School of Business and Management

Business Administration

Strategic Finance & Analytics

Automated data wrangling for the use of increasing cash flow analysts' work efficiency – a case study

Oskari Lauri

Examiners: Assistant Professor Jyrki Savolainen,

Professor Mikael Collan

Table of contents

ABSTRACT

Lappeenranta-Lahti University of Technology LUT

LUT Business School

Business administration

Oskari Lauri

Automated Data wrangling for the use of increasing cash flow analysts' work efficiency – A case study

Master's thesis

2024

50 pages, 6 figures, 8 tables and 2 appendices

Examiner(s): Assistant Professor Jyrki Savolainen and Professor Mikael Collan

Keywords: Robotic Process Automation, Cash Flow Analysis, Data Wrangling

The aim of this thesis is to study Robotic Process Automation's (RPA) effect on the efficiency and precision of menial data wrangling work carried out by cash flow analysts. The study consists of a literature review on the topic, and a real-life example of a data cleaning task carried out by analysis professionals working for a Helsinki based consulting firm is analyzed.

Previous studies on RPA's effects onto efficiency consist largely of interviews done on professionals. This thesis attempts a more quantitative and practical approach by measuring the performances of human subjects. The performance metrics of time and error rates are reported and compared agaist an automated Python script created in this project for wrangling data from a customer's online accounting platform. The results imply that in the

case study RPA was able to speed up the low-level analyst's tasks as the average time spent by experts is reduced by 98.6 per cent. This finding corroborates with earlier literature.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppatieteet

Oskari Lauri

Automatisoitu datansiivous kassavirta-analyytikoiden työn tehokkuutta ja tarkentamista varten – tapaustutkimus

Kauppatieteiden pro gradu -tutkielma

2024

50 sivua, 6 kuvaa, 8 taulukkoa ja 2 liitettä

Tarkastaja(t): Apulaisprofessori Jyrki Savolainen ja Professori Mikael Collan

Avainsanat: Ohjelmistorobotiikka, Kassavirta-analyysi, Datansiivous

Tämä tutkielma tarkastelee ohjelmistorobotiikan kyvykkyyttä parantaa kassavirta-analyytikoiden työn tehokkuutta ja tarkkuutta toistuvissa ja yksinkertaisissa työtehtävissä. Tutkielma koostuu aiheeseen pohjautuvasta kirjallisuuskatsauksesta, sekä tapaustutkimuksesta, jossa tarkastellaan helsinkiläisen konsultointiyrityksen työntekijöiden suorittamaa datansiivoustyötehtävää.

Aikaisemmat tutkimukset ohjelmistorobotiikan vaikutuksesta työtehokkuuteen koostuvat laajalti ammattilaisiin kohdistetuista haastattelututkimuksista. Tämä tutkielma hyödyntää käytännönläheisempää ja kvantitatiivisempaa lähestymistapaa tarkkailemalla ihmisiä työssään ja mittaamalla heidän suorituksensa. Tapausyrityksen asiakkaan datan muokkausta varten suunnitellun Python-koodin suoritusta verrataan kassavirta-analyytikoiden manuaaliseen työhön nopeuden ja tarkkuuden näkökulmasta. Tulokset viittaavat siihen, että ohjelmistorobotiikkaan perustuva ratkaisu voisi vähentää kassavirta-analyytikoiden toistuviin datansiivoustehtäviin käyttämää aikaa 98.6 prosentilla. Tulos on linjassa aiheesta aiemmin julkaistun kirjallisuuden kanssa.

ACKNOWLEDGEMENTS

My studies at LUT University have proven to be a boon for my professional skills and confidence as an analyst. I’ve picked up valuable skills during my studies in data analysis and financial theory. I extend my gratitude to the LUT staff for providing me with demanding yet insightful courses and a framework for academic knowledge. I would also like to seperately thank my thesis advisor Jyrki Savolainen, who guided me during this process by giving me practical pointers in searching for academic sources.

My gratitude also goes to my employer, who made this thesis possible, and by giving me support and understanding during the demanding times as I balanced between a wonderful yet demanding job and the making of this thesis.

Lastly, I would like to thank my spouse Jasmiina, for providing me with emotional support and courage during the challenging and stressful COVID-19 pandemic, which loomed during the making of this thesis.

I wish you a happy reading.

With gratitude,

Oskari Lauri

In Helsinki, Finland 31.01.2024

1 Introduction

Wrangling is the process which professionals partake in editing data in order to both make it easier to handle and to make it compatible for further analyses. The process of data wrangling often presents an obstacle for experts working in the finance, consultancy, and auditing sectors. Customer data is expected to be analyzed by professionals, but this data may not be available in a format used by these firms. This leads to them targeting time and manpower to wrangling. This task comes at a cost however, and analysts search for ways to increase the efficiency of wrangling – not only to save company resources, but both to eliminate mundane and repetitive tasks, and to give them more available time to carry out more complex tasks. Robotic Process Automation (RPA) is the use of automated software codes used for carrying out menial and repetitive tasks that would otherwise consume manpower.

This thesis deals with the automation of these processes with RPA, and how it can be utilized to speed up the data wrangling tasks carried out by the analysts. An empirical approach is taken where a group of analysts are compared against a tailor-made RPA-script.

1.1 Background and motivation

To be an analyst is to handle and refine information, and make deductions based on this process based on numerical analysis techniques. Analysis is becoming more important than ever - as more processes to produce data from our daily actions increases, there is more to analyze. The fact that more of what we do produces data to process raises a need to handle data more efficiently, as data becomes more abundant and available than ever before.

For analysts to gain information from data more efficiently, they employ a multitude of software and online platforms to perform automated calculations, and to check for errors and discrepancies. The use of these systems is reliant on data compatibility – that is, information must be in a format suitable for further use. Analysts must often obtain data from a source before carrying out their tasks, and more often than not, the data they acquire must be preprocessed into a predetermined format in order for them to use it further. This process is called wrangling, and it represents a significant use of time and effort for both businesses and professionals.

Wrangling presents an obstacle for experts working in the finance, consultancy and auditing sectors. Customer data is expected to be analyzed by professionals, yet this data may not be available in a format used by these firms. This leads to them targeting time and manpower to wrangling. This task comes at a cost however, and analysts search for ways to increase the efficiency of wrangling – not only to save company resources, but both to eliminate mundane and repetitive tasks, and to give them more available time to carry out more complex tasks.

There are many ways to approach wrangling. Larger firms with ample resources can invest in full cross system integration, removing any need for manual work. This approach is often costly and hence out of reach for smaller players in the private sector, who often have to rely on manual work that is both inefficient and difficult to standardize. A new option has recently entered the frey – robotic process automation, or RPA - the use of software robots that mimic manual human work.

Through my personal work experience in a consultancy startup, I have experienced both the need for wrangling and the manpower required for it, as raw data exported from various platforms must be processed into a predetermined format. Through this thesis, I have set out to research if RPA could improve the efficiency of analyst work by minimizing the time this step takes in building customer specific cash flow simulations.

1.2 Focus and methodology

This thesis is both a literacy review on RPA's effectiveness on finance professional work automation, and a case study centered around a common wrangling task carried out by cash flow analysts working for a Helsinki based consultancy firm. Datasets exported from an online accounting platform were given to cash flow analysts for manual wrangling. The data chosen represented real accounting data from a customer of the case company. The cash flow analysts' efforts was measured by clocking the time it took to perform this task, and the amount of errors they made was assessed. Errors in this context meant any failures to alter the structure of the data that did not account for the prerequisites for a successful data import demanded by the firm's online cash flow simulation platform.

1.3 Research questions

This thesis looks into the effect of RPA onto cash flow analyst work efficiency. Efficiency in this context is defined by the ability to carry out a given task in a sufficient quality in the least amount of time possible. Through this framework, the research questions of this thesis are;

1. What does the academic literature suggest on the benefits of RPA onto cash flow analyst work efficiency?
2. How can RPA be utilized in the context of this case study and does it increase the efficiency and precision of the analysts?

1.4 Structure of this thesis

First, a general view on the theory of cash flow analysis and and the current academic literature of robotic process automation is presented. Next, this thesis will detail a literature review on the effect of automation on analyst and accountant work effiency.

The theoretical framework is followed by a description of the case study and the setup of the research. The problem that the RPA solution set out to alleviate is detailed, and a summary of the Python script's main function is given. The results from the use of this code in relation to the case company's analysts manual performance is described, and the implications of this comparison will be detailed. The results of this comparison are further discussed, and the implications of the case study are also discussed. Lastly, the limitations of the study are detailed, and improvements for further study will be suggested.

The focus of this thesis is the effect of automation on the work efficiency of cash flow analysist's data wrangling tasks. The next two chapters will detail a the theoretical framework of subjects relevant to the focus of this case study. First, the fundamentals of cash flow analysis and reporting will be provided as they are stated in academic textbooks.

2 Theoretical framework

The focus of this thesis is the effect of automation on the work efficiency of cash flow analysist's data wrangling tasks. The next two chapters will detail a the theoretical framework of subjects relevant to the focus of this case study. First, the fundamentals of cash flow analysis and reporting will be provided as they are stated in academic textbooks.

2.1 Cash flow analysis

Cash flow analysis involves the monitoring and prediction of the outputs and inputs of a company's liquid cash reserve and the components that affect them. Carrying out cash flow analysis is a core part of maintaining business operations sustainably, as a company's operations cannot take place if it is unable to meet the demands of their creditors. Business leaders forecast their future cash flows in order to stay up to date on potential risks on their liquidity, operating capability and obligations, and to see if they will have the necessary resources for future planned investments. Companies are often required to report their cash flows in a statement as a part of their yearly reporting in order to message their stakeholders how they are spending money, as a responsible use of liquid capital is a key factor in upholding and growing the value of the firm and to ensure possible dividend payouts. (Fridson & Alvarez, 2011., p 80-81.)

Cash flows are typically reported in three main components, each representing the different types of utilizations of liquid assets. Each component can be calculated using entries of the financial statement. The cash flow statement is a useful tools for both shareholders and stakeholders, as a balance sheet and income statement only tell so much about whether or not a firm is maintaining liquidity in a satisfactory manner. A company may be profitable according to its profit and loss statement, but the cash made in carrying out core operations may not be retrieved timely enough to ensure payments for costs from overheads and salaries. The profit and loss statement may also include items that do not affect the cash flow, such as depreciations and amortizations. Likewise, it also does not list transactions that occur from investments and financing. The three core cash flow components are cash flow from operating activities, cash flow from financing, and cash flow from investments. (Fridson & Alvarez, 2011., p 80-86.)

