

Has the Pandemic Impacted my Workforce's Productivity?

Applying Effort Mining to Identify Productivity Shifts during COVID-19 Lockdown

Wolf-Dietrich Zabka^[0000-0002-4949-6495], Peter Blank, and Rafael Accorsi^[0000-0001-5620-561X]

PricewaterhouseCoopers, Switzerland
Competence Center for Process Analytics and Mining
Birchstrasse 160, 8050 Zurich, Switzerland
{wolf-dietrich.zabka, peter.blank, rafael.accorsi}@pwc.ch

Abstract. As working from home became mandatory during the COVID-19 lockdown, many businesses feared negative productivity shifts originating from the new way of working based on the same processes and technology. Process Mining as a data-driven process discovery, analysis and monitoring method offers the continuous monitoring of process performance. In practice, however, severe limitations arise when assessing the time effort required for process execution as many enterprise resource planning systems do not record the time required for the execution of activities. In this context, the contributions of this paper are twofold: first, we present *Effort Mining* as a process- and system-agnostic method that allows us to estimate the time required for process steps; second, we apply this method to compare the effort before and during the lockdown as a means to identify productivity gaps induced by the pandemic. Our investigation has identified no substantial change in the employees' productivity, indicating that the underlying business processes and corresponding technology were fit for purpose in both scenarios.

Keywords: Process & People Analytics · Productivity Analysis · Pandemics

1 Introduction

Improving operations is a promising investment in times of stability. In times of crisis, however, the focus shifts to survival and maintaining the current level of operational excellence as well as liquidity become the main goal. This was also the case when COVID-19 hit the economy, working remotely from home became mandatory for many firms, and the environment of the people executing the processes changed. To assess process performance, Process Mining has emerged as a holistic and data-driven methodology in the last decade [1]. Its main purpose is to increase the operational excellence of companies through data-driven

process discovery, initiating continuous process improvement, and creating process transparency [1]. A substantial part of business process execution is based on human interaction [2], a component that could be critically affected by a lockdown and mandatory working from home. This change imposed the questions of how well employees would adapt to the remote working, how seamless process execution could continue and how productive employees would be in their new environment. In particular, the assessment of human productivity or performance imposes difficulties on Process Mining. When it comes to the judgement of human effort for process execution, there is a crucial limitation in the practical environment: in the most common enterprise resource planning (ERP) systems the process steps are atomic, meaning the process steps are recorded with a single timestamp reflecting the end of an activity. The absence of a start timestamp makes it difficult to judge the true overall work effort in terms of time passed. Even though the existing literature covers the estimation of waiting times from transition times [3], human-generated uncertainty of logged events [4] or the relevance of missing life-cycle data by utilizing of predictive models [5], to the best of our knowledge existing literature does not tackle this challenge [1, 6]. In consequence, we sought a process-agnostic method to deduct the overall work effort from process mining logs, which only have the end timestamp of an activity. Based on this, the contributions of this paper are twofold:

1. We introduce *Effort Mining* as a method to deduct the start timestamp based on user behaviour and measure the systemic work effort for process execution.
2. We demonstrate the use of Effort Mining in a practical/industry setting by applying it to monitor changes in human behaviour during a lockdown.

Overall, Effort Mining is a natural complement to process mining, allowing the deep-dive into the time spent per activity. This is an impactful extension when the focus shifts from end-to-end process performance to more of a “people” perspective and the relevant questions are centred around work time. As a result, we provide our clients with data-driven information about how the daily working rhythm was impacted by the lockdown and how processing times of activities changed, even if only the end timestamp of activities was recorded by the ERP system.

The remainder of this paper is structured as follows: Sec. 2 briefly describes the situation our clients faced during the lockdown, which motivated the development of Effort Mining. Sec. 3 describes the technical steps required in Effort Mining. This includes the minimum data requirements, the executed data transformation and the assumptions required to deduct a start timestamp. Sec. 4 reports on the results that can be achieved by using Effort Mining with a focus on the comparison of data recorded during the lockdown in April/May 2020 and data recorded in the previous year. The results originate from a client active in Financial Services, but as the method is process-agnostic, the results can be seen as indicative for any process or industry. Sec. 5 outlines the additional experience we have gained when applying Effort Mining, the feedback we have received until now, and the kind of applications we see for Effort Mining in the future.

