Designing And Evaluating Critical Path Methods & Genetic Algorithms to Assess Critical Activities Within an Organization Using Process Mining

Abou Keita

FALL 2022

CAPSTONE RESEARCH

Zeanique L. Barber

Abstract

This paper addresses the problem of inefficiencies in business processes using Process Mining. The key idea of this paper is to propose different methodologies and solutions that can address the problem. More specifically, I introduce formal definitions of event logs, process models, Genetic Algorithms, genetic algorithm python library, and conformance-checking techniques that measure the fitness and deviation of the generated process model for a given event log. The model was generated on three different datasets: a synthetic event log and two real-life event logs. The proposed algorithms are compared with each other in terms of their fitness and deviation from the original dataset. The datasets differ in their lengths, sizes, and complexity. It is shown that the Genetic Algorithm provides superior performances, and a graphic representation of the process models are shown.

Introduction

What do every organized company have in common (beside of course the goal to make as much money as humanly possible and extract every cent out of their consumers)? They all have processes, or activities, that are routinely performed by either a human agent or computer-driven, to function daily. These processes and activities can become very complex depending on the task or as the organization scales. This can become an issue as some activities to complete a certain task can have redundancy and bottleneck that can lead to efficiency in the company's operation. These inefficiencies can present certain problems becomes they could cost the company time, money, and personal or machine resources that could be allocated elsewhere. We all know that at the end of the daytime is money.

This research aims to utilize two different methodologies employed in another different field that can resolve and address the inefficiencies mentioned above. The purpose and origin to do the research came from the following reasons:

- Identify key tasks, activities, and bottlenecks, in a business process and their relationship to each other
- Help businesses get a more objective insight into their internal/external processes
- Provide operational improvement suggestions based on the findings
- Reduce unnecessary tasks/activities in the current business process

In an article from Gibbons, a Forbes writer, he describes 5 red flags that create inefficiencies within an organization's business process. The first red flag is that employees depend [on] manual and redundant processes, meaning that the same task is repeated multiple times (2020). The second red flag is that information isn’t centralized or immediately available (2020). The third reason is that everyday processes are complex and require hoop jumping. The fourth reason is that tasks fall by the wayside because they’ve been abandoned. Lastly, an organization’s corporate systems barely communicate with each other (2020). All these reasons create an environment that is not very efficient and not cost-effective. Employee time is not maximized, and the limited resources are not distributed evenly. The wasted time of an employee doing one task already executed by another employee could cost a delay in the pipeline.

This paper will focus on using a process mining approach of event logs to discuss and develop an implementation of the Critical Path Method and Genetic Algorithm. With that said, we must define and examine what process mining consists of before tackling the implementations. Process Mining (PM) is a new discipline that emerged in the late 1990s at the Eindhoven University of Technology with the pioneering work of Professor Wil van der Aalst. PM is defined as [bridging] the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining (*Homepage Wil Van Der Aalst*, n.d.). PM discovers, monitors, and improves business processes as they are and not as people think they might be [through] extracting knowledge from event logs readily available in today's information systems, to visualize business processes — and their every variation — as they run *(Process Mining Introduction, n.d.)*. The reason I chose PM for this research is because of the benefits it offers (*Celonis Process Mining Software*, n.d.):

- Objectivity and fact-based insights derived from actual data to help audit, analyze, and improve existing business processes.
- Faster, cheaper, and more accurate than lengthy and often subjective process mapping workshops.
- No rip-and-replace is needed as PM works on top of existing systems, helping leverage existing technology investments

PM follows four simple steps in sequential order. The four steps are extracting data, process discovery, conformance checking, and finally performance analysis or enhancement *(Process Mining Introduction, n.d.)*. In the first step, the user collects event data through different Information Systems like a database, website, or application. In the second step, the user applies different process discovery algorithms and finds out the actual process happening of the collected data. In the third step, the user relies on conformance checking to evaluate the newly generated model or ideal model compared to what is happening from the process discovered. The user can replay the event data on this model to figure out if there are any deviations between the real-life process (happening every day) and the ideal process (from the model).

Lastly, performance analysis describes the different behavior of the generated models compared to the original model. In addition, it can be used to discover bottlenecks occurring in an event log *(Process Mining Introduction, n.d.)*. Enhancement goes beyond measuring and comparing. It's focused on extending the model with additional information, such as location data, costs, timing, etc. Enhancing the model with such attributes helps improve its performance and conduct more advanced analysis. Applying these techniques would allow us to map processes and discover gaps and inefficiencies in the workflow *(Editor, 2021)*.

An event log is defined as a collection of cases and a case can be seen as a trace/sequence of events that can be coming from a wide variety of sources *(Process Mining Introduction, n.d.)*:

- Database system (e.g., patient data in a hospital),
- Comma-separated values (CSV) file or spreadsheet,
- Transaction log (e.g., a trading system),

- Business suite/ERP system (SAP, Oracle, etc.),
- Message log (e.g., from IBM middleware),
- Open API provides data from websites or social media *(Process Mining Introduction, n.d.)*

The ontology of an event log consists of an activity, timestamp, resource, event, case, and trace. Each part is defined below (Hussain, 2022):

- Activity: A well-defined step in some workflow/process.
- Timestamp: The date and time at which something worthy of note happened.
- Resource: Staff and/or other assets used/consumed in the execution of an activity.
- Event: At a minimum, the combination of activity and a timestamp. Optionally, events may have associated resources, life cycles, and other data.
- Case: A related set of events denoted, and connected, by a unique identifier where the events can be ordered.
- Event Log: A list of cases and associated events.
- Trace: A distinct pattern of case activities within an event log where each activity is present at most once per trace. Event logs typically contain many traces

Activity Timestamp Resource

SessionID	Page	Timestamp	CookieID	DataCenter	SiteVersion
487434	portal.aspx	2016-01-01 15:34:01	A	phoenix	1.12
487434	dashboard.aspx	2016-01-01 15:34:15	A	phoenix	1.12
487434	purchaseorderreport.aspx	2016-01-01 15:34:30	A	phoenix	1.12
487435	portal.aspx	2016-01-01 14:01:10	B	phoenix	2
487435	help.aspx	2016-01-01 14:03:23	B	phoenix	2
487435	contactus.aspx	2016-01-01 14:04:07	B	phoenix	2
487436	portal.aspx	2016-01-01 17:11:17	A	phoenix	1.12
487436	myteam.aspx	2016-01-01 17:12:41	A	phoenix	1.12
487436	expensereports.aspx	2016-01-01 17:12:55	A	phoenix	1.12

