

Data-driven insights to Robotic Process Automation with Process Mining

Many organizations are making efforts in automating their mainly manual processes. However, there's currently a large amount of guess work and subjectivity involved in assessing the processes that might qualify for automation and keeping track of improvements. Process Mining can be a powerful ally in bringing data-driven insights to support and substantiate process owners' and developers' decisions as well as quantify enhancements brought about by automation.



Ir. Carolina Di Bisceglie
is a consultant at KPMG in the Data & Analytics team.
dibisceglie.carolina@kpmg.nl



Himalini Aklecha MSc
is a consultant at KPMG in the Data & Analytics team.
aklecha.himalini@kpmg.nl



Elham Ramezani PhD
is a manager at KPMG in the Data & Analytics team.
ramezani.elham@kpmg.nl

In every organization, process execution data is constantly been logged in different source systems

INTRODUCTION

Robotics Process Automation (RPA) has become an interesting topic within organizations, as it provides a quick and efficient method to implement and execute processes. There are many enterprise automation tools available for organizations. RPA uses software-based robots that are at the top of the IT infrastructure to perform high volume tasks without changing the existing architecture and allowing for agile implementation of an RPA project.

Organizations are moving towards standard back office processes to cut costs and improve efficiency. RPA is frequently implemented in such cases, with the idea that high volume and repetitive tasks can be automated. This leads to an increase of first-time right (limiting human error), a decrease in labor costs, and freeing resources to focus on more value adding activities where human creativity can bring competitive advantage for a business. However, several RPA projects fail to stay within budget, and time and return of investment is usually not delivered as expected. This is often caused due to false notion of process complexity and lack of transparency on how processes are being executed ([Kirc17]).

Process mining technology offers a set of novel tools and techniques for factual driven analysis of business processes. This technology uses the abundance of event data to provide an end-to-end and transparent view of processes. This paper explains how process mining can be leveraged to accelerate and improve the quality of RPA projects and measure its results.

In this article we will first introduce process mining along with its most common techniques. This is followed by an introduction to RPA and the different stages of a typical RPA project. We will then dive into how process mining can be applied for a successful RPA implementation. This is also further contextualized with the help of

a running example. Lastly, we will conclude this article with how RPA projects can benefit from process mining techniques.

Setting the Scene

In every organization, process execution data is constantly been logged in different source systems. This data, also known as event logs, contains information about the events that are being executed for each instance of a process. For example, a patient treatment process in a hospital may consist of the following events: the registration of the patient, first appointment, examination, diagnosis, preparation of the care plan, etc. Process mining learns a process by example from these event logs and provides insights into and transparency about how the business processes are being executed. One of the most common techniques of process mining is process discovery, where information from an event log is extracted to build a process model. These models represent the as-is process within the business. Using process mining, it is possible to detect different variations of a process or compare a process between regions, periods of time, suppliers, customers, etc.

Furthermore, process mining provides insights into the distribution of the different users active in a process (e.g. manual users versus, system users) and handover of work between them.

Other important process mining techniques are conformance checking and model enhancement. Conformance checking is done by mapping the extracted log against the discovered or hand-drawn process model. This mapping is used to detect and capture deviations that are caused due to the difference in the behavior of the logged data and the business process. Model enhancement is used to extend a process model with additional information extracted from the event log. For example, the additional information can be extracted from timestamp information (time perspective) or from data attributes that characterize the process (data perspective). This can be further used to repair and alter the process structure ([Aals16]). The presented set of techniques, when combined, can be leveraged to obtain valuable insights for RPA projects.

RPA is an umbrella term for tools that operate on the user interface of other computer systems in the way a human would ([Aals18]). In other words, technology is used to configure computer software robots, also named bots, to emulate human executions of business processes in digital systems. RPA bots use the user interface to capture data and manipulate applications in the same way humans do.

There is a range of processes that has been tried and tested, making them ideal options for RPA. Checking vendor invoices, handling routine insurance claims, or pro-

cessing loan applications are just a few examples where RPA has been used successfully. In general, all processes that are high-volume, business-rule-driven and repeatable are perfect candidates for RPA.