2.1.1 Cash flow from operating activities

Cash flow from operating activities refers to the cash generated and spent during a firm's core operations. These operations encompass all cash transfers related to revenue, overheads, personnel, and operating expenses. This cash flow is often a crucial component to monitor and forecast, as it signifies the sustainability of a company's value creation model. In addition to the mentioned transactions, cash flows associated with activities that cannot be allocated either to financial cash flows or to cash flows from investments can also be categorized as operational cash flows, such as other operating income. Cash flows resulting from a firm's core activities are primarily used to fulfill the main responsibilities of companies and cover the costs associated with raising equity and capital. (Brealey et. al. 2022)

Cash flows from operational activities can be reported in two ways: the direct method and the indirect method. The direct method reports cash flows as they happen when a company receives or transfers cash away as it conducts transactions. The direct method is often thought of as a more accurate method, as it produces information per every transaction in real time as documents are recorded. When employing the direct method, transactions are catecorized according to their primary utility, such as customer income and supplier costs. While the direct method requires more effort in producing cash flow information, it is more favored due to it providing more data for advanced cash flow forecasting methods. (Fridson & Alvarez, 2011., p 80-86.)

The indirect method involves the use of accrual accounting information and the net income stated in the income statement. Net income is adjusted for changes in the balance sheet in order to derive the total change in cash caused by operating activities, such as depreciations and amortizations. The indirect method is more widely used in cash flow analysis, as its application necessitates little effort in producing information, and because it stresses the easily available information on the difference between net profit and net cash flow, and its tendency to stress the impotance of changes in working capital. (Fridson & Alvarez, 2011., p 80-86.)

2.1.2 Cash flow from financing

Cash flows from financing refer to the inflows and outflows of cash resulting from a firm's activities in gaining capital to fund its activities. These cash flows involve obtaining and repaying capital from a multitude of different sources, such as buying back and issuing stocks, shareholder contributions, issuing or repaying debt, and distributing dividends to stockholders.

Analyzing cash flows from financing provides several benefits. First, analysts can monitor the changes in companies' capital structure and their reliance on different sources of capital. The degree to which a firm uses debt or equity to reach is goals in relation to other firms acting in the same market provides valuable information to analysts, who often compare companies acting in the same market. This component also allows analysts and stakeholders to evaluate the financial health of a firm, as positive cahs flows from financing may indicate strong confidence from investors. As dividend payouts are a part of this category, cash flows from financing also act as a indicator of how dividend payouts meet investor expectations.

Cash flows from financing can also reveal key changes in capital restructuring and financing strategies, and as such reveal details about financial strategies, such as efforts to reduce debts, optimize capital costs and strengthren the balance sheet. Comparing cash flows between competitors allows for the analysis of companies meeting investor preferences.

2.1.3 Cash flows from investing activities

Cash flows from investing are the third category typically presented in a cash flow statement. This group includes the sales and purchases related to long term assets and investments. The key components in this group include the cash inflows and outflows related to long term assets, acquisitions and mergers, and investments in financial instruments. By reporting cash flows from investing, companies provide information to investors that they can use to evaluate their capability to make long term decisions that help in keeping a company competitive in their respective market.

When analysts look at cash flows from investments, they are often looking at how companies are using their cash in making investment decisions, and whether those decisions lead to more sustainable and profitable practices. Investing decisions also act as an indicator of making strategic iniatives and plans for future growth and diversification. These decisions can be compared to other actors in the same market in order to provide insight which firms are managing their investments efficiently.

2.2 Literature review

The goal of this of this chapter is to provide a summary of RPA as it is discussed in academic literature. First, a general overview on robotic process automation and key concepts in relation will be provided. Secondly, this chapter will look through literature found from the EBSCO database that conveys the results and challenges of implementating of RPA in increasing the productivity of accountant and analyst work.

The papers for this review were found by typing the following search terms: ”robotic process automation”, ”accounting”, and ”empirical study OR empirical research OR data OR method OR experiment”.

This search yielded fourteen results after filtering for peer-reviewed papers written in the English language. These papers’ findings are summarized in the theoretical framework section, along with a general summary of cash flow analysis. Additional papers were found by looking through the sources of the papers found via this search.

2.2.1 Overview of Robotic Process Automation

Robotic process automation (RPA) is a term that encompasses the software based tools whose purpose is to replace human labor in carrying out data related tasks. ”Robotic” in this context doesn’t refer to physical robots, but to software that imitates humans interacting with different system interfaces. RPA is typically used to automate tasks that are seen as mundate and repetative, without making changes into the information systems themselves. These tasks may include data exporting, compilation from multiple systems and basic recurring analyses. (van der Aalst, Bichler & Heinzl, 2018.)

In 2020, the global revenues were of the RPA market was estimated to be more than four billion USD (Statista). Many sources forecast the market to grow at an annual rate between 30 % and 50 %. (Lacity & Willcocks, 2021)

RPA is often seen as a transitional lower cost middle ground in automation. This comes in contrast with cross system integration. While integration is done to repetetive data related tasks that are carried out in bulk numbers and frequently, RPA is often seen as best reserved for tasks that – while also repetetive, are not carried out as often. This is largely due to the high cost of system integration investments and the lower cost of applying RPA. (van der Aalst, Bichler & Heinzl, 2018.) McKinsey estimates that more than 69 % of data proceccing tasks and 64 % of data collection tasks could be handled with RPA (McKinsey 2019).

Robotic process automation has seen a significant spike in interest in the finance sector. Most large players players in the market are currently looking into the implementation of RPA or even impementing it in their daily operations. It is seen as likely that RPA will evolve into a standard practice necessary for staying in the field. (EY, 2016.) Cooper et al (2019) state that accounting services utilize RPA mainly in tax and assurance services.

Hoffmann, Samp and Urback (2019) characterize RPA by using a framework illustrated in Figure 1. They sum up RPA as "software robots following a choreography of technological modules and control flow operators while operating within IT ecosystems and using established applications." They recognize RPA as a fast and easier way to automate processes compared to other automation approaches, with the caveat that it may complicate governance structures, and that organizations applying it in their daily operations should account for this long term while adopting RPA.

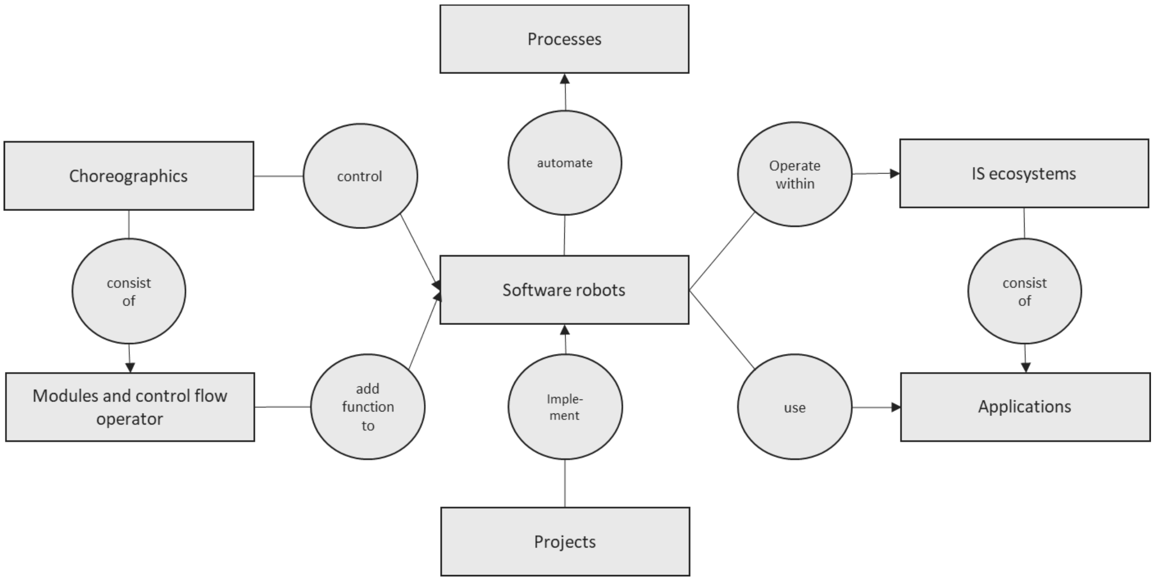

Figure 1. Illustration of RPA's nature in organizations. Adapted from Hoffmann, Samp and Urback (2019).

The successfullness and necessity of implementing RPA is dependant on a number of criteria. Effective automation via the use of software robots requires that the tasks can be written as instructions, that these instructions apply to said tasks, and that the utility of automating them is worth the effort. Asatiani and Penttinen (2016) recognize eight key criteria for RPA:

- High transaction volume – the actions to be automated are performed often, or they include a high number of sub tasks
- Need to access multiple systems – the task concerns accessing multiple systems, such as copying and compiling fata from a certain file format and transferring them to a database
- Stable environment – The systems in which the actions are carried out remain same whenever performed, and are not subject to large changes
- Low cognitive requirements – The need for creative thinking, subjectivity and complex interpretation is minimal
- Easy decomposition into unambihguous roles – the automated function can be divided into clear steps that follow logical rules
- Proneness to human error – the task will benefit from a computer carrying it out, eliminating human specific error

- Limited need for exception handling – The task is highly standardized and involves little risk of deviation
- Clear understanding of the current manual costs – The company is able to deduce how the benefits of RPA generated efficiency would compare to the investments necessary to automate the tasks

Hoffmann, Samp and Urback (2019) recognize two key performance indicator (KPI) areas that decision makers can use to measure the entrepreneurial impacts of implementing RPA. The first KPI group involves the internal affairs of the company. This includes employee efficiency enhancement, work satisfaction, process acceleration and saving costs. The other group of KPI considers software robots' influences on external factors. These include customer satisfaction, stakeholder cooperation, and the company's stock value. This framework considers the internal performance factors as bridges to external performance.