In this example, the traces for this event log are:

- *Portal, dashboard, purchase order report*
- *Portal, help, contact us*
- *Portal, my team, expense reports (Image by (Hussain, 2022))*

Background or Literature Review

There are four past research and literature covering the topic of applying genetic algorithms to process mining or optimization on the event log. In research van der Aalst et al, presented a genetic algorithm to discover a Petri net given a set of event traces (2005). They used a causal matrix as their internal representation. The difference between this work and mine is that van der Aalst et al use a matrix causal as their internal representation, which they claimed to be more expressive. Activities and their respective percentages are used as the internal representation of individuals in the population in this paper.

The second paper is a genetic algorithm-based approach to identify cost reduction opportunities in the event (Low et al., 2014). The motivation for the research was to identify a more efficient execution scenario where cases finish earlier and the utilization of the resources is more optimal… [by] applying the notion of cost towards optimizing time and resource (Low et al., 2014). On top of that framework, a comprehensive and adaptive cost structure that captures different cost-related dimensions was proposed and incorporated as the objective function. Finally, the optimization is realized with the implementation of a set of genetic algorithm variants. The overall cost of a business process (based on the process behavior represented in a log) is computed, not the cost of individual cases (Low et al., 2014). The method is similar to what I am trying to achieve, but overall, this paper aims to look at comparing the Critical Path Method and Genetic Algorithm.

The third paper addressed the problem of supervised classification of time event logs of two classes: positive and negative populations. Oliveira et al. determined some process model that fits well the positive event logs and poorly the negative ones (2020). The event log is set as a classification and as an optimization problem for the determination of a process model that maximizes its replayability for the positive population and minimizes its replayability for the negative one (Oliveira et al., 2020). The results provided superior performances and a graphic representation of the process model of the positive event logs (Oliveira et al., 2020). Lastly, the research from Effendi and Sarno proposed to development of a Non-Linear Optimization of the Critical Path Method (2017). After identifying the Critical Path Method, they used linear programming to optimize the data to get minimum duration with minimum additional cost. The proposed method as two steps. First, calculate optimization data and then, optimize the data using non-linear modeling (Effendi & Sarno, 2017).

Methods & Materials

This section provides the different technologies used for the research paper. The PM software used is PM4PY. PM4PY is a python library that supports (state-of-the-art) process mining algorithms in python. It is open source (licensed under GPL) and intended to be used in both academia and industry projects. It is a product of the Fraunhofer Institute for Applied Information Technology (Burattin et al., 2018). PM4PY was extensively utilized for this research in the following ways:

- Read the XES file extension event logs
- Convert the original XES file extension into Pandas data frame and vice-versa for building analysis
- Perform process discovery of the event log file through many different algorithms
- Analyze the event log through the extraction of events, activities, traces, trace variants
- Analyze the relationship among resources performed for each activity
- Perform a social network analysis among resources

Google Collaboration was used for writing the codes and importing the files. The research comprises comparing the original datasets with two different techniques. The two techniques are the Critical Path Method and Genetic Algorithm optimization Genetic Algorithm optimization will have three variations. This is done to identify the most suitable application of it in Process Mining. The first variation is my implementation code of a Genetic Algorithm on the datasets. The second variation consists of using EasyGa, a Genetic Algorithm python library, on the datasets. Lastly, I will modify the fitness function (more on that on that later) to resemble the flow of the activities. The last variation will be called EasyGA with BPMN modeling.

The research consists of four steps performed on each dataset. The four steps are as follows:

- Generate Process Discovery, performance analysis, and statistics on each separate original dataset.
- Apply Genetic Algorithm (GA), EasyGa, and EasyGA with BPMN modeling on the datasets to gather the result of what the new activities would become.
- Generate Process Discovery, performance analysis, and statistics for each of the newly gathered results.
- Evaluate and compare the newly generated models' deviation from the original dataset

The newly generated business process models will be compared to their original counterparts. In addition, the research will look at some other parameters such as business process time duration, the number of activities within the process, and the number of events involved in the process to see the changes in the newly generated business process models compared to their original counterparts. Conformance Checking will be used to test and evaluate the models. The Resource Profile from event logs is also possible meaning that the behavior of a resource can be measured over some time with different metrics (Fraunhofer Institute for Applied Information Technology, n.d.).

Methodologies

Critical Path Analysis (CPA) | Critical Path Method (CPM)

The Critical Path Method (CPM) is a technique that identifies necessary tasks for project completion and determines scheduling flexibilities. In the context of this project, it will be determining the necessary activities. It is the longest sequence of activities (cases) that must be finished on time for the entire project to be complete. Any delays in critical tasks will delay the rest of the project (process). It revolves around discovering the most important tasks in the project timeline, identifying task dependencies, and calculating task durations (Team Asana, n.d.). CPM is implemented in projects for the following reasons (Team Asana, n.d.):

- Helps identify the most important tasks
- Helps reduce timelines:
- Compares planned with actual: the critical path method can also be used to compare planned progress with actual progress.
- Improves future planning: CPM can be used to compare expectations with actual progress. The data used from current projects can inform future project plans.
- Facilitates more effective resource management: CPM helps project managers prioritize tasks, giving them a better idea of how and where to deploy resources.
- Helps avoid bottlenecks: Bottlenecks in projects can result in lost valuable time. Plotting out project dependencies using a network diagram, will give you a better idea of which activities can and can't run in parallel, allowing you to schedule accordingly.
- Provides valuable insight on how to plan projects, allocate resources, and schedule tasks

Team Asana provides the five steps involved in finding the CPM of a process (Team Asana, n.d.):

1. List all activities
2. Identify dependencies between activities
3. Create a network diagram
4. Estimate task duration
    1. Forward pass: This is used to calculate early start (ES) and early finish (EF) dates by using a previously specified start date. ES is the highest EF value from immediate predecessors, whereas EF is ES + duration. The calculation starts with 0 at the ES of the first activity and proceeds through the schedule. Determining ES and EF dates allows for the early allocation of resources to the project.
    2. Backward pass: This is used to calculate late start (LS) and late finish (LF) dates. LS is LF - duration, whereas LF is the lowest LS value from immediate successors. The calculation starts with the last scheduled activity and proceeds backward through the entire schedule.
5. Calculate the critical path

Step 1: Write down the start and end time next to each activity.

