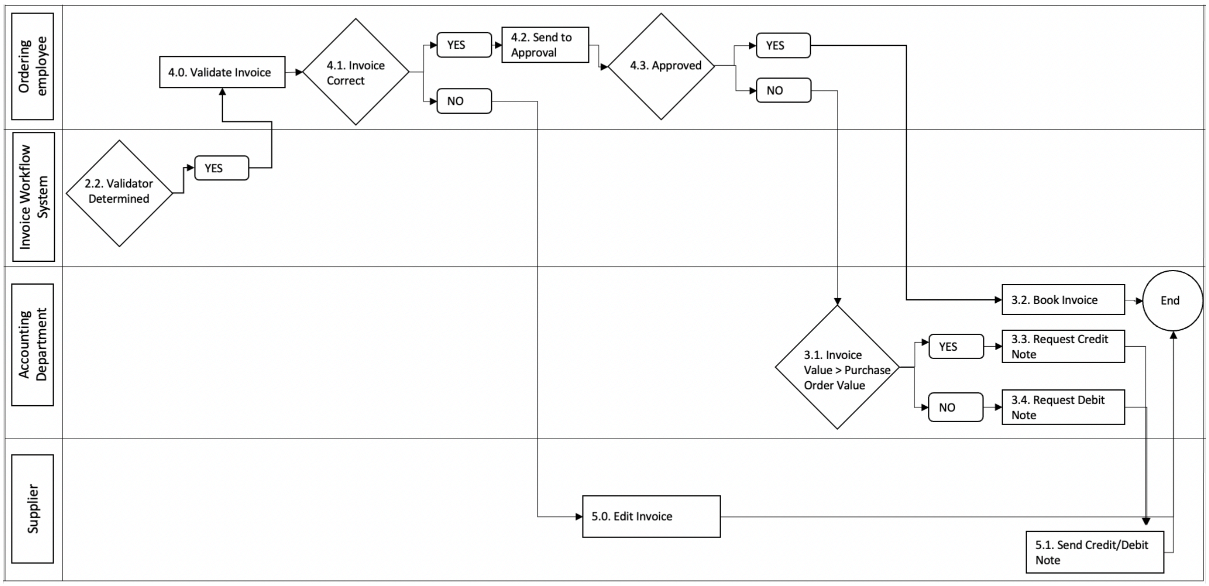
Looking back at the decision point (2.0), if the Purchase Order Number isn't found, the invoice goes straight to the accounting department for checking (3.0). Here is the focal point of the automation employed and studied by this thesis. The problem that is tackled is figuring out who will check the invoice. As of 2023, this step involved a lot of human intervention, and even the people who were taking care of it didn’t have much information to identify a precise validator, so guessing was a big part of it.

Moving to the decision point (2.1) and leaving the smooth process, if the adjusted invoice value doesn't match the purchase order value, the same action happens—the invoice goes to the accounting department to find out, once again, who will check it (3.0), and the process is finished. The process design then continues when the person who will check the invoice is found (2.2).

If the invoice is correct (4.1), the person who ordered sends it to the manager for approval (4.2). If the manager says it's okay (4.3), the invoice goes to the accounting department to be recorded (3.2), finishing the smooth process. If the invoice is wrong, it's considered not valid, and the supplier is told (5.0). This ends the process because the supplier must fix and resend the invoice, starting the process from the beginning.

Going back to the point where the checked invoice isn't approved by the manager (4.3), the accounting department needs to see if the invoice amount is more or less than the purchase order value (3.1). If the invoice amount is more than the purchase order value, they ask for a credit note (3.3). If the invoice amount is less than the purchase order total, they ask for a supplier debit (3.4). The process ends when the supplier confirms the debit or credit note (5.1).