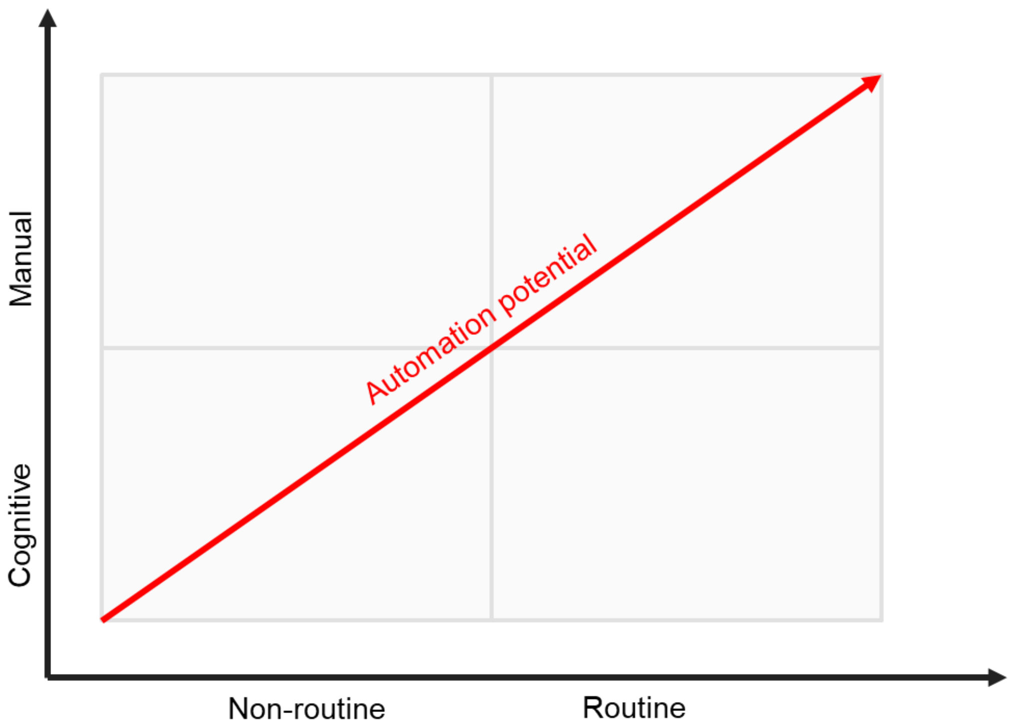

Figure 2. Framework for automating tasks. Adapted from Frey and Osborne (2013).

While RPA shows much promise for many companies, its adoption does not always go smoothly. Ernst and Young estimate that between 30 and 50 percent of initial attempts in RPA adoption end in failure (EY, 2016). The surge in RPA's popularity has also given a rise of demand for a more comprehensive outlook on its adoption to better understand its effect. Zhang et al. (2022) estimate that RPA processes have potential to underline existing troubles in the digitalization advances of the private sector while also introducing new problems, and that these issues will demand more attention in the coming years.

A possible pitfall in establishing RPA processes is short-sightedness. As RPA holds a promise of increased return on investment, it can lull companies into adapting it in a way that does not account for the long-term implications and sustainability of said solutions, leading to faulty implementations. (Zhang et al 2022.) To gain a steady flow of Return On Investment (ROI), companies should initially focus on using RPA to automate simple and stable processes of low complexity, working their way up to more complex processes once more experience is gained (Harrast, 2020).

2.2.2 RPA and efficiency of accounting and finance workers

This portion will list publications that have investigated the effects of RPA onto worker efficiency and company performance. Emphasis was put into case studies that aimed to measure the effects in firms in the field of accounting and finance.

Cooper et al. (2019) investigated the implementation of RPA software in public accounting by carrying out interviews of fourteen managers, directors and partners who worked for the Big 4 firms. The work experience of the respondents varied between 7 and 31 years. Interviewees were selected based on the degree of experience in RPA implementation. They find that RPA leads into significant efficiency and effectiveness increases, with one accounting company disclosing that it had saved over one million work hours due to RPA adoption. One respondent reported a 16-hour workload having been turned into a process that takes 17 seconds, and many firms increasing task accuracy rising from 90% to 99.9%. The research pointed towards the Big 4 firms sharing a high interest into the costs and benefits of implementing RPA, and that they are generally favoring buying licences to RPA enterprise platforms from providers like UiPath and Automation Everywhere.

The implementation of RPA in the Big 4 firms was further researched by Cooper et al. (2022), when they examined whether there were differences in perception between firm leaders and lower-level employees, and the factors that contributed to any possible differences. The study found that overall, both leaders and lower-level employees held positive perceptions of RPA and its potential benefits for the accounting industry. However, they noticed some differences – leaders tended to see RPA more as a strategic advantage due to its potential benefits for greater efficiency and cost savings, while lower-level employees were more likely to see RPA as a threat to employment due to need for additional training and job displacement. Lower-level employees were also less likely to report work satisfaction because of RPA implementation. The authors suggest that promoting an open and collaborative working environment can help to alleviate these differences.

Fernandez & Aman's (2018.) case study conducted eleven interviews among professionals employed in global accounting services to map attitudes about RPA in relation to individual work performance and organizational positioning. The respondents showed generally favourable and optimistic attitudes towards RPA's potential to increase accountants' efficiency in menial tasks, along with work quality and accuracy in handling large amounts of data, and that time saved from data wrangling can be used more for complex analysis task that necessitate human expertise.

Fernandez and Aman (2018.) also highlight the importance of adapting to RPA as a more prominent technology – future accounting workers should prepare to adopt IT expertise as a larger part of their toolkit as new tools will be introduced with increasing frequency, which necessitates wide support from management and IT workers. Although the respondents generally agreed that RPA will alleviate shortages in human resources, some saw that it may threaten employment of finance workers incapable of adapting, echoing the results of Cooper et al. 2022. One respondent in Fernandez and Aman's study reported that effort saved via RPA can enable analysts to use more of their time strategy analyses that draw most professionals into the field in the first place.

Hsiung et al. (2022) distributed 140 questionnaires to 70 different small accounting firms across Taiwan, out of which 102 were deemed valid for analysis. The findings of the study suggest that the introduction of RPA systems had a positive impact on the efficiency and productivity of small accounting firms in Taiwan. Specifically, RPA systems were found to reduce the time required for accounting tasks, minimize errors, and improve the accuracy of financial statements. Moreover, the implementation of RPA systems was associated with a reduction in labor costs, as firms were able to achieve more with fewer employees.

The authors (Hsiung et al. (2022) recognized that a higher degree of use tends to correlate with larger satisfaction in the used RPA system and perceived utility of an RPA system and the information it produces. They also recognized that a firm's CEO's support in system adoption has a significant positive correlation with factors that indicate successful RPA adoption. The authors also analysed the differences between genders of the respondents and found that male employees reported more positive experiences in RPA adoption.

The literature review by Jedrzejka (2019) concluded that RPA stands out among automation solutions in the accounting field due to its availability, driven by its affordability and easier applicability, and that the increases in efficiency RPA provides may lead to entry level accounting jobs to become more complex due to repetitive tasks being automated. Jedrzejka predicts that accountant jobs will transform due to increased automation, and that new jobs will be created due to the increased time available for more complex tasks that demand complex professional judgement, and that entry-level jobs may be lost to automation.

Kedziora & Kiviranta (2018) carried out interviews on seven informants working in RPA implementing companies in northern and central Europe. The interview included questions about the details of the RPA projects the informants had participated in, their views on what methods suited for which projects, the cost-benefit ratios of the different RPA implementations, and their views on how RPA may change their organizational role in the future. The informants they interviewed did not imply significant job losses due to RPA adoption, but rather the jobs of accountants being transformed from data handling to being more consulting oriented. Informants reported RPA as a viable option for increasing process efficiency due to its lower entry barrier and easier implementation.

Kokina and Blanchette (2019) conducted a multiple case study by carrying out semi-structured interviews to RPA adopters to shed light on the successes of RPA task suitability, the implementation of RPA in organizations, and its impact on company performance. They screened for RPA adopters with experience using UiPath, Blue Prism or Automation Everywhere, and who worked for organizations that either had already adopted RPA for the use of finance or accounting functions, or had strongly considered utilizing it in said areas. The participants were picked via two anonymous companies, whom the authors describe as (1) one of the largest RPA providers in the world, and (2) an international organization of financial professionals, landing sixteen interviews from eleven participant companies for the study.

The study found that the respondent firms had developed internal scorecards and Return On Investment (ROI) calculation methods specifically to assist in implementing RPA solutions and measuring their effectiveness, with said methods taking saved labour hours, effectiveness, and reductions in temporary workers into account. The authors note that as a measure of efficiency, reductions in work hours might be subject to bias due to process managers underestimating the time manual work in the absence of RPA takes. Participants reported that work carried out to assess the necessity of RPA implementation forced them to reassess the methods chosen to carry out tasks, and as an unintended consequence, some tasks were improved regardless of RPA adoption. For tasks that were automated, the participants reported more transparency in their processes, as RPA use is often recorded in data gathering purposes. On top of increasing efficiency, the companies reported many qualitative improvements, such as better understanding of internal processes and updating their job skills.

Kokina et al. (2020) continued the multiple case approach used in the aforementioned 2019 research to interview 18 participants from 8 different companies worldwide. The focus of this study was to assess how RPA is affecting the roles of accountants within their workplaces. Through their research, they were able to identify five key roles finance professionals tend to fall into after RPA has been implemented. The roles Kokina et al. recognized are Identifier, Explainer, Trainer, Sustainer and Analyzer (figure 3).

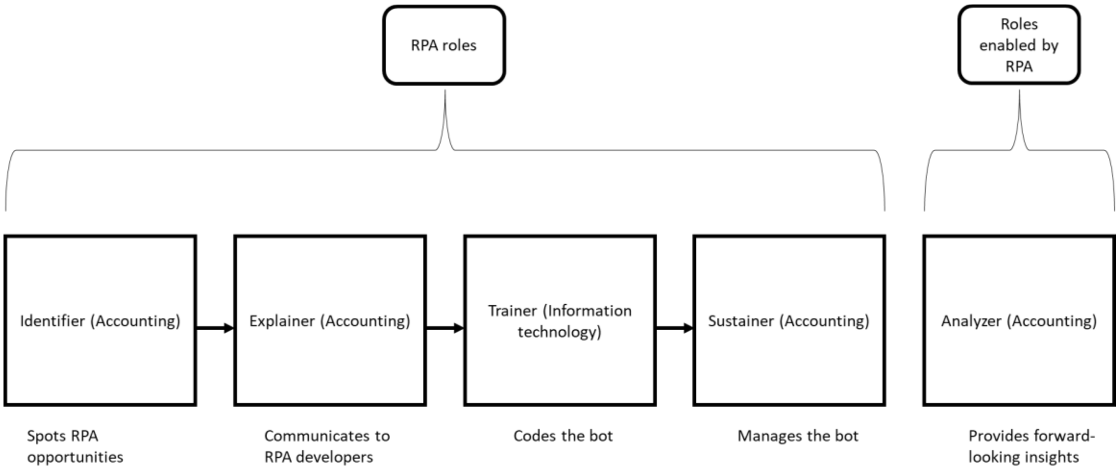

Figure 3. Roles within RPA adopting organizations. Adapted from Kokina et al. (2013).