- The first activity has a start time of 0, and the end time is the duration of the activity.
- The next activity's start time is the end time of the previous activity, and the end time is the start time plus the duration.
- Do this for all the activities.

Step 2: Look at the end time of the last activity in the sequence to determine the duration of the entire sequence.

Step 3: The sequence of activities with the longest duration is the critical path.

1. Calculate the float --> Total float vs. free float
    1. *Critical tasks with zero float mean that they are part of the critical path.* Tasks with positive float numbers belong in the non-critical path, meaning they may be delayed without affecting the project completion date. If short on time or resources, non-critical tasks may be skipped.
    2. Total float: the amount of time that an activity can be delayed from the early start date without delaying the project finish date or violating a schedule constraint. Total float = LS - ES or LF - EF
    3. Free float: Refers to how long an activity can be delayed without impacting the following activity. There can only be free float when two or more activities share a common successor. On a network diagram, this is where activities converge. Free float = ES (next task) - EF (current task)

Table 1 from (Usmani, 2022) shows the advantages and disadvantages of CPM. Whilst it is a great tool to use, it cannot be used for every situation.

Table 1 – Advantages and disadvantages of Critical Path Method

Advantages	Disadvantages
• **Shows a graphical view of the project.** • **You can discover and visualize dependencies.** • **It aids in project planning, scheduling, and controlling.** • **It helps in contingency planning.** • **You can see the critical path and identify critical activities.** • **It helps you assign the float to activities and flexibility to float activities.** • It shows you where you can take action to bring projects back on track	• An optimal planning tool assumes that all resources are always available for the project. • does not consider resource dependencies. • There is a chance of miscalculating float or slack. • Less attention is paid to non-critical activities, though sometimes they may become critical activities in the view of the organization • Projects based on the critical path often do not finish on time. • For a complex project, you will need to use project management software to develop the critical path scheduling that will add cost to your project.

Genetic Algorithm

A search heuristic that is inspired by Charles Darwin's theory of natural evolution. This algorithm reflects the process of natural selection where the fittest individuals are selected for reproduction to produce offspring of the next generation (Mallawaarachchi, 2020).

How does Genetic Algorithm Work

The notion of Natural Selection

The process of natural selection starts with the selection of the fittest individuals from a population. They produce offspring which inherit the characteristics of the parents and will be added to the next generation. If parents have better fitness, their offspring will be better than parents and have a better chance of surviving. This process keeps on iterating and in the end, a generation with the fittest individuals will be found. This notion can be applied to a search problem. We consider a set of solutions for a problem and select the set of best ones out of them. Five phases are considered in a genetic algorithm (Mallawaarachchi, 2020).

1. Initial population→Fitness function→ Selection→Crossover→Mutation→Termination

The process begins with a set of individuals which is called a Population. Each individual is a solution to the problem you want to solve. An individual is characterized by a set of parameters (variables) known as Genes. Genes are joined into a string to form a Chromosome (solution). In a genetic algorithm, the set of genes of an individual is represented using a string, in terms of an alphabet. Usually, binary values are used (a string of 1s and 0s). We say that we encode the genes in a chromosome: Population, Chromosomes, and Genes (Mallawaarachchi, 2020).

Fitness Function

- The fitness function determines how to fit an individual (the ability of an individual to compete with other individuals). It gives a fitness score to each individual. The probability that an individual will be selected for reproduction is based on their fitness score.

Selection

- The idea of the selection phase is to select the fittest individuals and let them pass their genes to the next generation.
- Two pairs of individuals (parents) are selected based on their fitness scores. Individuals with high fitness have more chances to be selected for reproduction.

Crossover

- Crossover is the most significant phase in a genetic algorithm. For each pair of parents to be mated, a crossover point is chosen at random from within the genes.

Crossover point

Offspring are created by exchanging the genes of parents among themselves until the crossover point is reached.

Mutation

In certain new offspring formed, some of their genes can be subjected to a mutation with a low random probability. This implies that some of the bits in the bit string can be flipped. Mutation occurs to maintain diversity within the population and prevent premature convergence.

Termination

The algorithm terminates if the population has converged (does not produce offspring which are significantly different from the previous generation). Then it is said that the genetic algorithm has provided a set of solutions to our problem.

Genetic Algorithms are ideal for finding an optimal solution for a large set of datasets. It is algorithmically done so it is faster than CPM and can handle large input dataset

There are three variations of Genetic Algorithms implemented with a different focus on the fitness function and population of the same dataset.

Table 2 - Advantages and disadvantages of Genetic Algorithms

Advantages	Disadvantages
• **Faster than other algorithms in running on a simple and complex fitness function** • Easier. If the vector representation of the individual is right, we can find out a solution without deep analysis work.	• Random heuristics sometimes doesn't find the optimum. • It is not a complete algorithm (not always the algorithm finds a suitable solution). Sometimes it can get stuck with a local maximum problem. Nevertheless, crossover operation (we will point out further down) helps to mitigate it, although this implies more iterations. • Difficult to understand and interpret, making it hard to know why a particular solution was found. • Sensitive to the initial conditions and may not find the global optimum solution. • Expensive to implement, and time-consuming to run with complex fitness functions. • Difficult to debug. • Difficult to optimize in a complex situation

Custom GA

In the implementation of Custom GA, I implemented my version of a Genetic Algorithm. In this variation, each activity is considered a single gene. The fitness function is focused on the percentage that each activity occupies in the entire process. The percentage is computed from the

organizational mining and the relative days total of each activity. If the combination of activities reaches 100 then it would stop. The fitness limit is set to stop the generation when a combination of activities reaches 100. The generation limit is the limit of how many times the process is repeated, think of epoch for machine learning. In this variation, the generational limit was intended to be 10000 generations, but an error was encountered and only ran for one generation before outputting a result.