In order to understand how Process Mining can aid an RPA project, it's necessary to understand how such projects are carried out and the stages they typically go through. Figure 1 shows the four different stages of an RPA project lifecycle and is explained in detail as follows:

1. **Assess:** Before starting an RPA project, it's important to understand the existing processes that potentially qualify for automation and how they unfold within the company. Once the candidates are chosen, a series of interviews with the process owners follows, aiming to map the process, its steps, and decision points. This phase can be very lengthy due to conflicting accounts from various process owners, which then need to be aligned to form a clear picture. This occurs because organizations don't always have a clear overview of how a specific process should be carried out, let alone how it unfolds daily. After a process structure has been pieced together, the clear and defined activities and their sequence are turned into the logical basis for the bots.
2. **Program & Test:** The following stage in the RPA lifecycle is to turn the devised process logic into a script that will be followed by the configured bots. The program is tested and the process owner and RPA team can assess whether its purpose is being fulfilled. As expected, a few iterations are needed to ensure the process is performed flawlessly by the bots. It's worth mentioning that it's not always going to be possible to automate 100% of the cases, because they might contain exceptions ruled by more complex logic. The bots are tested in a controlled environment, preferably using synthetic cases.
3. **Mobilize & Implement:** Once the testing is complete, the bots can be deployed to start handling day-to-day occurrences of the newly automated process. The deployment format depends heavily on the client's

preferred approach. It can be gradually implemented across departments or by switching the procedure for the entire enterprise overnight. Regardless of the chosen methodology, employees need to be trained in the new process and which actions they need to perform within this process.

4. **Measure & Sustain:** As seen in Figure 1, the project doesn't end with the implementation of the bots. Even after extensive testing, the programming of the bots is not impervious to errors caused by sudden changes to the process (e.g. software updates). Consequently, it is crucial to routinely monitor the bots' performance in order to detect such problems and quickly update the program to accommodate the changes. Furthermore, continuous monitoring of the amount of cases handled by bots and how that relates to their maximum usage is key in accounting for the project gains and computing return of investment.

How can Process Mining help?

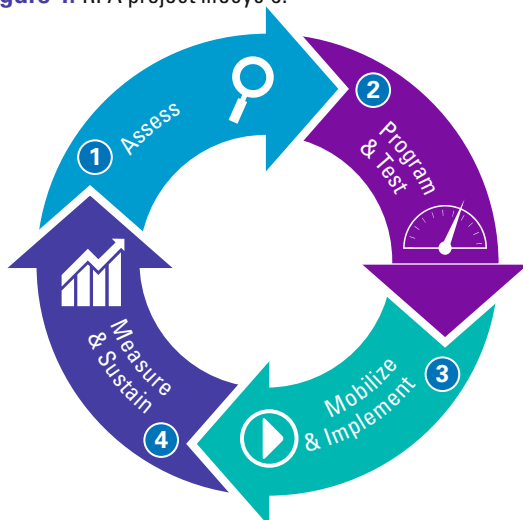
Process Mining can aid most phases of an RPA Project, generating valuable insights that reduce project time-frames and promote more informed decisions. In order to better showcase the added value, a running example with a Purchase to Pay process is given throughout this section. However, the methodology is process agnostic and therefore similar benefits can be achieved for other processes.

Process discovery removes the intrinsic subjectivity of the interviews when mapping the process and significantly shortens this lengthy step. By mining an event log instead, it's possible to get the most accurate representation of the as-is process as well as different variants that occur. This analysis is especially interesting when drilled down to relevant dimensions, because it allows to spot possible discrepancies. Once singled out, these can be either standardized or marked as exceptions to the main process.

The event log also provides information indicative of the current automation rate as well as the processing time of each activity. The current automation rate refers to the ratio between the number of activities performed by a system user and the total of activities performed. This can be calculated both for the whole process or by specific activity. The manual processing time of each activity refers to the time a manual user (employee) spends actively performing that task. This information paints a clearer picture about which parts of the process are in greater need for automation and would yield the highest returns: those with lower automation rates and higher manual hours. This preliminary analysis helps narrow down the activities that are worth inspecting further.