Four of the identified roles work within RPA implementation. Identifiers are accounting workers who recognize processes that are suitable for automation. Since they themselves work with said processes, accountants can provide insight into how rule based, mundane and prone to exceptions each task is, and recommend RPA adoption for where it would likely have a positive Return On Investment. Explainers are accountants who document the processes that are suitable for RPA, and who communicate and cooperate with software designers and developers in creating the automation solution for tasks. Trainers are directly involved in 'training' the software robots in handling the mundane tasks that are chosen for automation. The authors describe Trainers as accountants with strong a strong IT background who can account and even code for exceptions in automated data retrieval tasks, such as pop-up windows and password expiration messages. Sustainer refers to IT trained accountants who are tasked with bot upkeep and monitoring in case of automated steps becoming obsolete. The systems used in accounting tasks are subject to changes and updates, which a bot cannot distinguish on its own, causing a bot that has been programmed for a certain task to become unable to carry out its function. Sustainers account for these changes and assist in reprogramming and updating the bots so that they can fulfil their functions.

The four aforementioned roles deal with RPA implementation itself. The fifth and final role – Analyzer, is a role that is born once automation is successfully introduced in organizations. As RPA deals with mundane and repetitive tasks and frees up time from accountants and finance forkers, they are granted room to deal with non-manual tasks. Kokina et al. (2021) describe analyzers as workers who use the extra time provided via RPA to provide their employers with insights and qualitative information based on the data that, without RPA, would consume more of their time in processing related tasks, and who are therefore able to help their organizations better in reaching their strategic goals.

Korhonen et al. (2021) conducted an interventionist case study on a Finnish machinery manufacturer, and analysed how well their management accounting tasks could be automated. The authors found that the case firm pricing processes had initially seemed more suitable to automation than they had initially estimated. Experts working for the manufacturing firm had to use judgement to assess efficiency need and cost pressure in ways that could not be automated easily. The manufacturer was hesitant to use automation for large deals – officers in the firm were using different tools for deals in parallel instead of adopting the automated pricing system universally. The authors noted that this was because the parameters of automation were not completely agreed upon among the firm's hierarchy. The authors warn against attempting automation on processes that seem compatible at first glance and urge firms that delve into it to deeply familiarise themselves with the processes they aim to automize to avoid irrational investments into efficiency.

Yunus, Aman and Keliwon (2019) conducted interviews on business leaders who had implemented automation to find out key challenges and success factors in RPA implementation in accounting firms. They carried out six face-to-face interviews with managers utilizing a conceptual framework that stressed the importance of four key areas: idealized influence, inspirational motivation, intellectual stimulation, and individualized consideration. The respondents brought up challenges in IT innovation, saying they had faced resistance from workers who had grown accustomed to their existing working models, and that fear of losing their employment due to RPA adoption was a common obstacle in getting staff on board with change. The authors found multiple ways of alleviating these challenges. First, management should be involved in RPA adoption and actively act as ambassadors for the new technology. Second, employee education should be encouraged in an open minded environment in order to help them adopt RPA. Third, knowledge of the benefits and uses of new technologies should be shared openly inside the firms looking to adopt innovations.

2.3 Summary of the literature review

The literature review shows that RPA has been looked though its effect on daily finance professional work, and the efficiency and roles of said workers in their respective organizations. The review shows that the efficiency of software robot application can vary greatly based on how RPA adoption is communicated within firms, and how its full scale adoption can also change the roles of workers within an organization. When utilized correctly, RPA can be a boon for efficiency of finance professionals, but its incorrect application can also lead into unnecessary investments due to some processes being too reliant on subjective judgement.

Table 1. Summary of the literature review on RPA's potential to increase the efficiency of finance workers.

Title	Author(s)	Year of publication	Research topic	Key findings
Perceptions of Robotic Process Automation in Big 4 Public Accounting Firms: Do Firm Leaders and Lower-Level Employees Agree?	Cooper, Holderness Jr., Sorensen & Wood	2022	Perception differences in RPA adoption between higher ups and lower-level employees.	RPA is highly favored by managers and seen as a strategic advantage. Some lower level employees saw RPA as a threat to employment due to it lowering the need for manual work.
Robotic Process Automation in Public Accounting.	Cooper, Holderness, Sorensen & Wood	2019	RPA software adoption success. Interviews on managers working for the Big 4.	RPA leads into high savings in work hour allocation due to increases in efficiency.
Impacts of Robotic Process Automation on Global Accounting Services.	Fernandez & Aman	2018	Interviews on global accounting professionals regarding RPA's effect onto work performance and organizational positioning.	Advanced accounting jobs may become more complex due to RPA eliminating more menial tasks.
Research on the Introduction of a Robotic Process Automation (RPA) System in Small Accounting Firms in Taiwan	Hsiung & Wang	2022	Efficiency and productivity after RPA adoption. Carried out via an online questionnaire sent to small accounting firms.	RPA has a positive effect on worker efficiency in accounting firms.
Robotic process automation and its impact on accounting	Jędrzejka	2019	Literature review on RPA effectiveness.	RPA is easy to implement due to its availability. Entry level accountant jobs may become obsolete due to increases in efficiency.
Digital Business Value Creation with Robotic Process Automation (rpa) in Northern and Central Europe.	Kedziora & Kiviranta	2018	RPA implementation success in Europran firms, carried out via interviews on seven informants.	RPA adoption did not lead in the removal of jobs, but jobs were transformed due to increases in efficiency.
Early evidence of digital labor in accounting: Innovation with Robotic Process Automation.	Kokina & Blanchette	2019	RPA task suitability and implementation success. Carried out as interviews.	RPA adoption leads to internal validation methods for measuring RPA success, and it also leads to icreases in process transparency.
Accountant as Digital Innovator: Roles and Competencies in the Age of Automation	Kokina, Gilleran, Blanchette & Stoddard	2020	RPA's effect on the role of accountants. Carried out as interviews.	RPA adoption leads to the formation of specific roles that optimize its utilization.
Exploring the programmability of management accounting work for increasing automation: an interventionist case study.	Korhonen, Selos, Laine & Suomala	2021	RPA's success in automating management accounting tasks.	RPA adoption may fail when attempted upon processes that demand much human judgement, such as sales pricing.
The Role of Business Leaders in Information Technology Innovation in the New Era of Disruptive Technology.	Yunus, Aman & Keliwon	2019	Key success and obstacle factors in RPA implementation.	RPA adoption faces challenges from employees who feel threatened by its adoption, and organizations should adopt policies that alleviete doubts from workers who feel threatened.

3 Case description

Consulting Firm A is a consulting firm located in Helsinki, Finland. Focusing on serving small and medium sized firms, their main service is a monthly consulting session between their own representatives and their customers' decision makers. The service involves the use of a SaaS-style online platform, which the customers can access to view key metrics at any time. These metrics include key performance indicators, roadmaps, financial and managerial accounting data, and a modelled cash flow that can be used for forecasting the customers' cash balance. A key part of Consulting Firm A's service is to observe changes in management decisions, and how they may affect the future cash flows of their customers.

Consulting Firm A utilizes consultants and analysts in delivering their key service, and it is up to the analysts to extract data from the customers' information systems, and to use this data to model the customers' cash flows and KPI dashboards. Consulting Firm A's own platform can automatically extract accounting data from two of the most used accounting software in Finland. Other software, however, demand a more manual approach.

For systems that have not been integrated, analysts must log in to the systems and extract the data via data table file exports and convert these exports into compatible CSV files. The effort this demands is dependant on the amount of steps required to convert customer data into a compatible form. Consulting Firm A is currently looking into RPA and automated Python codes as a more cost-effective alternative to full system integration for every accounting software their customers use. They wish to analyze the possible ROI in using analyst time in creating a library of Python scripts that can automate data wrangling tasks.

An interventionist case approach is adopted for the purposes of this thesis. First, an accounting system that demands manual data wrangling to be compatible with Consulting Firm A's system was selected. The system was chosen on the basis of low overall analyst experience to factor out the effect of customer system familiarity on efficiency, as some analysts within the firm are more familiar with certain accounting software data. Fennoa has not been in use most of their customers, meaning the analysts chosen for this study are more evenly skilled in handling the data. A code was created to automate a manual data wrangling process using the data from said system. Analysts from Consulting Firm A were asked to perform the wrangling manually using their own preferred methods, and the time taken to perform said task was recorded. Their performance time was compared to the bot, and the errors in task performance were assessed in relation to it.

3.1 Cash flow system description

Consulting Firm A uses an online platform as a part of their core service, which their customers can access at any time with a username and a password. The platform is fed information in the form of CSV files that are required to contain data in a given format in order for a successful data import to occur. Analysts working for the firm export data from accounting software and wrangle it into the predetermined format before modelling a cash flow with these figures and analyse how different drivers affected by managerial decisions affect cash flow projections.

The system uses accounting data in a monthly format. Monthly figures for each account name are given in table form with rows representing variable/account names, and with columns representing months. The values must be expressed in raw numeric form, with commas separating decimals. This means that any unit symbols, such as kilograms, euro or dollar signs, or percentages are not allowed. A model of a correctly formatted dataset is illustrated in Figure 4. The system can record only one variable per given name. In case of a dataset containing multiple sets of monthly values for one value name, only the last row of said name will be recorded. This leads to a need for the case company's accountants to manually check each dataset for any repeated variable names for differing values so that any crucial information is not lost in translation.

	08/2021	09/2021	10/2021	11/2021
Variable 1	100,00	4223,10	4500	122,10
Variable 2	200,50	44,01	124	1256,10
Variable 3	330	500	2550	242,10
Variable 4	450,01	256	156,67	464,45
Variable 5	504	960	124,45	144,12
Variable 6	770	345,10	1001,20	5000

Figure 4. Structure of a correctly formatted dataset

The monthly values must be expressed differently depending on whether the account belongs in the balance sheet or the income statement. For values corresponding balance sheet accounts, a year-to-date cumulative value for the given month must be given. For values corresponding an income statement account, a monthly value must be given.