EasyGA

The EasyGA variation relies on using the EasyGA python library. EasyGA is a python package designed to provide an easy-to-use Genetic Algorithm. The package is designed to work right out of the box, while also allowing the user to customize features as they see fit (Wilczak & Nguyen, n.d.). EasyGA removes the complexity and difficulty of creating one's custom GA. It prints interfaces that show a graph of the population changes and fitness function

For this research, EasyGA is setup in the following way.

- The *chromosome length* and *population size* attributes specify the input length or the number of genes (attributes/activities) the user is putting in. In my case, the attributes are "Activities". The number of activities for each dataset is used as the chromosome length and population size. Each activity is treated as a single gene
- The *target fitness type* attribute defines the objective target. It is set to “max” as we are trying to maximize the number of tasks contributing to the percentage limit.
- The *generation goal* attribute defines how many generations will the GA run for. For this research, the generation goal is set to 10,000.
- Fitness Function
    - The fitness accepts the *chromosome_data* (collection of all genes) as input.
    - The fitness value is based on the percentage that each activity holds from the organization mining. If a gene is turned 'on', then it increments the gene's (aka activity) percentage into the overall fitness.
    - The fitness value tries all variation and combination that adds up to 100 percentage until the generation goal limit
    - Returns all variation and the best fitness (best combination)

EasyGA w/ BPMN Modelling

The EasyGA w/ BPMN Modelling is a variation of the EasyGA method. In this method, each activity is not treated as a single gene in the dataset. This method tries to replicate the original dataset BPMN modeling considering a single path where the models experience concurrency/parallelism and the exclusivity of activities into the EasyGA. For this research, EasyGA w/ BPMN Modelling is setup in the following way.

- The *chromosome length* and *population size* attributes specify the number of activities for each dataset used as the chromosome length and population size. A gene randomly selects a single activity from a list of activities and counts it as a single gene for activities

concurrently occurring. This variation considers the combination of all the activities within the list with the other activities. The same principle is applied to XOR joins.

- The *target fitness type* attribute is set to “max” as we are trying to maximize the number of tasks contributing to the percentage limit.
- The *generation goal* attribute is set to 10,000.
- Fitness Function
    - The fitness accepts the chromosome data (collection of all genes but will only consider one for parallelism and XOR joins) as input.
    - The fitness value is based on the percentage that each activity holds from the organization mining. If a gene is turned 'on', then it increments the gene's (aka activity) percentage into the overall fitness.
    - The fitness value tries all variation and combination that adds up to 100 percentage until the generation goal limit
    - Returns all variation and the best fitness (best combination)

EasyGA w/ BPMN Modelling attempts to build Genetic Algorithms that are more tailored and specific to the dataset model. This reduces the number of genes that might be duplicated activities and removes duplicates.

Sub-Methodologies

Process Discovery

Process Discovery focuses on utilizing the event log data to create a visual process model of it. PM4PY offers many differs model algorithms representing the flow of the data in a different aspects. The three different Process Discovery algorithms utilized for this research are Heuristic Miner, Direct Follow Graph (DFG), and the BPMN Model. It is recommended to use all three models to understand the process model much better as each has its advantages and disadvantages.

BPMN Model

BPMN Model is an algorithm that takes an event log as input and then tries to smartly identify patterns. It looks at sequential behavior among certain groups of activities and concurrent behavior among certain activities. Finds patterns that it applies to itself. Then, it recursively applies itself and displays the subgroup it has detected.

Table 3- Advantages and disadvantages of BPMN Model

Advantages	Disadvantages
• **Shows both concurrent and XOR relationships among activities.** • Activities that can be done concurrently are inside of the <+> and <+> symbols • Activities considered to be XOR means that only one of the activities inside the <X> and <X> can be done at a given time and then moves forward • Shows a more general and high-level relationship between activities	• Difficult interpretation with nested concurrencies and XOR symbol so the user must pay attention to how activities are arranged

Heuristic Miner

The Heuristics Miner is an algorithm that acts on the Directly-Follows Graph, providing a way to handle noise and to find common constructs (dependency between two activities, AND). The output of the Heuristics Miner is a Heuristics Net, an object that contains the activities and the relationships between them. The Heuristics Net can be then converted into a Petri net (Sarojag, 2021). The flow is interpreted from top to bottom, meaning the top activities occurs first and the bottom activities occur later.

The green circle represents the start of the process. The orange circle represents the end of the process. Activities that are adjacent to each other represent "concurrency/parallelism" meaning that these sets of activities can be performed at the same time. It can also mean "a single activity must be chosen from the set presented". The activities are represented in squares with the arrows dictating the flow of direction that a case is moving. The number next to the arrows indicates the number of cases following this path. Cases can skip certain activities, get roll-back to a previous activity, or normally moves to the next activity

Table 4 - Advantages and Disadvantages of Heuristic Miner

Advantages	Disadvantages
• **Decision-makers can see the flow of cases and which activity tends to have the most.** • Shows cases that are 'revisiting' the same activity	• Impossible to distinguish between which exact activities can be performed with XOR joins

Direct Follow Graph (DFG)

Directly-Follows graphs are graphs where the nodes represent the events in the log and directed edges are present between nodes if there is at least a trace in the log where the source event is followed by the target event (Sarojag, 2021). These directed edges work nicely together with some additional metrics, such as (Sarojag, 2021):

- Frequency — The number of times the source event is followed by the target event.
- Performance — aggregation, for example, the average time elapsed between the two events

On top of these directed edges, it is easy to represent metrics like frequency (counting the number of times the source event/activity is followed by the target event/activity) and performance (some aggregation, for example, the mean, of time inter-lapsed between the two events/activities) (Fraunhofer Institute for Applied Information Technology, n.d.).

Table 5 - Advantages and disadvantages of Direct Follow Graph

Advantages	Disadvantages
• Best use case in identifying local relationships among activities	• Very messy and complex to understand • Not a general representation of the overall process which leads to a false interpretation of what is going on • Only show the LOCAL relationship between activities • Concurrency in the process itself can lead to cycles in the process map not indicating clear relationships

Social Network Analysis

A Social Network Analysis will be performed for each dataset and method. A Social Network Analysis (SNA) analyzes the relationship and association of resources with each other (if at all) in a process or across processes. Some of its metrics are Handover of Work, Subcontracting, Working Together, Similar Activities, Roles Discovery, Clustering, and Organizational Mining.