For the selected activities, an individual analysis can be made. It is possible to visualize all paths going in or out

Figure 1. RPA project lifecycle.



of each activity, giving more insight into their suitability for automation. The paths going in and out of each activity provide insight on the possibility of automating a sequence of activities instead of just one. If the process is straight-forward in a way that several activities don't have

Figure 2. Process model discovered for four variants of the example Purchase to Pay process with highest frequency.

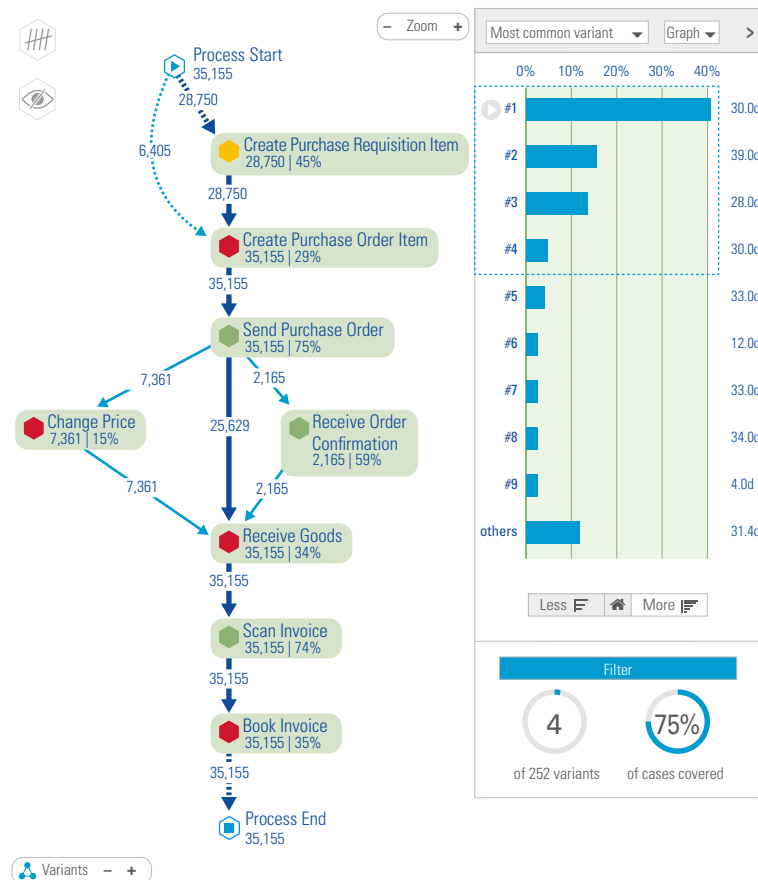
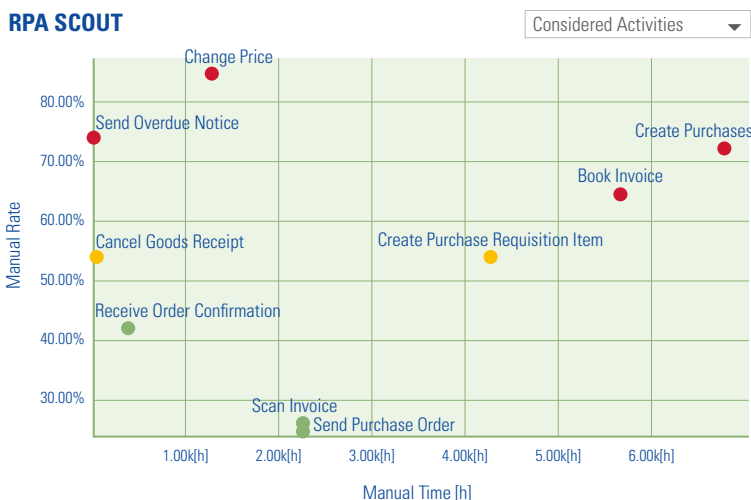


Figure 3. Scatterplot of the Manual Rate versus Manual Time for each activity in the process.



complex decision points and do not require human verifications, bots can be programmed to automate the entire batch of activities (or in some cases the whole process).

This analysis can be leveraged to create a business case for the automation of each activity or batch thereof. Consequently, it allows for quantitatively prioritizing which automations should be carried out and in which order they should occur so returns can be maximized. In sum, process mining is a powerful tool to accelerate project timeline and provide information for well-based, data-driven decisions in the assessment phase of RPA projects.