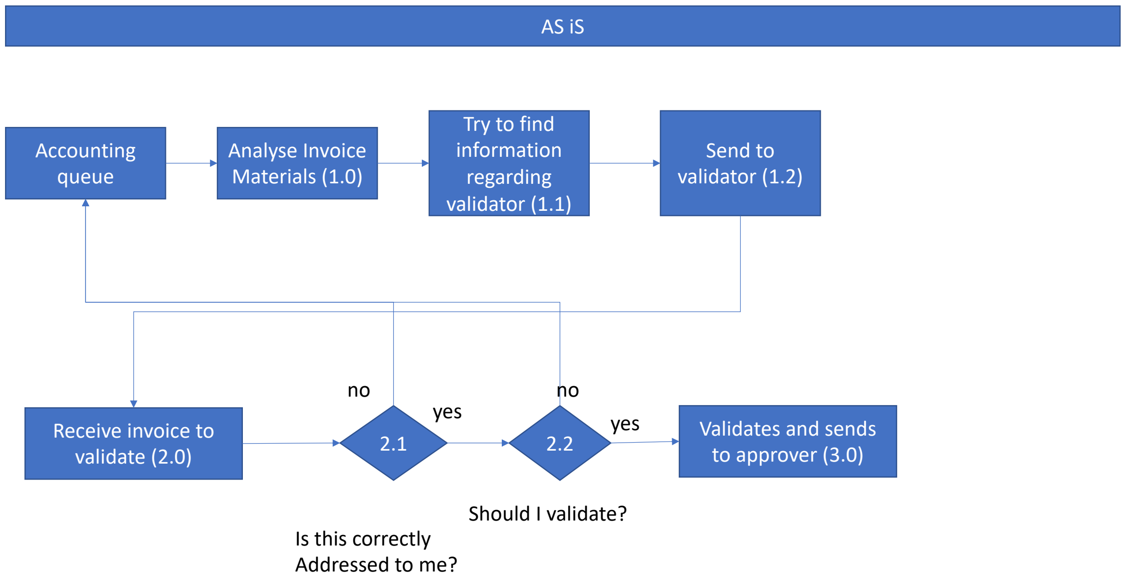


3.3. As-is validator identification

As previously mentioned, this particular thesis focuses on the validator identification. This process is needed every time the smooth process is abandoned for one of the problems previously explained and there's the need for manual intervention in order to validate the invoice.

The current state of the process requires manually analyzing the invoice material (1.0) to try and find information regarding the correct validator (1.1), based on the specific supplies ordered, the department that they will be sent to and other various information that might be contained in the invoice, and sending the invoice to the validator (1.2). However, it is an unreliable process that often implies a lot of guessing.

When the designed validator receives the invoice (2.0), he first checks whether it was correctly sent to him (2.1), if he is not responsible for the order, he sends it back to the accounting department and the invoice reenters the accounting cue. If he was correctly appointed as the validator for that invoice, he examines it and decides whether to validate it or not (2.2) based on the information reported in the invoice and on the purchase order. If he validates, the invoice is sent for approval (3.0), otherwise it flows back to the accounting cue.



4. Future validator identification process

The ability of the validator identification process to guarantee that invoices are appropriately validated and handled by the proper persons underlines its significance. The overall efficiency and accuracy of CTT's invoice validation process can be improved, which will ultimately result in better financial management and control, by improving the identification of invoice validators and lowering the need for guessing.

4.1. To-Be validator identification

The goal of the explained project is to increase the degree of digitalization and specifically automation of this step to make it not only faster and more efficient, but also more reliable. This requires important changes in many steps of the process.

A central purchase department will be instituted, where all the purchase requests will flow to. This will allow for better control over the spendings of the company. In addition, this department will hold a database where every purchase order is linked to a specific validator, so that if there's the need for manual validation, it will be substantially easier to identify the person responsible for it. The new process will go on as follows:

First, invoices will be fed into an accounting bot cue (1.0), the bot will extrapolate the following data (1.1):

- Was it already worked by bot?
- Supplier NIF
- Invoice Value
- Invoice Data
- PO number

NIF: a NIF (Número de Identificação Fiscal), also known as a tax identification number, is an identification number used in Portugal for tax-related purposes. It is sometimes also referred to as a VAT number or fiscal number.
This unique number is used to track all tax-related activities in Portugal, such as income tax payments, value-added tax (VAT) payments, and other tax obligations. Every individual or business entity in Portugal, whether a resident or not, is required to have a NIF for any financial transactions that might be subject to tax, such as purchasing real estate, opening a bank account, receiving income, or starting a business.
The NIF is a nine-digit number, and the structure varies depending on whether it's issued to an individual or a corporate entity. For individuals, the number starts with 1 or 2, while for companies, it usually begins with a 5. It is important in this process as it is one of the main tool that the bot can use to link a purchase order to the specific invoice.