The Handover of Work metric measures how many times an individual is followed by another individual in the execution of a business process. It considers the directly-follows relationships between resources during the work of a case. An edge is added between the two resources if such relationships occur. It uses the resource as a node of the graph, instead of the activity (Fraunhofer Institute for Applied Information Technology, n.d.). The Subcontracting metric calculates how many times the work of an individual is interleaved by the work of some other individual, only to eventually "return" to the original individual (Fraunhofer Institute for Applied Information Technology, n.d.). The Working Together metric calculates how many times two individuals work together for resolving a process instance (Fraunhofer Institute for Applied Information Technology, n.d.). The Similar Activities metric calculates how much similar is the work pattern between two individuals (Fraunhofer Institute for Applied Information Technology, n.d.). In Role Discovery, a role is a set of activities in the log that are executed by a similar (multi)set of resources. Hence, it is a specific function in the organization. Grouping the activities into roles can help (Fraunhofer Institute for Applied Information Technology, n.d.):

- understanding which activities are executed by which roles.
- understanding roles themselves (numerosity of resources for a single activity may not provide enough explanation)
- Initially, each activity corresponds to a different role and is associated with the multiset of his originators. After that, roles are merged according to their similarity, until no more merges are possible.

Given the results of applying an SNA metric, a clustering operation permits a group of resources that are connected by a meaningful connection in the given metric (Fraunhofer Institute for Applied Information Technology, n.d.). For example:

- Clustering the results of the *working together metric*, individuals that work often together would be inserted in the same group.
- Clustering the results of the *similar activities metric*, individuals that work on the same tasks would be inserted in the same group.

PM4PY provides a baseline method to get a list of groups (where each group is a list of resources) from the specification of the values of an SNA metric (Fraunhofer Institute for Applied Information Technology, n.d.).

Organizational Mining describes some group-related metrics (local diagnostics). Among these, we have (Fraunhofer Institute for Applied Information Technology, n.d.):

- Group Relative Focus: (on a given type of work) specifies how much a resource group performed this type of work compared to the overall workload of the group. It can be used to measure how the workload of a resource group is distributed over different types of work, i.e., work diversification of the group.
- Group Relative Stake: (in each type of work) specifies how much this type of work was performed by a certain resource group among all groups. It can be used to measure how the workload devoted to a certain type of work is distributed over resource groups in an organizational model, i.e., work participation by different groups.
- Group Coverage: concerning a given type of work specifies the proportion of members of a resource group that performed this type of work.
- Group Member Contribution: a member of a resource group concerning the given type of work specifies how much of this type of work by the group was performed by the member. It can be used to measure how the workload of the entire group devoted to a certain type of work is distributed among the group members.

Data

The research consists of three datasets. The datasets are fetched and downloaded from Figshare.com. Figshare is a repository where users can make all of their research outputs available in a citable, shareable, and discoverable manner (Figshare, n.d.). It is a reputable website to retrieve event logs datasets from real-life situations. The three datasets fetched are:

- Review Example Large: a synthetic event log published by the Eindhoven University of Technology (4TU.ResearchData, 2020)
- Helpdesk: an event log representing the process performed by help desk employees (Neu et al., 2021)
- BPI 2012 Challenge: an event log of a loan application process (Neu et al., 2021)

Some of the attributes in the datasets are:

- Concept:resources (Resources or users involved in the trace),
- Concept:name (Activity name)
- Timestamp:timestamp (Timestamp),
- Lifecycle:transition (Status of an event; start/end),
- Case:concept:name (Unique Case Id for each event)
- Case:description (Description of the event)
- Result (Decision taken by a given resource)

The Review Example Large (REL) is a synthetic event log that has a timeframe ranging from 2006 to 2252 (Figure A-1). The starting and ending activities of this dataset respectively are to invite reviewers. The process ends with accepting or rejecting a document. All traces started with the first activities but only 5068 were accepted and the remaining 4932 got rejected (Figure A-2). The total number of events in the data set is 236,360 and the total number of unique cases is 10,000 (Figure A-3). There are 14 unique activities (Figure A-4) with 10 different resources/people involved minus the "Invalid" entry (Figure A-5). As mentioned previously, this is a synthetic dataset, so the description reflects that. The result columns contain the decision taken by resources when some activities indicated a choice deciding if a document was rejected. In total, there were 17, 985 documents (Figure A-6). The median case duration is 30,721 days. In addition, I will be showing Exploratory Data Analysis (EDA), and SNA of this original dataset.

Figure 1 – Distribution of the Events over the Years for Review Example Large dataset

Figure 2 – Review Example Large dataset starting and ending activities

```
# imports the xes files and oupts the first and last activity
event_log = pm .read_xes(xes_path)
start_activities = pm.get_start_activities(event_log)
end_activities = pm.get_end_activities(event_log)
print("Start activities: {}\nEnd activities: {}".format(start_activities, end_activities))
```

```
parsing log, completed traces :: 100% 10000/10000 [00:11<00:00, 1246.51it/s]
Start activities: {'invite reviewers': 10000}
End activities: {'reject': 4932, 'accept': 5068}
```

Figure 3 - Review Example Large dataset number of event and unique cases

```
num_event = len(df_eventlog)
num_cases = len(df_eventlog['case:concept:name'].unique())
print("number of event: {} \n Number of unique cases: {}".format(num_event, num_cases))
```

```
number of event: 236360
 Number of unique cases: 10000
```

Figure 4 - Review Example Large dataset activities

```
[10] # unique activities
     df_eventlog['concept:name'].unique()
```

```
array(['invite reviewers', 'get review 3', 'get review 1', 'time-out 2',
       'collect reviews', 'decide', 'invite additional reviewer',
       'get review X', 'reject', 'time-out 1', 'get review 2',
       'time-out X', 'accept', 'time-out 3'], dtype=object)
```

```
[11] #14 unique activites
     df_eventlog['concept:name'].nunique()
```

```
14
```