Running Example

Using process discovery, it's possible to immediately see the as-is process, with no room for subjectivity about the order in which the activities were performed as showcased in Figure 2. On the left side, the process is displayed with all activities and paths that occur in the selected variants on the right side (four variants with highest frequency are chosen). The numbers on the arrows in the process mark the number of cases within the current selection in which that connection occurs. The percentages below the label of each activity indicate its automation rate and the colors are associated with an arbitrary scale that marks red when the automation rate is lower than 45%, yellow between 45% and 55% and green above 55%. For a good balance between process variant representation and comprehensive visualization, we chose to display only the top 4 variants, which cover 75% of cases as seen in the lower right corner. It also shows an astounding 252 variants, most of which are unwanted.

Also noteworthy is the occurrence of change activities. The mined process flow shows that "Change Price" occurred more than 7.000 times just in the considered variants. Change activities are often a byproduct of human error and indicate process rework and consequently lengthier process throughput times. Automating the creation of the purchase order could help reduce the number of change activities significantly, and, therefore, the rework needed.

Figure 3 sorts activities based on the manual execution rate and the total time spent on manual execution of the activities. By having a closer look at the manual execution rate per activity and the total hours spent on them, it's possible to narrow down the number of automation candidates to 'Create Purchase Order Item' and

'Book Invoice'. As illustrated in Figure 3, activities 'Create Purchase Order Item' and 'Book Invoice' are executed often manually (70% and 65% respectively) and the total time spent on manual executions of the these activities are around 6.800 hours for 'Create Purchase Order Item' and 5.750 hours for 'Book Invoice'.

For these two activities, a more in-depth analysis of the paths going in and out of each was made. The following analysis will focus on 'Create Purchase Order Item'. As seen in the top table of Figure 4, activity 'Create Purchase Order Item' is executed 39.244 times after activity 'Create Purchase Requisition Item' which is in accordance with how the process should be carried out. Examining the activities following the creation of the purchase order item in the bottom table of the same figure, it's clear there are no complex decisions to be made: the next activity should be 'Send Purchase Order Item' (for 45.588 purchase order items, the sequence <Create purchase order item, Send purchase order> is observed). It also shows a significant portion of the created purchase order items (1.123) get refused. If the refusals are caused by human errors, automating the purchase order item creation could help bring down the refusal occurrence. However, if the refusals are governed by a more complex logic that would require human interference, it already indicates that automating a sequence of activities following the creation of the purchase order item might not be feasible.

Finally, a rough Business Case was put together for the 'create purchase order item' based on:

- number of manual executions;
- targeted automation rate;
- average processing time;
- full time employee (FTE) yearly hours;
- FTE annual salary average.

As seen in Figure 5, the last four criteria are customizable which allows for more accurate projections of FTE and monetary savings. Based on the Business Case, a decision was made to simulate the automation of 'Create Purchase Order Item' with a targeted automation rate of at least 80%. This number considers that it might not be possible to automate 100% of cases due to input from a different source or in a different format from those the bots are programmed to handle. Possible fluctuations in the automation rate due to external factors such as software updates are also considered.

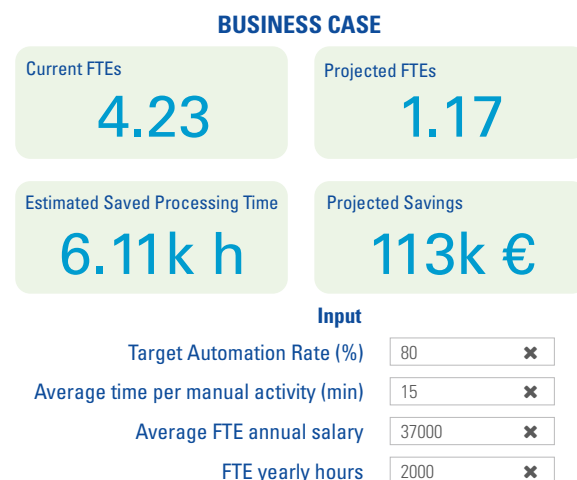
Using process discovery, it's possible to immediately see the as- is process, with no room for subjectivity

Figure 4. Tables showing activities preceding and succeeding 'Create Purchase Order Item'.