Invoice value: the invoice value plays a crucial role as it is possible to establish a direct link between a specific invoice and its corresponding purchase order. By examining both documents, it becomes possible to search for a direct correspondence in terms of the invoiced amount. Through the utilization of automated processes, such as optical character recognition (OCR) and data extraction techniques, it becomes possible to extract the invoice value and compare it with the purchase order data. This automated matching capability reduces manual effort, minimizes errors, and streamlines the reconciliation process.

Invoice data: other invoice data is analyzed as it is also possible to use it to link an invoice to a purchase order (PO) and, consequently, identify a validator. This additional data provides valuable context that can aid in the matching process between invoices and POs. By examining and analyzing the invoice data, we can identify key elements such as supplier details, product descriptions, dates, and other relevant information. These details serve as reference points to establish a connection between the invoice and its corresponding PO.

PO number: the inclusion of a PO number, or purchase order number, in an invoice significantly simplifies the process of linking the invoice to its corresponding purchase order (PO) and enables the automatic identification of the validator. When an invoice contains a purchase order number, the correlation between the two becomes much more straightforward. As each purchase order should be linked to a specific validator, the presence of the PO number facilitates the automated identification of the validator.

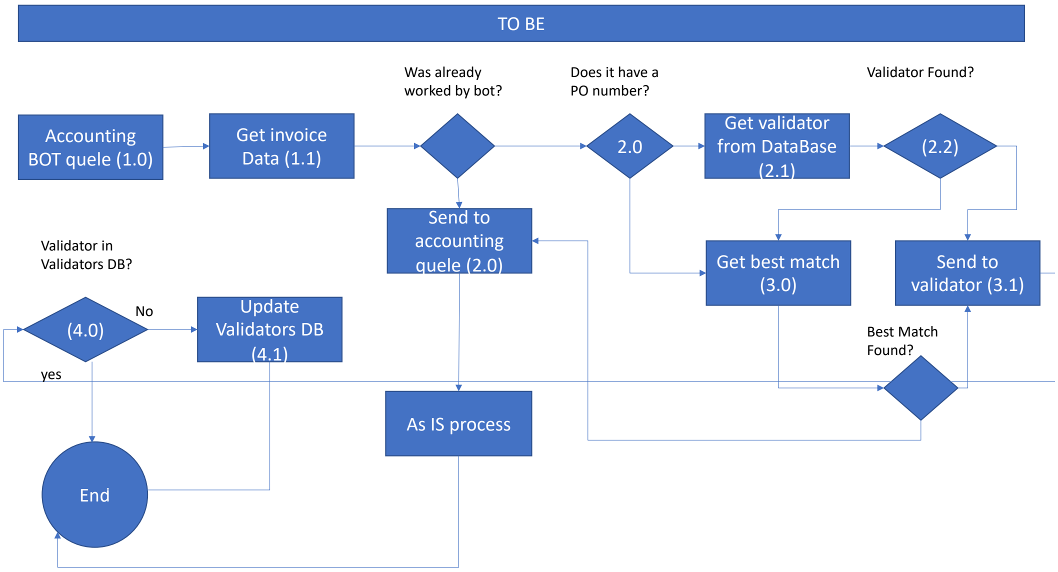
Continuation of the process:

If the answer to the initial question "Was the invoice already worked by a bot?" is yes, it means that a bot has previously attempted to find a suitable validator for the invoice but without success. In this case, the invoice is directed to the regular accounting queue (2.0) and follows the standard process explained earlier. However, if the answer is no, meaning the bot has not worked on the invoice yet, it proceeds to the next step.