Figure 5 - Review Example Large dataset resources

```
df_eventlog['org:resource'].unique()
```

```
array(['Mike', 'Mary', 'Pam', '__INVALID__', 'Anne', 'Wil', 'Carol',
       'Pete', 'Sara', 'John', 'Sam'], dtype=object)
```

```
df_eventlog['org:resource'].nunique()
```

```
11
```

Figure 6 - Review Example Large dataset dataframe

```
df_eventlog[df_eventlog['result'] == 'reject']
```

1 to 25 of 17985 entries Filter

index	org:resource	time:timestamp	concept:name	lifecycle:transition	case:concept:name	case:description	result
2	Mary	2006-01-05 00:00:00+01:00	get review 3	complete	1	Simulated process instance	reject
3	Pam	2006-01-06 00:00:00+01:00	get review 1	complete	1	Simulated process instance	reject
11	Mary	2006-01-24 00:00:00+01:00	get review X	complete	1	Simulated process instance	reject
18	Pete	2006-02-14 00:00:00+01:00	get review 2	complete	10	Simulated process instance	reject
34	Pete	2007-01-09 00:00:00+01:00	get review X	complete	10	Simulated process instance	reject
40	Sara	2008-01-11 00:00:00+01:00	get review X	complete	10	Simulated process instance	reject
60	Mary	2012-01-08 00:00:00+01:00	get review X	complete	100	Simulated process instance	reject
72	John	2019-07-20 01:00:00+02:00	get review 3	complete	1000	Simulated process instance	reject
73	Pam	2019-07-21 01:00:00+02:00	get review 2	complete	1000	Simulated process instance	reject
103	Sam	2141-09-16 01:00:00+02:00	get review 2	complete	10000	Simulated process instance	reject
124	John	2192-12-02 00:00:00+01:00	get review X	complete	10000	Simulated process instance	reject
142	John	2019-08-27 01:00:00+02:00	get review 1	complete	1001	Simulated process instance	reject
153	Sara	2078-03-31 01:00:00+02:00	get review X	complete	1001	Simulated process instance	reject
159	Pam	2019-08-17 01:00:00+02:00	get review 1	complete	1002	Simulated process instance	reject
170	Pete	2056-03-06 00:00:00+01:00	get review X	complete	1002	Simulated process instance	reject
184	Pam	2019-08-09 01:00:00+02:00	get review 1	complete	1003	Simulated process instance	reject
193	Sara	2022-12-10 00:00:00+01:00	get review X	complete	1003	Simulated process instance	reject
207	John	2029-09-22 01:00:00+02:00	get review X	complete	1004	Simulated process instance	reject
222	Pete	2019-08-12 01:00:00+02:00	get review 3	complete	1005	Simulated process instance	reject
239	Carol	2080-01-15 00:00:00+01:00	get review X	complete	1005	Simulated process instance	reject
242	Sara	2088-09-18 01:00:00+02:00	get review X	complete	1005	Simulated process instance	reject
248	Sam	2019-07-12 01:00:00+02:00	get review 3	complete	1006	Simulated process instance	reject

The Helpdesk (HD) dataset shows the task taken by a helpdesk for a company. HD has a timeframe ranging from 2010 to 2014 (Figure B-1). The starting and ending activities of this dataset respectively are mostly starting on *'Assign seriousness'* with 4384 events (Figure B-2). The event log ends with most events on *'Closed'* with 4384 events (Figure B-2).

The event log contains 21,348 events and 4580 unique cases (Figure B-3). There are 22 unique resources (Figure B-4) and 14 different activities involved (Figure B-5). The name of the resources was randomly generated using the Faker library (Faker's Documentation, n.d.) as the original dataset had invalid values in that attribute that were not compatible with performing social network analysis. The median case duration is 40 days. In addition, I will be showing the EDA, and SNA of this original dataset.

Figure 7 - Helpdesk dataset distribution of the events over the years

Figure 8 - Helpdesk dataset starting and ending activities

Get starting and ending activities

```
[ ] xes_start_activites = pm4py.get_start_activities(df_log)
    xes_end_activites = pm4py.get_end_activities(df_log)
    print("starting activity: {} \n ending activity: {}".format(xes_start_activites, xes_end_activites))

    starting activity: {'Assign seriousness': 4384, 'Insert ticket': 118, 'Take in charge ticket': 74, 'Resolve ticket': 2, 'Create SW anomaly': 1, 'Wait': 1}
     ending activity: {'Closed': 4557, 'Resolve ticket': 10, 'Wait': 8, 'Require upgrade': 3, 'VERIFIED': 1, 'Take in charge ticket': 1}
```

Figure 9 - Helpdesk dataset number of event and unique cases

```
[ ] num_event = len(df_log)
    num_cases = len(df_log['case:concept:name'].unique())
    print("number of event: {} \n Number of unique cases: {}".format(num_event, num_cases))

    number of event: 21348
     Number of unique cases: 4580
```

Figure 10 - Helpdesk dataset resources

```
[ ] df_log['org:resource'].unique()

    array(['Kelly Williams', 'Angelica Smith', 'James Rogers',
           'Martin Burton', 'Bradley Acosta', 'Richard Perry',
           'Ricardo Parker', 'Megan Ortiz', 'Mark Valencia',
           'Nicholas Fowler', 'Angela Kirk', 'Linda Brooks', 'Kristin Burton',
           'Wendy Thompson', 'Olivia Henry', 'James Cole', 'John Nichols',
           'Jennifer Young', 'Omar Collins', 'Alex Haley', 'Donna Pearson',
           'Michelle Glenn'], dtype=object)
```

Figure 11 - activities

```
[ ] #the unique activies taking places in the log
    df_log['concept:name'].nunique()

    14
```

The BPI 2012 dataset is a real-life event log of a loan application process. HD has a timeframe ranging from 2011 to 2012 (Figure C-1). The starting and ending activities of this dataset respectively are mostly starting on *''W_Completeren aanvraag'* with 4852 events and *'W_Afhandelen leads' with* 4739 events (Figure C-2). The event log ends with most events on *'W_Valideren aanvraag'* with 2751 events, *'W_Completeren aanvraag'* with 2355 events, and *'W_Afhandelen leads'* with 2235 events (Figure C-2).