The cash flow model's hierarchy is designed to reflect a standard Finnish balance sheet and income statement. The names of each account in the customers' balance sheet and income statement acquired from imported raw data is inputted into a hierarchy tree (illustrated in Figure 4). For example, if the customer was to record their sales in two separate accounts numbered 3010 and 3020, the sums of accounts 3010 and 3020 would be added into bin "Revenue", with the same being carried out for each account. Alternatively, if the data has a summary value for all revenue, that too can be placed in the bin instead of each individual account. The analysts usually favour individual accounts instead of summary values due to each account having their own specific behaviour, and therefore, a unique forecasting logic. For firms with a light accounting structure with a small amount of accounts however, summary values can also be used. During a new customer's onboarding process, analysts working for the case company must match each customer accounting data variable name with each predetermined account bin in order to run a successful cash flow simulation. For the simulation to occur successfully, the accounts must be tied to each bin correctly, and the account names tied to each bin must hold the correct monthly values after they're imported into the system.

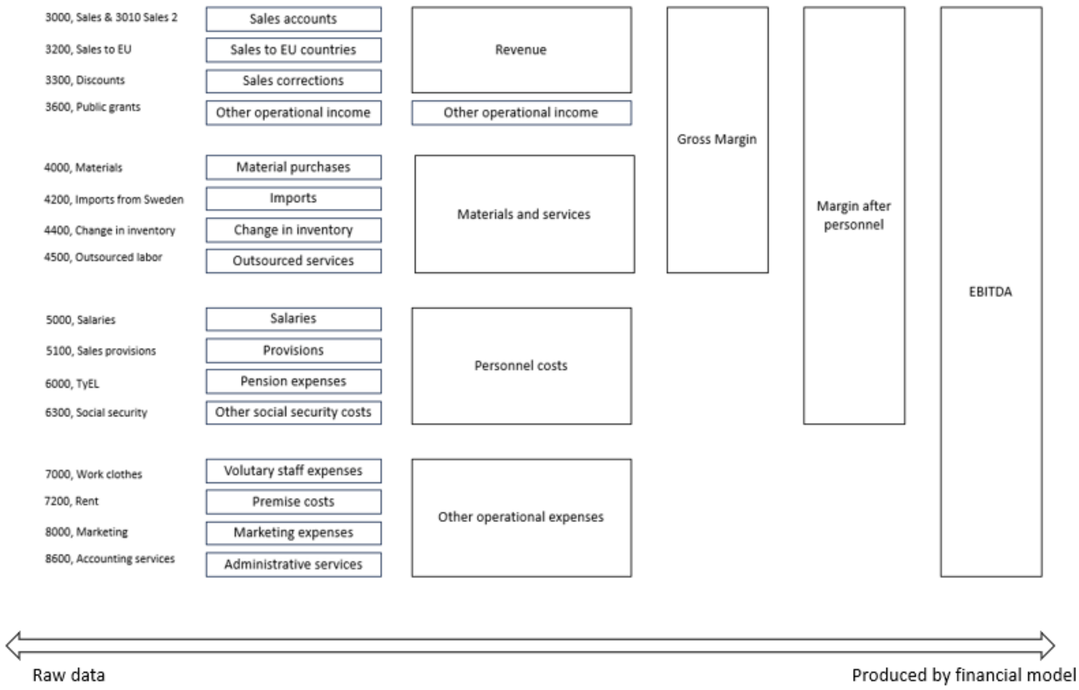

Figure 5. Illustration of how raw data account values are modelled into summary and derivative values in the case company's cash flow simulation platform. Raw data variables are illustrated as text on the left corner. Bins that either raw data or values produced by the financial model are placed into are represented as boxes.

The sum values represented by the bins are subjected to additional calculations during simulations to calculate derivative key figures, such as equity ratio, days sales outstanding and the three cash flow components. The case company's analysts perform validity tests to simulation results by checking that the results don't indicate any anomalies, such as mismatching sums for assets and liabilities, or unaccounted changes in equity. During a new customer's onboarding process, any errors in data imports would be noticed during these validity tests, and the error would need to be traced back via these tests and by additional scanning of source data. These tests consume analyst time, and this time could be reduced by ensuring that the data is imported in the correct format before the initial run.

3.2 Customer accounting data system and wrangling process description

The customer accounting data used for the purposes of this thesis was gathered from Fennoa, a Finnish online accounting platform. Fennoa allows its users to export Excel files of monthly data in a fixed, rolling twelve-month time frame format, with both balance sheet and income statement account values in the same file. The monthly values for the balance sheet accounts represent monthly changes in account balance, with the accounts' total balance being shown only on a starting month of an accounting period, and the income statement accounts are given in a monthly format. The fixed twelve-month format and the way monthly values are represented for balance sheet data represent a persistent obstacle for analysts working for the case company, who must manually wrangle the data into the correct format.

First, the analysts must convert the monthly balance sheet account values from changes in total balance to monthly total balance values. Second, they must scan all the names of the accounts for any duplicates and ensure that no two variables that represent different accounts share the same name. Failure in carrying out this step could potentially result in a distorted balance sheet and cash flow simulation, as an account value could end up gaining a false value, or a crucial value might be lost due to another account sharing the same variable name during a data import. This obstacle stems from the case company's system only accepting one value per variable name. Analysts working for them circumvent this obstacle by manually changing the account names in the exported raw data so that they are all unique. While this method allows data to be imported successfully, this adds an additional task for the analysts, therefore increasing the amount of steps where a time consuming error might occur.

Table 1. Illustration of raw data from Fennoa. The light gray column represents the start of the fiscal year, while the dark gray column represents the summary column.

Account	1.12.2020 - 31.12.2020	1.1.2021 - 31.1.2021	1.2.2021 -	1.3.2021 - 31.3.2021	1.4.2021 - 30.4.2021	…	1.11.2021 - 30.11.2021	1.12.2020 - 30.11.2021
BALANCE SHEET								
BALANCE SHEET ASSETS								
LONG TERM ASSETS								
Tangible assets	-1000	195000	-1000	-1000	-1000		-1000	185000
Land	0	195000	0	0	0		0	195000
1100 Real estate	0	95000	0	0	0		0	95000
Buildings	0	100000	0	0	0		0	100000
1150 Warehouse	0	100000	0	0	0		0	100000
Machinery	-1000	95000	-1000	-1000	-1000		-1000	85000
1161 Machines	-1000	95000	-1000	-1000	-1000		-1000	85000
…								
BALANCE SHEET LIABILITIES								
EQUITY								
Share capital	0	50000	0	0	0		0	50000
2001 Share capital	0	50000	0	0	0		0	50000
Profit from previous fiscal periods	75000	95450	0	0	0		0	95450
Profit (loss) of the period	20450	4500	3160	4360	4260		400	7060
SUM OF EQUITY								
LIABILITIES								
Long term								
Loans from financial institutions	-2500	100000	-2500	-2500	-2500		-2500	75000
2600 Bank loan	-2500	100000	-2500	-2500	-2500		-2500	75000
…								
Short term								
Loans from financial institutions	-1000	25000	-1000	-1000	-1000		-1000	15000
2800 Bank loan	-1000	25000	-1000	-1000	-1000		-1000	15000
...								
PROFIT AND LOSS STATEMENT								
REVENUE								
Sales accounts	55000	36700	36700	36700	36700		36700	403700
3000 Sales	55000	36700	36700	36700	36700		36700	403700
Materials and services	-12000	-15000	-15000	-15000	-15000		-15000	-165000
Purchases								
4000 Purchases	-12000	-15000	-15000	-15000	-15000		-15000	-165000
…								
9900 Prepaid income taxes	-200	-200	-200	-200	-200		-200	-2200
PROFIT (LOSS) OF THE PERIOD	1100	4500	-1340	1200	-100		-100	3560

3.3 Data and methodology

To explore the potential benefits of RPA to analyst efficiency in data wrangling, the performances in efficiency of three analysts were measured. The three analysts included two full-time employees and an intern, who all had experience in wrangling accounting data for the purposes of creating a cash flow model in Consulting Firm A’s system. Two of the analysts had no experience in wrangling data from the chosen customer accounting system, and the analyst who did had only built one cash flow model with data exported from said system. The low amount of experience with Fennoa was factored in the analyst selection in order control for analyst familiarity with the system, and to account for the time analysts would need in order to formulate an approach in carrying out the task. The analysts would time their performance in Toggl, an online work hour recording platform used by the case company in their efficiency tracking.

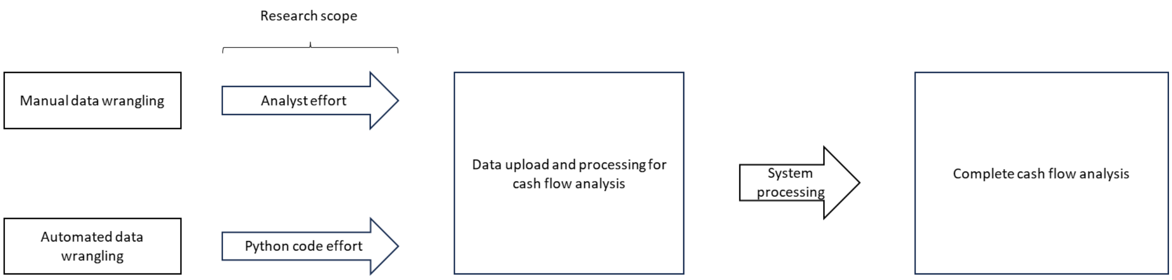

Figure 6. The thesis research setting.

The analysts were given real customer accounting data from Fennoa from the time frame of three years. They were given a list of instructions detailed in Appendix 2.

The only priming carried out in the setup was the choice to inform the participants that the individual files contained both balance sheet and income statement data. This approach was chosen because by directly giving the customer data files to the participants, the time the analysts would spend in acquiring the data from the online platform was eliminated. By eliminating this step, the opportunity to see the data's content in the system prior to exporting it in Excel files was also removed. The choice to eliminate the step of data retrieval was carried out so that any time used for it didn't obscure the time used for data wrangling, which was the focus of this case study.