Coming From		
Coming From	Case Count	Processing Time
Cancel Goods Receipt	1	0.25 hours
Change Quantity	544	136.00 hours
Create Purchase Requisition Item	87,483	21,870.75 hours
Scan Invoice	863	215.75 hours

Going To		
Going To	Case Count	Processing Time
Cancel Goods Receipt	104	26.00 hours
Change Quantity	348	87.00 hours
Refuse Purchase Order Item	2,635	658.75 hours
Sent Purchase Order	92,407	23,101.75 hours

Figure 5. Process mining supported Business Case.



Bots can be used for calculating gains and computing return on investment

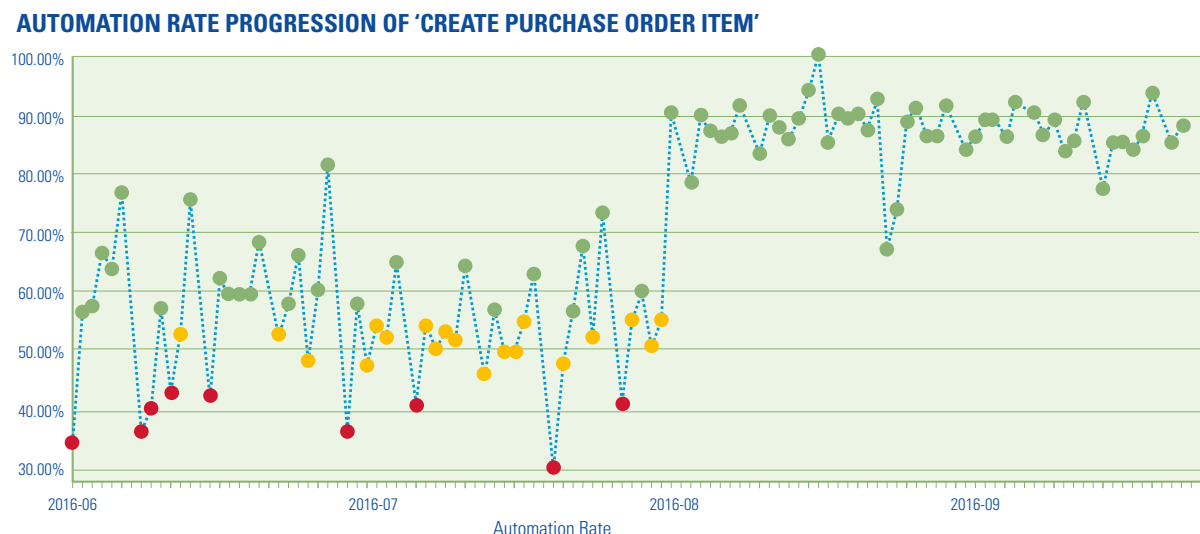
Moving forward in the project, process mining can be used to compare the bots to non-RPA supported executions of the process during testing. This gives a better overview of case coverage and process changes. The latter encompasses unexpected desirable and undesirable alterations caused by the automation. The positive ones can be incorporated whereas the negative ones can be used for improving the bot scripts in order to be avoided. However, bot-handled cases at this stage are limited by constraints of the chosen test method. As they cannot be trained for all scenarios, more insights will be gained once they start operating real cases and are confronted with situations that could not have

been anticipated. Those may be caused by different factors, for example software updates.

As multiple iterations are needed to perfect the final script the bot will run on, benchmarking the executions of each script to the other allows for comparing throughput time, case coverage and key process indicators (KPIs) on each iteration. Once the bots are ready to go live, it's possible to visualize the progression of the main KPIs and the process itself throughout it. Process mining facilitates managing the implementation progress with greater refinement and precision of the meaningful dimensions for each process and enterprise.

After the bots are fully operational, the process can be monitored live to guarantee the RPA benefits are consistently upheld and to immediately spot alterations in the KPIs that suggest a need for adjustments in the bot scripts. Additionally, end-to-end monitoring evinces unforeseen or previously unmeasurable benefits, such as sharp drops in the amount of rework by reducing human error. Furthermore, as event logs may also contain information regarding users (both humans and bots), it is possible to keep track of how many activities are being performed by each bot and the extent of the new processing time. Finally, since all these insights are backed by solid numbers extracted from the event log, they can be used for calculating gains and computing return on investment.