The event log contains 72,413 events and 9658 unique cases (Figure C-3). There are 56 unique resources (Figure C-4) and 6 different activities involved (Figure C-5). The name of the resources was randomly generated using the Faker library (Faker's Documentation, n.d.) as the original dataset had invalid values in that attribute that were not compatible with performing social network analysis. The median case duration is 40 days. In addition, I will be showing the EDA, SNA, and VISPM of this original dataset. The median case duration is 9 days. In addition, I will be showing the EDA, and SNA of this original dataset.

Figure 12 – BPI 2012 dataset distribution of the events over the years

Figure 13 - BPI 2012 dataset starting and ending activities

```
[ ] xes_start_activites = pm4py.get_start_activities(df_log)
    xes_end_activites = pm4py.get_end_activities(df_log)
    print("starting activity: {} \n ending activity: {}".format(xes_start_activites, xes_end_activites))

    starting activity: {'W_Completeren aanvraag': 4852, 'W_Afhandelen leads': 4739, 'W_Beoordelen fraude': 67}
     ending activity: {'W_Valideren aanvraag': 2751, 'W_Completeren aanvraag': 2355, 'W_Afhandelen leads': 2235, 'W_Nabellen offertes': 1803, 'W_Nabellen incomplete dossiers': 457, 'W_Beoordelen fraude': 57}
```

Figure 14 - BPI 2012 dataset number of event and unique cases

```
num_event = len(df_log)
num_cases = len(df_log['case:concept:name'].unique())
print("number of event: {} \n Number of unique cases: {}".format(num_event, num_cases))

number of event: 72413
 Number of unique cases: 9658
```

Figure 15 - BPI 2012 dataset resources

```
df_log['org:resource'].unique()

array(['Jennifer Thomas MD', 'Robert Peters', 'Christina Barnes',
       'David Adams', 'William Arroyo', 'Anna Schroeder', 'Brianna Bell',
       'Brandon Powell', 'Aaron Ray', 'Scott Little', 'Ashley Daniel',
       'Lisa Henson', 'Jennifer Wilson', 'Jane Turner',
       'Brianna Martinez', 'Laura Morris', 'Wesley Mills',
       'Brittany Valencia', 'Rhonda Brady', 'Emily Rice', 'Patrick Cole',
       'Paul Alvarado', 'Justin Thompson', 'Vanessa Shepherd',
       'Matthew Walker', 'Miguel Reeves', 'Kathy Ingram', 'Derrick Watts',
       'Sarah Hogan', 'Martha Costa', 'Emily Rojas', 'Steven Jones',
       'Rebecca Guerrero', 'David Doyle', 'Michelle Brennan',
       'Donald Wilson', 'Austin Jones', 'Joshua Phelps', 'Nathan Thomas',
       'Mary Thompson', 'Patricia Lloyd', 'Kayla Woods', 'Luis Jackson',
       'Donald White', 'Shawn Barnes', 'Heather Patterson',
       'Pam Martinez', 'Jesse Moore', 'Travis Edwards', 'Ruth Davies',
       'Nancy Clayton', 'Sherry Klein', 'Jacob Wilson',
       'Michael Anderson MD', 'Jennifer Collier', 'Kevin Schultz'],
      dtype=object)
```

Figure 16 - BPI 2012 dataset activities

```
#the unique activies taking places in the log
df_log['concept:name'].nunique()

6
```

Process Discovery Model

Now, I will be performing the Process Discovery for each dataset and applying the methodologies to them. The outputs for the methods are shown below. The Social Network Analysis (SNA) showed are only for the original datasets. The reason is to demonstrate the capabilities of PM4PY in creating resource analysis and showing an SNA for each method and each dataset will take more space. However, the actual coding of the research contains SNA for each methodology of each dataset.

Review Example Large Dataset

Original

Figure 17 - Review Example Large Dataset EDA Activities

Figure 18 - Review Example Large Dataset EDA Resources

Figure 19 - Review Example Large Dataset BPMN Model

Figure 20 - Review Example Large Dataset Heuristics Model

Figure 21- Review Example Large Dataset DFG Model

Social Network Analysis

Original Dataset 1 - Handover of Work

Figure 22 - Review Example Large Dataset Handover of Work

Every resource is connected and at some point, follows each other in sequence at some points in the event log.

Original Dataset 2 - Subcontracting

Figure 23 - Review Example Large Dataset Subcontracting

Pete has work that is interleaved by John, Mary, and Pam that eventually comes back to him. On the other hand, the three people who do not interleave with other people's work are: Mike, Anne, and Carol

Original Dataset 3 - Working Together

Figure 24 - Review Example Large Dataset Working Together

All resources worked together to resolve a process instance at some point.

Original Dataset 4 - Similar Activity

Figure 25 - Review Example Large Dataset Similar Activity

Mike and Anne are the only two sharing the same similar work pattern. The other resources share a similar work pattern

Roles Discovery

- Activities: ['accept', 'collect reviews', 'invite additional reviewer', 'invite reviewers', 'reject'] Originators importance {'Mike': 72082, 'Anne': 72158}
- Activities: ['decide'] Originators importance {'Wil': 20000}
- Activities: ['get review 1', 'get review 2', 'get review 3', 'get review X'] Originators importance {'Pam': 5180, 'John': 5157, 'Carol': 5146, 'Sara': 5195, 'Pete': 5233, 'Mary': 5175, 'Sam': 5143}
- Activities: ['time-out 1', 'time-out 2', 'time-out 3', 'time-out X'] Originators importance {'__INVALID__': 35891}

There are four roles in the original dataset outputted at the start of each "Activity" with the resources involved the most in each role

Clustering

clustered similar activities metric:

{1': ['Sara', 'Pete', 'Carol', 'John', 'Mary', 'Sam', 'Pam'],

'0': ['Wil', 'Anne', '__INVALID__', 'Mike']}

working together clustered metric:

{'3': ['Sara', 'Pete', 'Carol', 'John', 'Mary', 'Sam', 'Pam'],

'0': ['Wil'],

'1': ['Anne'],

'2': ['__INVALID__'],

'4': ['Mike']}

We can observe that Sara, Pete, Carol, John, Mary, Sam, and Pam are in the same group because they work together very often. They are also in the same group in the similar activities metric meaning that they work on the same tasks, so it makes sense for them to be working together frequently. In addition, it makes sense that Mike and Anne are grouped in the similar activities metric as we previously established that they had a connected in the "Similar activities" graph. It is interesting to note that though they do the same task, they do not often work together. The same argument is observed with Mike not working together with no one either.