On top of measuring the time the analysts took in the wrangling task, the performances of each analyst in producing compatible data were assessed. As stated previously, the cash flow modelling system accepts data only in a certain format, and possible values that share a name with other values may be lost when importing data. A list of prerequisite wrangling goals was drafted to assess the successfulness of the wrangling work, and the datasets produced by the study participants were checked for the listed criteria. The criteria were:

- The data is expressed in table format with rows representing variables and columns representing months
- The column titles must be expressed in the format 'MM/YYYY'
- Every row and column with a value must hold a title
- The values must be expressed in numeric form with no letters or unit characters

- Decimals must be separated with a comma
- All values that are to be separated into bins in the cash flow model must have a unique variable name

The code used to simulate the RPA wrangling task was made with Python utilizing its ‘pandas’ data table manipulation package. The script was programmed to combine information from several CSV export data tables from Fennoa, and to wrangle the information into a compatible form for the analysts to use. The code, detailed in Appendix 1, carries out the following operations:

- Access the CSV files in a set file path and combine the information into a single data table
- Remove the final sum column
- Add a column named “Tili”
- Fill the column with text values representing the balance sheet and profit and loss statement category names.
    - The names of account categories are represented in a separate list specified in a separate Excel sheet
    - If the row title column representing account names holds a category name, add the same text value to the “Tili” column. If it doesn’t, add the same value to the column as before
- Check the account name column for any names that appear more than once. For account names that appear more than once, add the value from “Tili” column to the end of the account name, inside brackets
    - This leads to account names that appear twice to have a distinct marker indicating which part of the financial statement the account belongs to. For example, the row “Loans from financial institutions” will be renamed “Loans from financial institutions (long term liabilities)” and “Loans from financial institutions (short term liabilities)”
- Remove the “Tili” column
- Rename the column titles to the date format “MM/YYYY”

- Add a column named “Sijainti”. All rows before the row title “Revenue” will be balance sheet values, everything after will be profit and loss statement values
- For balance sheet values, change the monthly values into a cumulative sum, starting from the first numeral value column
- Remove the “Sijainti” column
- Remove all rows that contain only zeroes

This procedure results in a data table where the monthly values will represent cumulative balances for balance sheet values per accounting month, and monthly changes for profit and loss statement values. Any account names that appear twice will be renamed so that the names are unique to prevent any data loss during the import.

Table 2. Fennoa data that has been wrangled with the Python code.

	12/2020	1/2021	2/2021	3/2021	4/2021	…	11/2021
Tangible assets	196000	195000	194000	193000	192000		185000
Land	195000	195000	195000	195000	195000		195000
1100 Real estate	95000	95000	95000	95000	95000		95000
Buildings	100000	100000	100000	100000	100000		100000
1150 Warehouse	100000	100000	100000	100000	100000		100000
Machinery	96000	95000	94000	93000	92000		85000
1161 Machines	96000	95000	94000	93000	92000		85000
…							
Share capital	50000	50000	50000	50000	50000		50000
2001 Share capital	50000	50000	50000	50000	50000		50000
Profit from previous fiscal periods	20450	95450	95450	95450	95450		95450
Profit (loss) of the period	-15950	4500	7660	12020	16280		19080
SUM OF EQUITY	104500	199950	203110	207470	211730		214530
Loans from financial institutions (Long term)	102500	100000	97500	95000	92500		75000
2600 Bank loan	102500	100000	97500	95000	92500		75000
…							
Loans from financial institutions (Short term)	26000	25000	24000	23000	22000		15000
2800 Bank loan	26000	25000	24000	23000	22000		15000
…							
Sales accounts	55000	36700	36700	36700	36700		36700
3000 Sales	55000	36700	36700	36700	36700		36700
Materials and services	-12000	-15000	-15000	-15000	-15000		-15000
4000 Purchases	-12000	-15000	-15000	-15000	-15000		-15000
…							
9900 Prepaid income taxes	-200	-200	-200	-200	-200		-200
PROFIT (LOSS) OF THE PERIOD (PROFIT AND LOSS STATEMENT)	1100	4500	-1340	1200	-100		-100

3.4 Results

The three case company analysts were each given three csv files containing monthly accounting information in the Fennoa format described before. The three analysts performed the wrangling task in their respective ways. After they deemed the task done, they returned spreadsheet files they viewed as sufficiently wrangled through either Slack or email to be assessed. There was great variance in the time that the participants took to carry out the task.

The fastest performance from Participant 1 took ten minutes, and the other two participants required a longer time. Participant 3 was the only one who had any experience in using data exported from Fennoa.

Table 3. the performance data clocked by each manual wrangling participant through Toggl.

Participant	Duration (h:mm:ss)	Duration (seconds)	Placement
Participant 1	0:10:10	610	1
Participant 2	3:40:00	13200	3
Participant 3	2:55:42	10542	2

The participants were instructed to wrangle the data into the desired format as they saw fit to match the target format, which did not specify how many files each participant would have to produce. The fastest performance was carried out by the intern, who had the least experience in wrangling accounting data for Consulting Firm A's platform compatibility. Participant 2, however, chose a similar approach and took considerably more time

Participant 1 chose a time saving angle in wrangling the data. Instead of going through extra steps to produce a single data file out of the three exports, he chose to wrangle the three datasets separately with the same processes in order to produce three separate datasets that could be imported into the case firm system. This led to him saving a considerable amount of time in the wrangling, as no step in combining the datasets was taken. The two other participants chose to make a singular table out of the three exports, causing them to use more time and effort. This points to the conclusion that in the context of the case firm's cash flow analysis platform, choosing to combine multiple non-wrangled datasets is not a relatively efficient way to wrangle customer data when multiple datasets are given to analysts.

Table 4. Participant 1's performance in wrangling the data into the desired format.

Participant 1	True/False	This step is critical for successful data import
The produced file/s are returned in the csv format	TRUE	TRUE
Balance sheet values represent monthly balances	TRUE	TRUE
The last column representing the sum of P&L statemet rows is deleted	TRUE	TRUE
Month names are expressed in the m/yyyy format	TRUE	TRUE
The balance sheet 'profit of the period' value is not overwritten	FALSE	TRUE
The balance sheet individual account value names are unique	TRUE	TRUE*
The balance sheet summary value names are unique	FALSE	TRUE*
The profit and loss statement summary values are unique	FALSE	TRUE*
The profit and loss statement account value names are unique	TRUE	TRUE*
The balance sheet summary values are unique	TRUE	TRUE*
Is the cash flow simulation successful?	FALSE	FALSE
** the cash flow simulation is successful if either the summary value or account value names are all unique*		

Participant 1 was successful in creating three datasets that were accepted by the import process of the case firm's platform, as each of the three files each contained a dataset with the correct basic structure of account names sorted by row, and columns named by month. The balance sheet values were correctly identified as monthly changes instead of cumulative balances in contrast to profit and loss statement account values and changed accordingly. The data, however, did not contain the correct information to produce a successful cash flow simulation, as a crucial account was not named in a unique manner. As the value name "profit (loss) of the period" was used in two instances in each dataset, in both the profit and loss statement and the balance sheet - with the instances referring to a different value, a crucial figure was lost due to the system overwriting the balance sheet equity account value. This would lead to a failed cash flow simulation, due to equity ending up being miscalculated as the "profit (loss) of the period" not being registering a correct value.

Participant 1, therefore, failed to conduct a successful data wrangling despite having recognized the most efficient approach out of the three participants. Even though their approach served the most time saving way of transforming the numerical values of each exported dataset, they failed to account for the uniqueness of account names for crucial accounts demanded by the cash flow simulation platform, leading to a critical value not being imported into the case firm's system correctly.

Like Participant 1, Participant 2 chose to produce separate files for each Fennoa export. Unlike Participant 1, however, Participant 2 took a significantly longer time to finish his task. Later questioning revealed that the participant had attempted to produce a single file at first by combining the data into a single file so that they could more easily browse the correctness of the accounting data, before deciding that separate files were a more efficient solution. As with Participant 1, this points to the conclusion that creating a separate CSV file for each given export is a more efficient angle in this task.

Table 5 Participant 2's performance in wrangling the data into the desired format.

Participant 2	**True/False**	**This step is critical for successful data import**
The produced file/s are returned in the csv format	TRUE	TRUE
Balance sheet values represent monthly balances	TRUE	TRUE
The last column representing the sum of P&L statemet rows is deleted	TRUE	TRUE
Month names are expressed in the m/yyyy format	TRUE	TRUE
The balance sheet 'profit of the period' value is not overwritten	FALSE	TRUE
The balance sheet individual account value names are unique	TRUE	TRUE*
The balance sheet summary value names are unique	FALSE	TRUE*
The profit and loss statement summary values are unique	FALSE	TRUE*
The profit and loss statement account value names are unique	TRUE	TRUE*
The balance sheet summary values are unique	TRUE	TRUE*
Is the cash flow simulation successful?	FALSE	FALSE
** the cash flow simulation is successful if either the summary value or account value names are all unique*		

Apart from the significant difference in the time taken to finish the wrangling task, Participant 2's performance was near identical to Participant 1's. He had successfully transformed the balance sheet data into the monthly cumulative value format and ensured that the month was expressed in the correct form. While the balance sheet summary value names were not unique, each account name within the balance sheet was unique. He had also recognized that the profit and loss statement monthly values did not need changing. Just like Participant 1 however, Participant 2 had not picked up on the value name "Profit (loss) of the period" occurring twice in the exported datasets. This led to the balance sheet's cumulative equity value being overwritten by the profit and loss statement's similarly named value. Due to this, the balance sheet could not have been successfully imported into the case company's cash flow simulation system.

The results of Participant 2's efforts suggest that a manual approach predisposes the wrangling process to inefficiency due to the analysts not initially recognizing the most efficient approach in transforming raw data into the desired form, on top of them not recognizing the necessary steps to wrangle the data in order to produce a successful cash flow simulation in the case firm's platform.

Participant 3 had taken a more meticulous approach in relation to the other participants. While both Participant 1 and 2 had separated the exports into separate csv files, Participant

3 had chosen to produce an Excel file (.xlsx) with separate tabs representing the balance sheets and profit and loss statements for each year detailed in the Fennoa exports. While this allowed him to produce a separate spreadsheet for each yearly profit and loss statement and balance sheet in a single file, this was not the file format required by the case company's cash flow simulation system, as the data was expected to be delivered in the csv format. This means Participant 3 failed to produce files that were accepted by the case company's platform, and that his file would needed to have been converted into separate csv files in order to have been successfully imported into the case company's cash flow simulation system, even if all other steps had been carried out successfully.

Table 6 Participant 3's performance in wrangling the data into the desired format.