Figure 6. Automation rate progression throughout RPA project implementation.



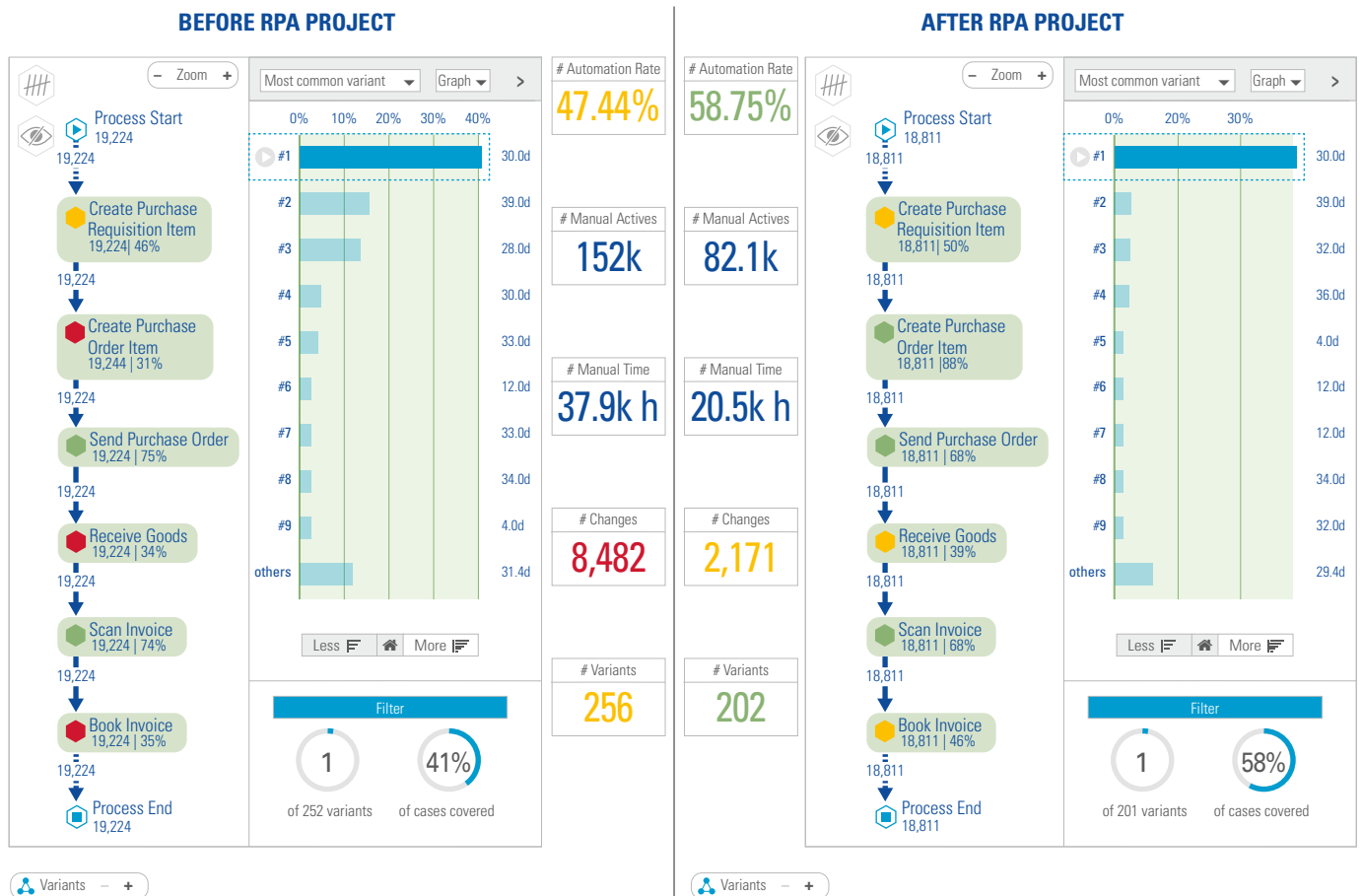


Figure 7. Process variants and KPI comparison before and after an RPA project.