The bot begins searching for the purchase order (P.O.) number (2.0). If the P.O. number is found, the bot automatically searches for the corresponding validator in the new database (2.1). If the appropriate person is found, the invoice is smoothly and automatically sent for validation (3.1).

In cases where the correct validator is not present in the database or if the bot fails to locate the P.O. number, the program continues its efforts to find the best match based on other information available on the invoice (3.0). If a probable correct validator cannot be identified, the invoice is sent to the accounting queue (2.0). However, if a likely match is found, based on other information on the invoice the program once again automatically sends the invoice for validation (3.1).

Once the correct validator is determined, either through the P.O. number or through best match identification, the robot also verifies if the validator's name is already included in the new validators database (4.0). If the name is not found, the robot automatically adds it to the database (4.1).



4.2. KPIs

To measure the success of our project and establish clear performance indicators, we applied the SMART methodology. SMART stands for Specific, Measurable, Attainable, Relevant, and Time-bound, which helps in setting and evaluating objectives effectively. We used this framework to assess the efficiency of our centralized and automated validator identification process.

Firstly, we set a specific goal to precisely define the objective of monitoring the effectiveness of the centralized validator identification process.

Next, we needed to create measurable objectives to determine how we would measure the effectiveness of the process. This involved using metrics such as the time required to find a validator, the average duration of the validation process when the ideal scenario is not possible, the number of validators identified per month, and the frequency of manual intervention, among others.

To ensure attainability, we set objectives that were within a realistic time frame and took into account available resources and the complexity of the task. It was important to establish relevant objectives that aligned with the organization's broader goals and could positively impact overall performance.

To enhance efficiency, we set a specific time frame for achieving the objectives and established interim benchmarks to monitor progress along the way.

Here's an example of how we applied SMART to our project:
Specific: Implement a centralized and automated RPA solution for identifying invoice validators.
Measurable: Measure success by tracking metrics such as the time required to find a validator, the average duration of the validation process when the ideal scenario is not possible, the number of validators identified per month, and the frequency of manual intervention.
Attainable: This goal is achievable if we have the necessary resources (funding, technical expertise, stakeholder support) and a well-designed and implemented RPA solution.
Relevant: This objective aligns with our organization's goals of improving process efficiency and reducing costs.
Time-bound: Complete the goal within a specific timeframe, such as six months from the project's start, to ensure progress and keep the project on track.

5. Analysis, limitations and improvements

To measure the efficiency benefits brought about by the use of robotic process automation (RPA), we want to use a regression model in our study, specifically, a logistic regression. We will be able to scientifically analyze the association between RPA usage and several organizational efficiency measures using the regression model. By doing this, we want to forecast any efficiency gains that might be seen when RPA is used in certain scenarios. This strategy will give us a more solid scientific foundation for comprehending how RPA boosts productivity and will direct future choices about RPA adoption and its application across various jobs and operations.

5.1. Limitations

Data on the outcome of the automation are few because it was not fully implemented at the time of the dissertation deadline. However, despite the sparse data, this approach can still offer insightful information and act as a rough predictor of possible outcomes.

The findings from the current dataset could not accurately reflect the true impact of automation. The findings' precision and dependability may be constrained by the small sample size and quick data gathering period. When adequate data becomes available, it is advised to run the regression model again in order to get more reliable and insightful findings. Additionally, our data doesn't contain nearly enough parameters to take into account all the factors that affect how long the validation process takes.

However, even though the results of the current investigation might not be statistically significant, they still help to provide a preliminary knowledge of the possible effects of automation. It can serve as a starting point for spotting trends, patterns, and possible connections between variables. These preliminary conclusions can direct future study and provide as a springboard for a more thorough review.

The sample size will grow over time when the automation is fully implemented and more data is gathered, enabling more precise and trustworthy analysis. It will be possible to get more reliable and significant results by running the model again with a bigger dataset, which will make it easier to come to useful conclusions.

5.2. Improvements for next studies

Future research must take into account the constraints of the current study and make plans to go around them in order to produce more reliable, accurate, and insightful results.