2 Situation faced

On 11th of March 2020, the World Health Organisation announced in a press conference that COVID-19 was a pandemic and that immediate actions needed to be taken by governments [7]. In the following, many countries introduced lockdowns to decrease mobility and retain the spread of the virus. For companies in these countries, this meant that work should be conducted from home whenever possible, and remote working became the new standard operating mode. The move to working from home raised two primary concerns: first, about the productivity of employees since employees could not be monitored as closely as before, and second, concerns about employee wellbeing were raised as the boundary between home and workspace faded.

After the initial phase of COVID-19, a common question was if and how remote working impacted the productivity of employees. For several clients, our team had previously executed process mining projects and created the required data models based on data generated and stored in the clients' ERP systems. The process mining data models were extensive, well-developed, and validated in previous projects. Thus, it was regarded as a reliable and high-quality data source. Further, data visualisation with process mining dashboarding tools was already implemented. Previously, the main purpose of the process mining infrastructure had been to ensure compliance and to identify process improvement potentials, but now a new scope was added: the goal of these studies was to use the existing process mining data models to identify changes in employee performance and behaviour. To achieve this, we have created a process-agnostic approach that is based on process mining event logs and can be used to assess processing times for the execution of activities, gives insight into the daily rhythm in which activities in the ERP system are executed, and estimates the overall involvement of employees in terms of overall working hours.

3 Action taken

The starting point for these projects were existing process mining event logs as a data source and the existing process mining infrastructure. To assess productivity within the firm, we deduct the number of manual activities executed in time intervals. Another relevant productivity metric is the typical execution time of specific activities: this metric is not available directly as the ERP systems record only the end timestamp of an activity. However, we have developed a statistical procedure to estimate the start timestamp. Further, we estimate the duration of workdays based on the first and last timestamp submitted by users on each day.

3.1 Data Transformation from Event Log to Effort Log

The table structure of the process mining event log is depicted in Table 1, having the typical columns of the CaseID, the timestamp, and the occurring event. This is the typical minimum data required for process mining. Further, the

CaseID	Timestamp	Event	UserID
CaseNo1	2020-06-04 08:35:45	EventA	User1
CaseNo1	2020-06-04 08:36:45	EventB	User1
CaseNo2	2020-06-04 08:37:00	EventB	User1
CaseNo2	2020-06-04 09:05:30	EventA	User1
CaseNo3	2020-06-04 09:05:30	EventC	User1
CaseNo3	2020-06-04 09:05:30	EventD	User1
CaseNo3	2020-06-04 09:05:30	EventD	User1
CaseNo3	2020-06-04 10:00:15	EventD	User1
...
CaseNo9	2020-06-04 17:25:15	EventA	User1

Table 1: Structure of the process mining event log, with the typical columns CaseID, Timestamp and Event. Further, the UserID is required.

UserID of the user executing the respective process step is a piece of the required information for Effort Mining because it is needed to add information about a possible activity start.

Based on the hypothesis that users perform multiple activities in a day for the process in scope, we transform the process mining eventlog and create the effort log as shown in Table 2. This table has one entry for each manual activity performed in the respective process. Here, we define one activity as a distinct combination of the UserID and the recorded timestamp. One activity can affect several cases and can create several events in the event log (e.g. compare ActivityID 4 in Table 2 and entries at the timestamp ‘2020-06-04 09:05:30’ in Table 1). Further, we make an estimate for the processing time based on the previous activity executed by the same user on the same day. A direct consequence is that the processing time of the first activity of each user on each specific date is missing and will be estimated in a subsequent step.

3.2 Estimation of Start Timestamps of Activities

The effort log (Table 2) generated from the event log (Table 1) provides us with a set of possible activities being executed in the process. For each distinct activity, we obtain a set of observed processing times from which we construct a respective distribution. Figure 1 shows the distribution of observed processing times for three different activities. The mode of this distribution we account as the most likely or typical processing time of the respective activity inside the ERP system. If the observed processing time exceeds the typical processing time by far, the respective user likely executed activities outside of the ERP system. These activities outside of the ERP system could include preparations for the respective activity (e.g. making inquiries), executing activities in different processes, administrative tasks or taking a break. Even though a detailed decomposition of this effect is beyond the scope of this paper and might even not be possible without additional information, the observed processing time can serve as an

ActivityID	UserID	Timestamp	Activity	Observed Processing Time
1	User1	2020-06-04 08:35:45	EventA	<i>Null</i>
2	User1	2020-06-04 08:36:45	EventA	60 sec
3	User1	2020-06-04 08:37:00	EventB	15 sec
4	User1	2020-06-04 09:05:30	EventA,EventC, EventD	1710 sec
5	User1	2020-06-04 10:00:15	EventD	3285 sec
...
10	User1	2020-06-04 17:25:15	EventA	1000 sec

Table 2: Structure of the effort log: The ActivityID represents a distinct UserID-timestamp combination. The observed processing time is deducted based on the timestamp of the previous timestamp made by the same user.

indication for the overall working time required to execute the respective action. Further, we can assume that the mixture of the activities outside of the ERP systems remains consistent over time and that it is distributed homogeneously over the users and activities if the number of users and activities is large.