Running Example

Using information extracted from the event log, it was possible to create a clear view of the progression of the automation rate throughout RPA implementation and afterwards. Figure 6 shows the simulated trend of the automation rate of activity “Create Purchase Order Item” during the 2-month implementation phase and over the following two months. We observe that during the implementation phase the automation rate fluctuates, as it would be when the bots would be implemented across departments, considering possible final fixes needed in the bot’s scripts. After implementation is completed, the automation rate of the activity increases and remains relatively stable.

Mining the event log also brings solid numbers regarding the status of KPIs that are otherwise hard to measure, such as the number of process variants and the manual hours spent on the process or the automated activity. This can be seen in Figure 7 where we compare the as-is process before the RPA implementation against the process discovered from the event log that was generated after the RPA implementation. The automation rate has

increased by 11,3%, total frequency of manual activities decreased by 70.000, and total (manual) processing time has decreased by 17.400 hours. Moreover, it sheds light into a by-product gain of the automation initiative: reduction of change activities throughout the rest of the process and the consequential increase in process standardization. This is evidenced in Figure 7 by a 74,4% decrease in the number of change activities and 21% decrease in the number of process variants. By bringing data-based numbers regarding the occurrence of change activities, it’s easy to quantify the improvements made in process efficiency.

Finally, the new number of manual executions combined with the input given before the automation effort allowed us to monitor the business case created before the RPA project and keep track of the financial gains obtained. For this running example, Figure 8 shows a decrease in the number of FTEs currently needed to perform this activity 4,23 (as shown in Figure 5) to 1,5, a drop of 64,5%. Similarly, the figure shows the saved processing time amounted to 5.460 hours and roughly estimated monetary savings achieved € 101k after the activity “Create Purchase Order Item” was automated.

Process mining techniques can provide a basis for assessing where business processes can be automated

BUSINESS CASE

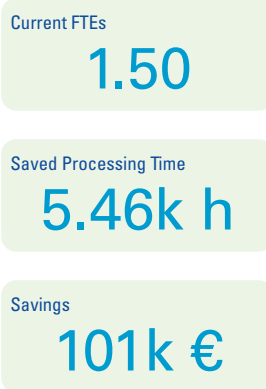


Figure 8. Updated process mining-supported Business Case.

CONCLUSION

In conclusion: process mining techniques can provide a basis for assessing where business processes can be automated. These techniques allow for data-driven decisions at different stages of an RPA project, eliminating guess work and reducing failures. By analyzing the as-is process for automation, a fact-based assessment can be done to analyze the process fragments that can benefit from automation. Furthermore, these techniques can be extended to analyze the automated process during the testing and implementation phase. Such data-driven decisions in combination with the continuous monitoring of RPA implementation leads to reduced costs and risks. This has been shown with the help of a running example on a purchase to pay process which has been implemented within the KPMG RPA Scout app in the Celonis platform.

References

- [Aals16] Aalst, van der, W. (2016). *Process mining: data science in action*. Heidelberg: Springer.
- [Aals18] Aalst, van der, W., Bichler, M., & Heinzl, A. (2018). *Robotic Process Automation*. Bus Inf Syst Eng.
- [Kirc17] Kirchmer, M. (2017). *Robotic process automation – pragmatic solution or dangerous illusion?* Retrieved from BPM-D: <https://bpm-d.com/bpm-d-exhibiting-at-btoes-2/>.

About the authors

Ir. Carolina Di Bisceglie is a consultant at KPMG in the Data & Analytics team. Carolina is a member of the Business Analytics unit at the Technology practice of KPMG. She co-developed the KPMG RPA Scout app, which directly applies the methodology described in this article.

Himalini Aklecha MSc is a consultant at KPMG in the Data & Analytics team. Himalini is a member of the Business Analytics unit at the Technology practice of KPMG. She is mainly involved in Process Mining activities within the group.

Elham Ramezani PhD did an innovative research in Process Mining from theory to application and developed methodology and techniques for a comprehensive data-driven analysis of business processes. Elham leveraged process mining to analyze and improve processes in different business sectors including the public sector, health care and banking.