Data diversity and accessibility remain crucial priorities. The lack of post-automation data in the current investigation constrained our ability to conduct an analysis. This restriction has highlighted the requirement for a balanced and thorough data set, with both pre- and post-automation data, in order to enable an appropriate assessment of the automation's impact.

Beyond time and the dummy variable of whether a procedure was automated, it is important to investigate additional parameters. Future research can think about including factors that highlight the complexity of activities, the variety in how processes are carried out, the workload or demand, and more. These more variables would make it possible to understand automation efficiency and its effects on the many aspects of operations in a more complex way.

We've also discovered how crucial it is to keep the data collection procedure consistent. The accuracy and comparability of the data gathered can be improved by ensuring the use of uniform measuring instruments and methodologies during both the pre- and post-automation phases.

Finally, widening the study's focus to include qualitative factors like employee happiness could give a comprehensive picture of automation's effects. We may acquire a thorough picture of the automation's effects by looking at both the quantitative effectiveness and the qualitative experiences of persons dealing with the automated procedures.

5.2.1. Increase the number of parameters

While our suggested model, which is centered on time and the dummy variable of automation, provides a good place to start for this study, we also need to take into account the possible benefits of including additional elements. These variables will be crucial in improving our model so that it more accurately reflects the intricate realities of our business operations. Other possible variables to include might be:

The intricacy of the task. Regardless of whether the process is automated or not, varying task complexity and time requirements are a given in any business activity. Our model would be able to take into consideration these factors, for example, by implementing a complexity score.

The amount of work put into a process is the second parameter we may take into account. The length of time it takes to complete a process, whether automated or not, could be significantly impacted by how much work is involved, such as a vast dataset or a large number of objects.

Thirdly, we need to consider the day of the week or time of day the task was completed. Both human and automated productivity levels can change over time, with human workers generally being less productive toward the conclusion of shifts and automated systems being impacted by fluctuating server load levels.

The level of expertise or experience of the people carrying out manual tasks constitutes a fourth potential factor. This allows us to distinguish between the effects of automation and unique worker characteristics on process time.

Finally, system performance indicators would be crucial additions to our model should our automation incorporate software or hardware systems. System uptime, error rates, CPU utilization, and network speed are some variables that might help us better understand how these systems might affect our outcomes.

5.2.2. Increase the number of observations

The amount of data needed for a regression model is an important concern, and it frequently depends on a variety of variables, including the model's complexity (i.e., the number of parameters), the level of precision desired, and the variability of the data. The pre-automation period and the post-automation period are the two unique time periods we are comparing in our situation with the implementation of automation.

A substantial basis for understanding how processes functioned before to the shift is provided by the number of observations from the pre-automation period (15.000+) in our research. However, the post-automation period, in which we have a just a few observations, is a substantial issue. This disparity might change our perception of how automation affects the process. It's like contrasting a high-resolution pre-automation image with a low-resolution post-automation image. The image is roughly the same, although the latter case's features can be blurrier, making exact comparison challenging.
We require a substantial number of observations from the post-automation phase in order for our regression model to produce results that are trustworthy and generalizable. The sample size required in the post-automation era relies on the degree of precision that we want to achieve and how variable the data are. To avoid the model being overly biased by the group with more data, it is generally advised that the sample sizes for the compared groups be about equal, or at least comparable. We would therefore advise to get at least 2000/3000 more observations form the second group.

Additionally, having additional observations following the adoption of automation would help us determine whether or not the advantages of automation endure over time. Therefore, gathering more information from the post-automation phase would be extremely valuable before we can confidently move forward with the analysis.

5.2.3. Include qualitative aspects

It's important to include qualitative factors in future study on the effects of automation. This change reflects an effort to comprehend both the quantitative effects on productivity and efficiency as well as the less obvious but no less significant consequences on human dynamics within the business. A more complete picture of the overall impact of automation can be obtained by looking into employee experiences, satisfaction levels, workload perceptions, and attitudes. This qualitative data, which offers insights into how automation is viewed and how it influences the work culture, can be acquired through interviews, focus groups, or questionnaires. Finding effects, such stress brought on by change or higher job satisfaction owing to fewer boring activities, might be helpful. Future research that takes into account these factors will provide a more comprehensive understanding of the advantages and drawbacks of automation, ultimately helping businesses who are thinking about automating more of their activities in developing their business strategies and making more informed decisions.