To estimate the start timestamp of a specific activity, we define a threshold in dependence on the typical time of the respective activity. If the threshold is not exceeded, we use the observed processing times to deduct the start timestamp of the respective activity. If the threshold is exceeded, we consider the typical processing time to deduct the start timestamp. Further, we use the typical processing time to deduct the start timestamp for the first activity of each day for each user, where no observed processing time is available (see Table 2). It is essential to acknowledge that this method is an approximation and does not deliver exact values for each activity. However, for commonly executed activities in large organizations sufficient data points are available to calculate the typical times accurately and deduct the start timestamp of the first activity of a user in a reasonable manner. Overall, this method gives solid evidence about how much time employees spend in the ERP system for all common activities of the process.

3.3 Duration of a Workday and Overall FTEs

A relevant question is whether behavioural changes occurred during the lockdown. To tackle this question, we visualise the distribution of actions over the day. Further, we can deduct the length of a working day based on the first and the last timestamp of a day. E.g., for User1 in Table 2, we would deduct a starting time of the workday on the 4th June 2020 at 8:35 and an end time of the workday at 17:25, resulting in a workday duration of 8.8 hours. Starting from the workday duration in hours of individual users, it is possible to deduct the overall working hours required to execute the respective process and the number

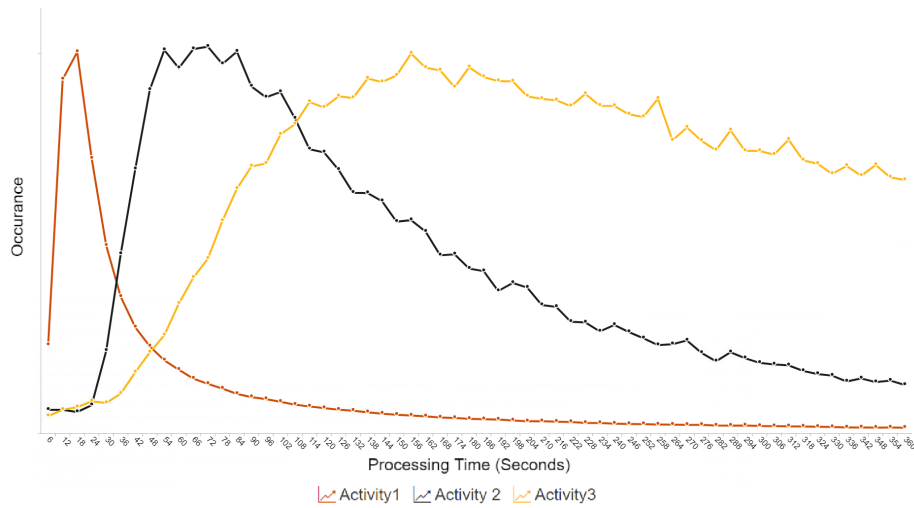


Fig. 1: Distribution of processing time for three different activities, with one being executed typically in less than 1 minute (orange), one typically executed within 1 to 2 minutes (black), and one typically requiring more than two minutes (yellow). For each activity, we calculate the typical processing time, which is the mode of the distribution.

of FTEs involved. It should be noted that this approach neglects the duration of the first activity of a user on a respective day. A solution for this is to approximate the start of the first activity using the average observed processing time of the respective activity.

4 Results achieved

4.1 Overall Process KPIs

Our standard Effort Mining analysis results in a table which summarises the most relevant KPIs describing the work time consumption of the process in scope, as depicted in Table 3: for each activity, we quantify how often it is executed, we deliver the typical processing time required to execute the activity in the ERP system, the total time spent in the organisation to execute the respective activities in the ERP system, as well as the overall working time required. If required, it is also possible to break down the results for different organisational units and compare, for instance, the processing times for certain activities across the organisation.