Participant 3	**True/False**	**This step is critical for successful data import**
The produced file/s are returned in the csv format	FALSE	TRUE
Balance sheet values represent monthly balances	TRUE	TRUE
The last column representing the sum of P&L statemet rows is deleted	TRUE	TRUE
Month names are expressed in the m/yyyy format	TRUE	TRUE
The balance sheet 'profit of the period' value is not overwritten	TRUE	TRUE
The balance sheet individual account value names are unique	TRUE	TRUE*
The balance sheet summary value names are unique	FALSE	TRUE*
The profit and loss statement summary values are unique	FALSE	TRUE*
The profit and loss statement account value names are unique	TRUE	TRUE*
The balance sheet summary values are unique	TRUE	TRUE*
Is the cash flow simulation successful?	FALSE	FALSE
** the cash flow simulation is successful if either the summary value or account value names are all unique*		

Later questioning revealed that Participant 3 had misunderstood the study objective, and that he had intended to produce separate spreadsheet tabs for each year's profit and loss statement and for each balance sheet, and had joined each spreadheet into the Excel spreadsheet file he had returned. For the purposes of this study, we can treat each separate tab as a separate csv file for the rest of the analysis to analyze how Participant 3 had carried out the other wrangling objectives.

Participant 3 had successfully recognized the balance sheet format mismatch and transformed the monthly values into cumulative values for each yearly balance sheet. Like the other participants, he had not recognized the "profit (loss) of the period" value names in the balance sheet as having the same title as the profit and loss statement value and failed to rename them in order to produce unique variables. Due to his method however, if the separate spreadsheets he had produced were imported in a particular order into the case company's cash flow simulation platform, his files could have been imported in a way that would have yielded a successful cash flow simulation. If the separate tabs on the Excel spreadsheet had represented individual csv files, the files could have been imported in a particulate order so that the overwriting of the equity "profit (loss) of period" values were not overwritten. If the profit and loss statement files were imported before the balance sheet files, the balance sheet 'profit (loss) of period' values would have been imported into the system successfully. This approach would have yielded an export effort that produced sufficiently unique values in order to build a complete model of the customer firm's cash flows utilizing individual account variable names.

Because the returned file format was incorrect however, Participant 3's effort cannot be interpreted as a fully correct approach. His choice to combine each profit and loss statement and balance sheet into separate tabs in an Excel file likely contributes to his performance time, which was the second highest. This points to the conclusion that failing to recognize the correct format and the effort to combine multiple files into a single file contributes to lower efficiency. He was however, the closest to producing a successful wrangling, as simply saving each separate spreadsheet tab as their own CSV file and importing them into the cash flow system in the correct order would have resulted in a successful cash flow simulation, while the other two participant's files could not have been imported successfully in any order.

3.5 Python code performance

The code written to carry out the task that the participants attempted was able to produce a single csv file that satisfied all the prerequisites. The balance sheet values were successfully transformed into monthly balance values into the cumulative format while keeping the profit and loss statement values in the monthly format. All summary value names were given a unique name in order to account for the case firm's system overwriting the numerical values of variables sharing the same name, and the summary value column was successfully deleted. The profit and loss variable names in both the balance sheet and the profit and loss statement were renamed with a suffix that indicated which part of the financial statement the value belonged to.

The code successfully carried out the prerequisite wrangling tasks while creating a single csv file, meaning that the data could be imported with a single file import instead of multiple file imports. This removed the need to import the data in multiple files, which was shown to be the most efficient way when wrangling customer accounting data manually. This points to the conclusion that an automated wrangling code allows analysts to skip several manual steps even if the analysts would had chosen the most efficient way to approach wrangling manually.

The code was run by three analysts working for the case firm. The analysts who ran the Python code were separate people than the ones chosen for the manual wrangling task. Each of the analysts who ran the code were able to produce a sufficient csv file in less than a minute. This points towards the conclusion that a sufficiently written automated Python script can handle wrangling tasks with a significantly greater speed and success rate than an analyst that is inexperienced with Fennoa data.

Table 7. The time taken by the second set of analysts in running the wrangling code

Participant	Duration (h:mm:ss)	Duration (seconds)	Placement
Participant 4	0:02:12	132	3
Participant 5	0:01:27	97	1
Participant 6	0:01:51	111	2

3.6 Comparison

The analysts' manual performance in producing datasets that satisfied the case company's cash flow system's requirements was lacking in several aspects when compared to the wrangling code. The analysts required significantly more time in order to perform the task, and each of them failed to account for all of the necessary cleanup steps in order to produce data that would have yielded a successful cash flow simulation without any additional intervention. A more lenient interpretation of the results with Participant 3's data being treated as separate CSV files would have yielded one successful simulation out of three, but only if the datasets had been imported in an order that would have resulted in the overwriting of redundant values. Even with this more lenient way of looking at the results, each of the analysts' datasets would have required at least three separate imports into the case company's cash flow simulation system, since even the most time saving approaches the analysts took produced three separate data files in comparison to the single file the wrangling code produced.

Table 8. Manual wrangling performance times in comparison to the Python wrangling code

Performance time, Manual			
Participant	**Manual or automated**	**Duration (seconds)**	
Participant 1	Manual	610,00	
Participant 2	Manual	13200,00	
Participant 3	Manual	10542,00	
Participant 4	Automated	132,00	
Participant 5	Automated	97,00	
Participant 6	Automated	111,00	
Difference	**Manual performance time (sec)**	**Automated performance time**	**Difference in %**
Min	610,00	97,00	529 %
Max	13200,00	132,00	9900 %
Mean	8117,33	113,33	7062 %

Table 9. Manual wrangling errors in comparison to Python code performance

Precision				
Manual	**No. of critical errors**	**No. of non critical errors**	**Total errors**	**No. of data exports needed**
Participant 1	1	1	2	3
Participant 2	1	1	2	1
Participant 3	1	1	2	6
Total	**3**	**3**	**6**	**5**
Python code	**No. of critical errors**	**No. of non critical errors**	**Total errors**	**No. of data exports needed**
Participant 4	0	0	0	1
Participant 5	0	0	0	1
Participant 6	0	0	0	1
Total	**0**	**0**	**0**	**3**

The results of this comparison are heavily skewed towards the conclusion that an automated wrangling code will outmatch manual work in efficiency in several aspects, and greatly increase efficiency of this wrangling process. The time taken by the code is several times smaller than the fastest analyst in carrying out the wrangling task, with the fastest manual performance time taking more than six times more seconds in carrying out the task. The code also provides several key benefits on top of the time taken by manual wrangling. First, the code can perform the wrangling tasks at a significantly lower error rate. Each of the analysts failed to account for all necessary steps in order to produce datasets that would have been fully compatible, as failures to account for account name uniqueness and correct file formatting were not an issue with the code runs. Moreover, the code also accounts for summary value names' uniqueness, which gives the analysts an option to run the cash flow simulation by putting the summary values into the cash flow simulation bins instead of individual account names.

The code also gives a significant advantage that cannot be accounted for in analyst efficiency. Due to the Python code producing a single file, an analyst working for Consulting Firm A would be left with only one file instead of the minimum of three files produced by the most efficient way the analysts chose to employ. Due to this, even with the analysts' wrangling manually producing three files without critical flaws, their best effort would have still resulted in them needing to import three files separately instead of a single file. This means that even at the maximum observed efficiency with all critical errors ignored, the wrangling code would have reduced manual work by reducing the amount of steps the analysts would have needed for importing the data into Consulting Firm A's cash flow simulation system.

3.7 Relevancy against academic literature

The results of the comparison concur with the overall consensus of the literature review. Automated RPA codes and bots have been observed to increase efficiency of menial and repetitive work in several contexts and give time for professionals for more thought requiring work to such a degree, that it may even transform their roles within their respective organizations. While the effects onto professional roles are not included within the scope of this thesis, the implications of the results point towards the conclusion that the case company would greatly benefit from the use of RPA in reducing the time and errors associated with data wrangling, especially when they wish to employ analysts who are not familiar with a customer's online accounting software.

The literature review indicated that RPA holds the most potential for menial tasks in cases where full cross-system integration was not financially viable, and where the individual steps of a task followed an easily predictable pattern that could be broken down to component tasks (Asatiani & Penttinen, 2016.). This pointed towards testing the wrangling code for accounting data from a software that offers output files in a standard form, and which was not commonly used enough by the case firm's customers to justify investments into integration. Fennoa's accounting data fit into this box for this case study, as it has not been used by many of the case company's customers, and necessities a considerable amount of repetitive wrangling tasks to become compatible for their processes. Another indicator for RPA as a viable solution was that each of the participants had little experience with Fennoa, which contributed to the amount of errors that would have yielded a failed cash flow simulation on initial try, as an automated code would not have – saving an inexperienced analyst a substantial amount of time due to them not having to reassess their wrangling efforts. As financial statement data from the software can be expected to stay in a particular format, this lends wrangling codes as a viable solution for software robots that handle data cleaning and cross system wrangling tasks.

The literature review revealed that implementing RPA can lead to significant errors in cases where a part of the automated task requires subjective decision-making from experts who had to use their personal judgement honed by experience in their working life. The wrangling task chosen for this thesis required little subjectivity or experience in order to be carried out successfully, since the requirements for the data format in the case firm's cash flow simulation system are static and non-changing. While experience with wrangling Fennoa data may contribute to wrangling the csv files into a compatible format more successfully and efficiently, it was not a prerequisite for carrying out the task observed in this thesis, as the steps required for the wrangling were based on the case firm's system requirements, and therefore not subject to experience or subjectivity. Moreover, each of the analysts chosen to carry out this task had very little to no experience in handling Fennoa data, and there was still great variance in performance time.

4 Conclusions and discussion

The aim of this thesis was to observe and measure the potential effects of RPA onto a menial accounting data wrangling task carried out by cash flow analysts in a consultancy firm, and to measure the changes in performance time and precision.

1. What does the academic literature suggest on the benefits of RPA onto cash flow analyst work efficiency?

The implications of this thesis largely correspond with the current literature on RPA's effect on the efficiency on the time and quality of the work of professionals working in the accounting and finance sector. While the effect on the changes on work roles was not within the scope of this case study, the results correspond with already existing research on the fact that RPA greatly reduces the time taken by finance professionals in carrying out menial work, and that it greatly reduces the number of errors that occur in manual data wrangling. The results of this case study also affirm that RPA is largely suitable for processes that do not justify large investments in cross system integration.

2. How can RPA be utilized in the context of this case study and does it increase the efficiency and precision of the analysts?

While full scale RPA adoption was not used in the context of this thesis work, a simple Python code was able to enhance the efficiency of a single wrangling process in a manner that would show a large increase in the work efficiency of financial professionals who dealt with raw accounting data. Analysts were able to cut their time used in wrangling significantly, reduce the amount of errors was reduced to zero, and the time the analysts would need to use in uploading data was reduced by producing a single dataset instead of multiple sets that would be needed to be uploaded seperately. This shows that RPA automation has great potential in reducing the hours professionals have to sacrifice in carrying out menial wrangling tasks related to data wrangling.