6. The model

In the following section, we will talk about the details of how the model has been constructed, despite the limited reliability of the results. Although the analysis may lack statistically significant findings due to the recent implementation of automation and a shortage of sufficient data, it is still important to understand the methodology and processes involved.

6.1. Libraries required

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion
_matrix, roc_auc_score, log_loss
from imblearn.over_sampling import RandomOverSampler
from imblearn.pipeline import Pipeline
import warnings
warnings.filterwarnings('ignore')
```

These imported libraries and functions were essential for conducting data analysis and implementing machine learning algorithms:

1. The "pandas" library was used to handle and manipulate structured data, which played a crucial role in processing and analyzing the data related to the purchase-to-pay process.

2. The "matplotlib.pyplot" module enabled the creation of visualizations, allowing for a clear presentation of the findings and insights derived from the data analysis.

3. The "LogisticRegression" class from the "sklearn" library was utilized to build a predictive model to assess the impact of automation on operational effectiveness. Logistic regression served as a suitable algorithm for binary classification tasks, making it applicable in evaluating the influence of various factors on the outcomes.

4. The "train_test_split" function was used to split the dataset into training and testing subsets. This enabled the evaluation of the model's performance on unseen data, providing an estimation of its predictive capabilities.

5. The imported evaluation metrics, such as accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, roc_auc_score, and log_loss, allowed for the assessment of the model's performance in terms of classification accuracy, precision, recall, and other important metrics. These metrics helped in understanding the effectiveness of the model in predicting outcomes and provided insights into the quality and reliability of the results.

6. The "RandomOverSampler" class from the "imblearn" library was employed to address the issue of imbalanced classes in the dataset. It helped takle potential bias caused by imbalanced data, trying to ensure that the model's predictions were not skewed towards the majority class.

7. Finally, the "Pipeline" class from the "imblearn" library facilitated the organization and execution of a sequence of data processing and modeling steps. It streamlined the workflow by combining multiple transformations and modeling techniques into a cohesive pipeline.

6.2. Data Preparation and correlation

Through an elaboration of the provided dataset, we were able to calculate the time needed to complete all the processes where the automation was not involved and the few ones where it was involved.

Understanding how automation affects process length requires making a distinction between processes with automation and those without it. This way, it is possible to evaluate the efficiency and time-saving advantages provided by automation by comparing the time necessary for procedures where automation was not applied to those where it was.

```
data.corr()
```

	Time	Worked by a bot
Time	1.000000	-0.003641
Worked by a bot	-0.003641	1.000000

The correlation coefficient shows that the two variables, "Time" and "Worked by a bot," are correlated. The correlation coefficient between the two variables in this instance is -0.003641. This indicates that there is almost no linear relationship between the amount of work done by a bot and the time required to finish procedures.

Although this is not the outcome we were hoping for, additional data must still be gathered in order to draw significant conclusions.

6.3. Logistic regression

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Apply Random OverSampler to the training data
ros = RandomOverSampler(sampling_strategy='minority', random_state=42)


# Train the Logistic Regression model using 5-fold cross-validation
pipeline = Pipeline(steps=[
    ('ros', ros),
    ('classifier', LogisticRegression(solver='saga', random_state=42, max_iter=250))
])
model = pipeline.fit(X_train, y_train)
```

```
print(model.named_steps['classifier'].coef_[0])
```

The train_test_split function is used to divide the dataset into training and testing subsets, ensuring that the model's performance is evaluated appropriately. The training data's features and labels are represented by the variables X_train and y_train, whereas the testing data's features and labels are represented by X_test and y_test. The random_state option assures that the results may be replicated, while the test_size argument sets the percentage of data (20%) that will be used for testing.