4.2 Evolution of Processing Times

When looking at changes in the working behaviour caused by the lockdown-induced remote working, it makes sense to look at temporal changes of KPIs

Activity	Typical Processing Times	Count Activities	Overall Time in ERP System	Overall Working Time
EventA	45 sec	10.0 k	125 hours	431 hours
EventB	50 sec	8.3 k	115 hours	398 hours
EventA,EventC, EventD	75 sec	13.5 k	281 hours	616 hours
EventD	60 sec	14.2 k	232 hours	533 hours
...

Table 3: The standard Effort Mining analysis yields the displayed information.

listed in Table 3. In the following, we will look at the temporal evolution of processing times, the distribution of workday duration, and the number of executed activities. We will specifically compare the results from April/May of 2019, the year before the first lockdown, and the results of April/May 2020 when the first and most strict lockdown was imposed on our clients.

Figure 2 a) depicts how often Activity 3 was executed in each month in scope: overall, we do not observe a significant increase or decrease in activities in any period. In Figure 2 b) we aim to learn whether the typical processing time in the ERP system for activity execution changed over time. We define, as described in subsection 3.2, an upper threshold to define a sample and calculate the mean of the sample as well as its standard error. When looking at the resulting plot, we do not observe an increase in the processing time from the moment employees had been sent home to work remotely. We rather observe a slight constant decrease over time, while the number of executed activities remains at a similar level (Figure 2 a)). This trend is underlined by Figure 2 c), which compares the distribution of observed processing times in April/May 2020 for Activity 3 with the observed processing times observed in the previous year: the distribution in 2020 is shifted towards faster processing times, which might be an effect of a more routinised activity execution by the users. But a significant effect on performance induced due to the lockdown starting in March 2020 was neither observed for the shown activity, nor for any other activity.

4.3 Daily Rhythm of Employees

In the next step, we aim to understand whether the daily routine of our clients’ employees changed during the lockdown. To achieve this, we examine the distribution of timestamps of the recorded activities over the day, which carries information about the activity of the ERP system users and their daily rhythm. Further, we will look at the duration of workdays as defined in subsection 3.3 to elucidate if users had longer or shorter working days.

Figure 3 a) compares the distribution of occurring activities over the day for April/May 2019 and April/May 2020. The distributions appear overall very similar: the number of recorded activities in the ERP system increases in the

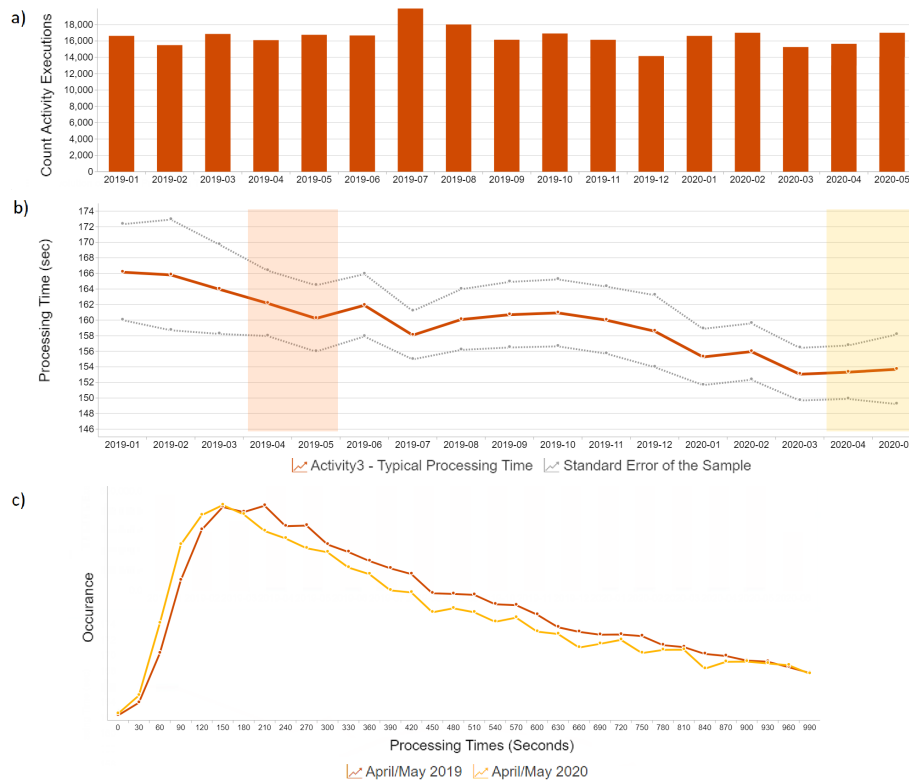


Fig. 2: a) Depicts the how often a certain activity is executed for each month. b) Visualises the evolution of the mean processing time of the taken sample with the corresponding standard error. The shaded areas mark the time ranges used in c). c) Compares the distributions of recorded processing times for Activity 3 in April/May 2019 and 2020. The shift to faster execution times underlines the effect observed in b).