The case study had several limitations. As the number of analysts who carried out the automated task was limited, their performance time is not generalizable. This however is unlikely to affect the significance in the difference in performance time between manual labour and the automated wrangling script, as the difference in time was quite large. The study's setting was also limited to a singular and a contextual wrangling task that does not lend itself to large scale conclusions.

For further case studies with a similar approach, I would suggest larger observation settings with more analysts and finance professionals conducting the same task manually in comparison to an automated wrangling code. I would also suggest a longer timeframe for further studies into RPA adoption's effect into professionals' roles, in order to look into the thresholds of the scale of RPA adoption that leads into changes in worker roles within organizations.

While RPA shows great potential for increasing efficiency in menial tasks, adopting RPA requires both time and investment. An interesting topic to consider for future research would be to look for factors that affect the financial feasibility of RPA adoption projects, and the level of investment that results in successful RPA adoption projects.

5 References

Asatiani & Penttinen. 2016. Turning robotic process automation into commercial success – Case OpusCapita. Journal of Information Technology Teaching Cases 6, 2, 67-74.

Brealey, R., Myers, S. Allen, F. 2022. Principles of Corporate Finance. McGraw Hill LLC, New York. 217-224.

Cooper, Holderness, Sorensen & Wood. 2019. Robotic Process Automation in Public Accounting. Accounting Horizons, 33, 4, 15-35.

Cooper, Holderness, Sorensen & Wood. 2022. Perceptions of Robotic Process Automation in Big 4 Public Accounting Firms: Do Firm Leaders and Lower-Level Employees Agree? Journal of Emerging Technologies in Accounting (2022) 19, 1, 33–51.

Ernst & Young, 2019. Driving impact at scale from automation and AI. pp. 11.

Eulerich, Pawlowski, Waddoups & Wood. 2020. A Framework for Using Robotic Process Automation for Audit Tasks. Contemporary Accounting Researchm 39, 1, 691-720.

Fernandez & Aman. 2018. Impacts of Robotic Process Automation on Global Accounting Services. International Journal of Innovative Research and Advanced Studies (IJIRAS), 10, 7, 81-88.

Fridson,M & Alvarez. F. 2011. Financial Statement Analysis: A Practitioner's Guide, Fourth Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

Harrast, S. Robotic process automation in accounting systems. Corporate Accounting & Finance, 31, 4, pp. 209-213.

Hsiung & Wang 2022. Research on the Introduction of a Robotic Process Automation (RPA) System in Small Accounting Firms in Taiwan. Economies, 10, 8, 200.

Jędrzejka 2019. Robotic process automation and its impact on accounting.
Zeszyty Teoretyczne Rachunkowości, 105, 161, 137-166.

Kaya, Turkyilmaz & Birol 2019. Impact of RPA Technologies on Accounting Systems. Muhasebe ve Finansman Dergisi. 82, 235-250.

Kedziora & Kiviranta 2018. Digital Business Value Creation with Robotic Process Automation (rpa) in Northern and Central Europe. Management, 12, 2, 161-274.

Kokina & Blanchette 2019. Early evidence of digital labor in accounting: Innovation with Robotic Process Automation. International Journal of Accounting Information Systems, 35.

Kokina, Gilleran, Blanchette & Stoddard 2020. Accountant as Digital Innovator: Roles and Competencies in the Age of Automation. Accounting Horizons, 35, 1, 153-184.

Korhonen, Selos, Laine & Suomala 2021. Exploring the programmability of management accounting work for increasing automation: an interventionist case study. Accounting Auditing & Accountability Journal, 34, 2, 223-280.

Lacity, M., & Willcocks, L. 2021. Becoming Strategic with Intelligent Automation.

van der Aalst, W.M.P., Bichler, M. & Heinzl, A. 2018. Robotic Process Automation. Business & Information Systems Engineering, 60, 269-272.

Yunus, Aman & Keliwon 2019. The Role of Business Leaders in Information Technology Innovation in the New Era of Disruptive Technology. Asian Journal of Accounting and Governance, 12, 1, 1-10.

Zhang 2019. Intelligent Process Automation in Audit. Journal of Emerging Technologies in Accounting, 16, 2, 69-88.

Appendices

Appendix 1. Python code

```
import csv
#ASETA ALLE TILIKAUDEN ALKAMISKUUKAUSI
tilikauden_alku=1
if tilikauden_alku==1:
   tk=12
else:   tk=tilikauden_alku-1

import numpy as np
import glob
import pandas as pd
import os as os
import openpyxl
import datetime as datetime
import itertools
#Poimitaan kansiosta kaikki csv-tiedostot. Tiedostot oltava aikajärjestyksessä, aikaisimmasta aloittaen
location= #
csv_files=glob.glob(location)
start = {'Tili': [0]}
df1=pd.DataFrame(data=start, dtype=object) #Tyhjä datataulukko "Tili"-nimisellä sarakkeella, johon lisätään loopissa muokattuja csv-tiedostoja merge-funktiolla
#luettelo tiliryhmistä, jotka toimivat tarkentavina lisätermeinä. Indikoivat ne tiliryhmät joissa esiintyy tili, joka on samalla nimellä muualla, ja onko tiliryhmä tuloksessa vai taseessa.
# Mikäli havaitaan uusi tiliryhmä, on se lisättävä luetteloon
#luettelopath = *aseta tiliryhmäluettelon osoite tähän
luettelo=pd.read_csv(luettelopath, sep=';')

#CSV:t yhdistävä for loop
for csv_file in csv_files:
   test1=pd.read_csv(csv_file ,encoding='ISO-8859-1',sep=";", engine='python',skiprows=2, parse_dates=True, decimal=",") #lukee csv:n
   test1["Tili"].iloc[-1]=test1["Tili"].iloc[-1] + ' (Tuloslaskelma)' #viimeinen termi on aina tulos tilikauden aikana, joka esiintyy myös taseessa, mutta tarkoittaa eri arvoa
   #Fennoan taulukoissa 14. sarake on aina kooste, poistettava
   test1=test1.drop(test1.columns[13], axis=1)
   #Fennoan tilinimissä välilyöntejä alussa, poistettava merge-funktiota varten
   test1['Tili']=test1['Tili'].str.strip()
   #luodaan sarake, jossa ilmestyy aina erottava termi luettelosta
   test1=test1.merge(luettelo, how="left")
   #Täytetään sarakkeen solut edeltävällä arvolla
   test1['Tilit']=test1['Tilit'].fillna(method='ffill')
   test1['Tarkennettu tili']=test1['Tili'] + '§'+test1['Tilit']
   test1['Tili']=test1['Tarkennettu tili']
   test1=test1.drop(['Tarkennettu tili','Tilit'], axis=1)
```

```
    df1=df1.merge(test1, how='outer').fillna(0) #jatkaa kaikki siivotut csv:t yhteen

df1=df1.groupby(['Tili'], sort=False).sum() #summaa kaikki sarakkeet Tilin ainutkertaisten arvojen mukaan

df1=df1.loc[(df1.sum(axis=1) != 0)] # Poistaa rivit, joiden summa on 0

df1.index.name = 'Tilit'
df1.reset_index(inplace=True)
df1[['Tilit','Sijainti']] = df1['Tilit'].str.split('§',expand=True)
df1['Esiintyvyys'] = df1.groupby('Tilit')['Tilit'].transform('count')

df1.loc[ df1['Esiintyvyys'] > 1, 'Esiintyvyys'] = df1['Sijainti']
df1['Esiintyvyys'].replace(['1', 1], '', inplace=True)
df1['Tilit']= df1['Tilit'] + " ¤" + df1['Esiintyvyys'] + "&"
df1['Tilit'] = df1['Tilit'].str.replace(' ¤&', '').str.replace('¤','(').str.replace('&',')')

df1.rename(columns={'Sijainti':'Tili'}, inplace=True)
df1 = pd.merge(df1, luettelo, on="Tili", how='inner')
df1.drop(['Esiintyvyys','Tili', 'Tilit_y'], axis=1, inplace=True)
df1=df1.set_index('Tilit_x')
kuukaudet=([col for col in df1])

kuukaudet=[i.split(' - ', 1)[0] for i in kuukaudet]
del(kuukaudet[-1])
kuukaudet = [e[2:] for e in kuukaudet]
kuukaudet = [e.replace(".","/") for e in kuukaudet]
kuuindeksit=[i.split('/', 1)[0] for i in kuukaudet]
idxs = [i for i, v in enumerate(kuuindeksit, 1) if v == str(tk)]   # calculating indices
result = [kuukaudet[i:j] for i, j in zip([0]+idxs, idxs)]
kuukaudet.append('Sijainti')

df1.columns=kuukaudet
dftulos=df1[df1['Sijainti'] == 'Tuloslaskelma']
dftase=df1[df1['Sijainti'] == 'Tase']

#Sovelletaan cumsum-toiminto taseen tileille, Result-lista määrittelee sarakkeet joihin cumsum-funktiota sovelletaan.
#
for i in len(result):
    dftase.loc[:, result[i-1]] = dftase.loc[:, result[i-1]].cumsum(axis=1)


df1=pd.concat([dftase, dftulos], axis=0)

df1=df1.drop(columns=['Sijainti'])
df1.to_csv('test10.csv', sep = ';', decimal = ',',encoding='ISO-8859-1')
```

Appendix 2. Instructions for participants

Kiitos osallistumisestanne.

Tehtävänänne on muokata liitteenä mukana olevat tiedostot yhteensopiviksi kassavirtajärjestelmään. Luokaa tiedosto/tiedostot jotka voi ladata järjestelmään sellaisenaan niin, että työtä voi jatkaa onnistuneen kassavirta-analyysi n luomista varten. Kukin tiedosto sisältää yhdessä taulukossa sekä tuloksen että taseen.

Toimikaa kuten kenen tahansa uuden asiakkaan onboarding-prosessin kanssa. Voitte muokata tiedostoja millä vain keinoilla, jotka näette sopivaksi, ja etsiä myös ohjeita internetistä. Älkää kuitenkaan kommunikoiko muiden osallistujien kanssa.

Tavoitteena on saada data muotoon, jossa sen voi ladata sisään järjestelmään ilman lisämuokkauksia kassavirtamallinnoksen onnistunutta simulointia varten.

Kun olette muokanneet tiedostot yhteensopivaan muotoon jossa sen voi ajaa järjestelmään, palauttakaa tiedostot Oskarille. Ajoittakaa suorituksenne Togglissa.