The RandomOverSampler is used to resolve any class imbalance in the training data. This method oversamples the minority class in order to balance the distribution of the classes, improving model performance and reducing biases.

Using a pipeline, the logistic regression model is trained. The pipeline combines the classifiers LogisticRegression and RandomOverSampler (ros). The logistic regression model uses the optimization algorithm'saga' as its solver parameter. While the max_iter option provides the maximum number of iterations for convergence, the random_state parameter ensures the repeatability of random processes within the model.

The logistic regression model's coefficients are displayed using the print statement once the model has been fitted using the training set of data. The coefficients connected with the features used in the model are retrieved using the model.named_steps['classifier'].coef_[0] command. The given example's coefficient, which denotes the correlation between the target variable's log-odds and its features, is displayed as -0.0095028.

6.4. Model evaluation

The model.predict() function is used to forecast the labels for the features in the X_test dataset in order to evaluate the model's performance on unobserved data. The y_pred variable contains the expected labels.

The performance of the model is further assessed by calculating a confusion matrix. The confusion_matrix() method creates a matrix that represents the classification results by taking as inputs the real labels (y_test) and the predicted labels (y_pred). Conf_matrix, the resulting matrix, contains data on predictions that were true positive, true negative, false positive, and false negative. According to this matrix, there are 0 genuine positives, 1 false positive, 1180 false positives, and 1346 true negatives.

```
# Check the performance on the test set
y_pred = model.predict(X_test)

# Calculate the confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)
```

```
Confusion Matrix:
 [[1346 1180]
 [   1    0]]
```

To evaluate the model's predictions beyond simple classification accuracy, additional metrics are calculated. First, the model.predict_proba() function is used to calculate the predicted probabilities of the positive class (1) for each instance in the X_test dataset. The predicted probabilities are stored in the y_pred_proba variable.

```
# Calculate the predicted probabilities for ROC-AUC score
y_pred_proba = model.predict_proba(X_test)[:, 1]

# Calculate the metrics
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
logloss = log_loss(y_test, y_pred)
roc_auc = roc_auc_score(y_test, y_pred_proba)

print("Accuracy: {:.2f}".format(accuracy))
print("Precision: {:.2f}".format(precision))
print("Recall: {:.2f}".format(recall))
print("F1 Score: {:.2f}".format(f1))
print("Log_Loss: {:.2f}".format(logloss))
print("ROC-AUC score: {:.2f}".format(roc_auc))
```

```
Accuracy: 0.53
Precision: 0.00
Recall: 0.00
F1 Score: 0.00
Log_Loss: 16.85
ROC-AUC score: 0.36
```

The metrics are shown as follows:

Accuracy: 0.53

Precision: 0.00

Recall: 0.00

F1 Score: 0.00

Log Loss: 16.85

ROC-AUC score: 0.36

6.4.1. Final considerations

As we could infer from the data preparation part of our notebook, the classes are too imbalanced. Even when applying oversampling techniques to address this issue, we can see that they are still not enough to help the model in its fitting phase. This results in a model with almost no prediction capabilities. Despite the discouraging results, the task could be repeated once enough data is collected to make the dataset more balanced for what concerns the field of interest. It would also be advised to collect data about different potential predictors, to further help the model in understanding the underlying pattern that could be hiding within the data.

References:

1) Häusser, Jan & Schulz-Hardt, Stefan & Schultze, Thomas & Tomaschek, Anne & Mojzisch, Andreas. (2014). Experimental evidence for the effects of task repetitiveness on mental strain and objective work performance. Journal of Organizational Behavior. 35. 10.1002/job.1920.
2) CTT Portugal Integrated Report
3) Lacity, M. C., & Willcocks, L. P. (2016). Robotic process automation at Telefónica O2. MIS Quarterly Executive.
4) Davenport, T. H., & Kirby, J. (2016). Just How Smart Are Smart Machines? MIT Sloan Management Review, 57(3), 21-25.
5) https://itrexgroup.com/blog/robotic-process-automation-cost/