morning between 7:00 and 8:00, which corresponds to the start of the working day of the users. Activity in the ERP system decreases drastically between 12:00 and 13:00, which corresponds to the lunch break, and rises after that to an activity level slightly lower than in the morning. Between 17:00 and 18:00, the number of activities decreases again as people finish their workday. When comparing the distribution of April/May 2020 during the lockdown with employees mostly working from home with the one from the previous year, we do not observe remarkable changes. This is indicative of the stable daily rhythm of the employees. It is worth noting that it was possible to identify organisational units for which late work after 20:00 increased during the lockdown (up to 3 % of all activities were executed after 20:00). However, this is in the range of typical fluctuations of late work observed within the organisation.

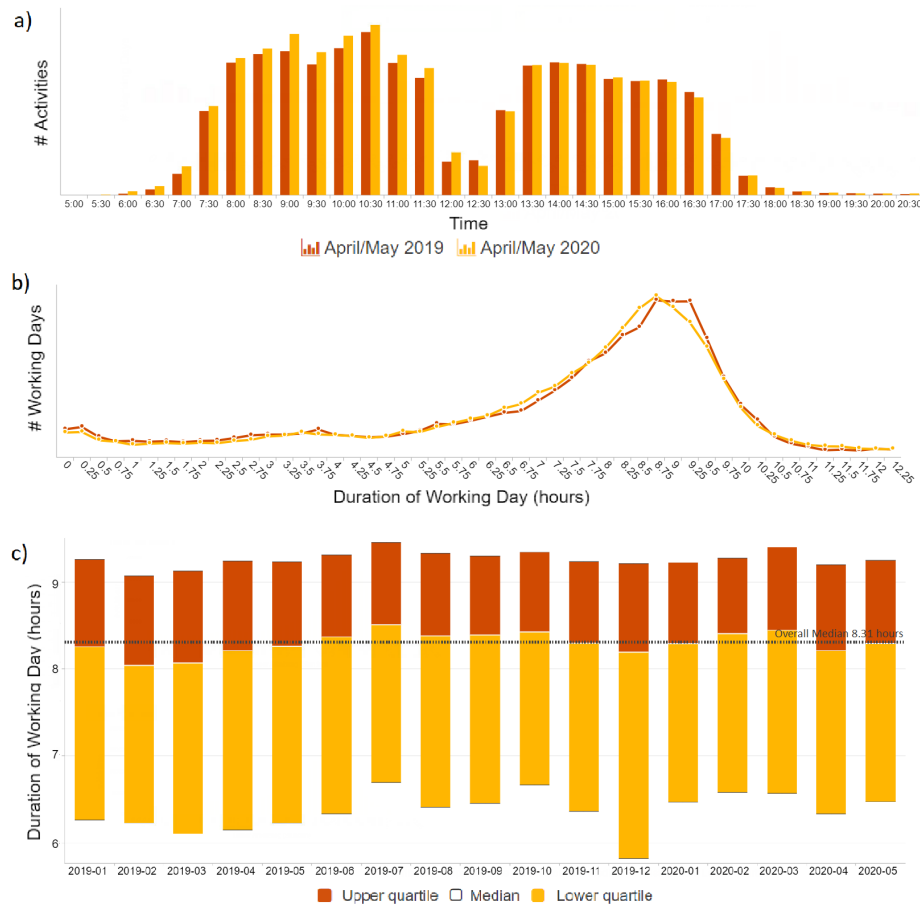


Fig. 3: a) Compares the total number of activities occurring at different times of the day for April/May 2019 and April/May 2020. b) Compares the duration of workdays during the same time periods. c) The boxplot depicts the temporal evolution of the distributions shown in b). The dashed line indicates the median workday duration.

Figure 3 b) compares the distributions of the workday duration for both periods. Many users participate in activities outside of the respective process and execute only for a fraction of their actual workday activities in the process. These users create short working days between 0 and 7 hours. Both distributions have a clear maximum between 8 and 10 hours, which reflects the typical workday duration. Longer working days are rather an exception for both periods in scope. It could be that for 2020 the distribution is shifted by 0.2 to 0.3 hours to a slightly shorter workday. However, when evaluating the temporal evolution of

the workday duration (Figure 3 c)), it becomes obvious that this is within the range of natural fluctuations.

Overall, this one-off analysis provided the respective client with the security that the productivity and the daily working rhythm of the employees did not change measurably due to the lockdown and that no immediate control action was required after two months of home office.

5 Outlook and lessons learned

This paper has presented Effort Mining, a new approach to measuring the effort required by activities in a process, thereby adding a novel perspective to process mining and allowing the precise quantification of effort in terms of time. We illustrate the use of Effort Mining in a case study based on real industry data, comparing the productivity of employees before and during a lockdown. As a result of this analysis, the client had the security that the process execution and employee performance was not impaired measurably by the home office. However, as the scope of this one-off analysis was limited to two months after the lockdown, long-term effects are not covered here. Future work might extend the time period covered by this analysis.

Overall, the feedback we received on this method was excellent and the calculated values of typical processing times and overall working times were in good agreement with the expectations of our clients. Another validation of the methodology was delivered by evaluating the overall working time of the most active UserIDs: They delivered a number of working hours to be expected by one FTE. After the initial application of effort mining in the here presented usecase, we used the methodology for several clients on different processes, such as order-to-cash, purchase-to-pay, record-to-report and enterprise asset management based on SAP data. We found that the methodology appears to be process-, industry- and system-agnostic. The only requirement is that the ERP system captures a sufficiently large amount of similar activities from sufficiently active users, which is typically given for core systems of large organisations.

One of the most relevant opportunities that arise through Effort Mining is judging the systemic work effort present in any given ERP system. As we utilise the execution pattern of users through all process steps, we can understand how users spend their time in the system. From that, we derive the typical processing time for each process step, allowing us to judge the business impact of each process execution. With effort mining, the time spent on process steps can be made transparent without using any additional, intrusive software component. We utilised this approach to create transparency in terms of work effort executed at home during the COVID-19 induced lockdown. It became evident that relocating employees to their own home for working did, in this case, not affect processing times or the number of activity executions. Further, the daily rhythm of the system users showed no drastic change in the remote working setting. A key message for the client was that the shift to working from home did not compromise human efficiency in the process execution. However, one should be

careful to generalise from this case to the situation of other firms as the results will depend on the employee behaviour and the culture of the organisation.

An additional opportunity for Effort Mining is the quantification of automation opportunities. Quantifying the systemic work effort in hours required to execute each process step enables us to associate a clear monetary value with each manual activity execution. This translation of systemic work effort into costs allows companies to optimise their spending on business improvement opportunities, e.g., robotic process automation initiatives [8].

An analytical area heavily impacted by Effort Mining could be business process simulation and prediction: these fields demand adequate information about resources used as an input, including human resources [9]. A systematic deduction of the time required to execute activities that can be scaled to all activities required for process execution could improve the quality of the output of these methods. Applications for this include but are not limited to the simulation of process improvement scenarios as well as workload predictions.

References

1. van der Aalst, W.: Process Mining Manifesto. In: Business Process Management Workshops, pp. 169–194. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
2. Arias, M. et al.: Human resource allocation in business process management and process mining: A systematic mapping study *Management Decision*, **56**(2), 376–405. (2018)
3. Nogayama, T. et al.: Estimation of Average Latent Waiting and Service Times of Activities from Event Logs. In: Business Process Management. BPM 2016. Lecture Notes in Computer Science, pp. 172–179 (2015)
4. Goncalves, R. et al.: Estimation and Characterization of Activity Duration in Business Processes. In: Business Process Management. BPM 2016. Lecture Notes in Computer Science, pp. 729–740 (2016)
5. Berkenstadt, G. et al.: Queueing Inference for Process Performance Analysis with Missing Life-Cycle Data. In: 2nd International Conference on Process Mining (ICPM), pp. 57–64 (2020)
6. Graafmans, T. et al.: Process Mining for Six Sigma. *Bus. Inf. Syst. Eng.* (2020)
7. WHO Homepage, <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>. Last accessed 27th May 2021
8. Schuler, J. et al.: Implementing robust and low-maintenance Robotic Process Automation (RPA) solutions in large organisations, *SSRN 3298036* (2018)
9. Tumay, K.: Business process simulation. In: Proceedings Winter Simulation Conference, pp. 93–98. IEEE, Coronado, CA, USA (1996)