Organizational Mining

group_relative_focus

{'Mike': {'invite reviewers': 0.4975, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.5028, 'decide': 0.0, 'invite additional reviewer': 0.4985280151946819, 'get review X': 0.0, 'reject': 0.5097323600973236, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.4984214680347277, 'time-out 3': 0.0},

'Mary': {'invite reviewers': 0.0, 'get review 3': 0.14037216828478966, 'get review 1': 0.1429140127388535, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14313264346190027, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14397120575884823, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Pam': {'invite reviewers': 0.0, 'get review 3': 0.14826051779935276, 'get review 1': 0.14709394904458598, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14167450611476953, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13917216556688664, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'__INVALID__': {'invite reviewers': 0.0, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 1.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.0, 'reject': 0.0, 'time-out 1': 1.0, 'get review 2': 0.0, 'time-out X': 1.0, 'accept': 0.0, 'time-out 3': 1.0},

'Anne': {'invite reviewers': 0.5025, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.4972, 'decide': 0.0, 'invite additional reviewer': 0.5014719848053182, 'get review X': 0.0, 'reject': 0.4902676399026764, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.5015785319652722, 'time-out 3': 0.0},

'Wil': {'invite reviewers': 0.0, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 1.0, 'invite additional reviewer': 0.0, 'get review X': 0.0, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Carol': {'invite reviewers': 0.0, 'get review 3': 0.13875404530744337, 'get review 1': 0.14530254777070065, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14242709313264346, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14037192561487702, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Pete': {'invite reviewers': 0.0, 'get review 3': 0.14603559870550162, 'get review 1': 0.1449044585987261, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14397930385700847, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14437112577484504, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Sara': {'invite reviewers': 0.0, 'get review 3': 0.13794498381877024, 'get review 1': 0.14470541401273884, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.143414863593603, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14737052589482103, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'John': {'invite reviewers': 0.0, 'get review 3': 0.14724919093851133, 'get review 1': 0.13992834394904458, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14256820319849484, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13897220555888823, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Sam': {'invite reviewers': 0.0, 'get review 3': 0.14138349514563106, 'get review 1': 0.1351512738853503, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.14280338664158043, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14577084583083383, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0}}

group_relative_stake

{'Mike': {'invite reviewers': 0.13803723537082768, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.1395077828029189, 'decide': 0.0, 'invite additional reviewer': 0.5826142448877667, 'get review X': 0.0, 'reject': 0.06975389140145945, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.07008684553702728, 'time-out 3': 0.0},

'Mary': {'invite reviewers': 0.0, 'get review 3': 0.13410628019323673, 'get review 1': 0.138743961352657, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5880193236714976, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.1391304347826087, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Pam': {'invite reviewers': 0.0, 'get review 3': 0.1415057915057915, 'get review 1': 0.14266409266409266, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5814671814671815, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13436293436293437, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'__INVALID__': {'invite reviewers': 0.0, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.13928282856426402, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.0, 'reject': 0.0, 'time-out 1': 0.13864199938703295, 'get review 2': 0.0, 'time-out X': 0.5812042016104316, 'accept': 0.0, 'time-out 3': 0.14087097043827143},

'Anne': {'invite reviewers': 0.1392776961667452, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.13780869758030986, 'decide': 0.0, 'invite additional reviewer': 0.5854375121261676, 'get review X': 0.0, 'reject': 0.06701959588680396, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.0704564982399734, 'time-out 3': 0.0},

'Wil': {'invite reviewers': 0.0, 'get review 3': 0.0, 'get review 1': 0.0, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 1.0, 'invite additional reviewer': 0.0, 'get review X': 0.0, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.0, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Carol': {'invite reviewers': 0.0, 'get review 3': 0.1333074232413525, 'get review 1': 0.14185775359502525, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5884181888845705, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13641663427905168, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Pete': {'invite reviewers': 0.0, 'get review 3': 0.13797057137397287, 'get review 1': 0.13911714121918595, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5849417160328684, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13797057137397287, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Sara': {'invite reviewers': 0.0, 'get review 3': 0.1312800769971126, 'get review 1': 0.1399422521655438, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5869104908565929, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.1418671799807507, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'John': {'invite reviewers': 0.0, 'get review 3': 0.14116734535582703, 'get review 1': 0.13631956563893738, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5877448128757029, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.13476827612953268, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0},

'Sam': {'invite reviewers': 0.0, 'get review 3': 0.13591289130857476, 'get review 1': 0.13202411044137663, 'time-out 2': 0.0, 'collect reviews': 0.0, 'decide': 0.0, 'invite additional reviewer': 0.0, 'get review X': 0.5903169356406767, 'reject': 0.0, 'time-out 1': 0.0, 'get review 2': 0.14174606260937198, 'time-out X': 0.0, 'accept': 0.0, 'time-out 3': 0.0}}

group_coverage

{'Mike': {'Mike': 1.0},

'Mary': {'Mary': 1.0},

'Pam': {'Pam': 1.0},

'__INVALID__': {'__INVALID__': 1.0},

'Anne': {'Anne': 1.0},

'Wil': {'Wil': 1.0}

'Carol': {'Carol': 1.0},

'Pete': {'Pete': 1.0},

'Sara': {'Sara': 1.0},

'John': {'John': 1.0},

'Sam': {'Sam': 1.0}}

group_member_contribution

{'Mike': {'Mike': {'invite reviewers': 9950, 'invite additional reviewer': 41996, 'reject': 5028, 'collect reviews': 10056, 'accept': 5052}}, 'Mary': {'Mary': {'get review 3': 694, 'get review X': 3043, 'get review 1': 718, 'get review 2': 720}}, 'Pam': {'Pam': {'get review 1': 739, 'get review 2': 696, 'get review X': 3012, 'get review 3': 733}}, '__INVALID__': {'__INVALID__': {'time-out 2': 4999, 'time-out 1': 4976, 'time-out X': 20860, 'time-out 3': 5056}}, 'Anne': {'Anne': {'collect reviews': 9944, 'invite reviewers': 10050, 'invite additional reviewer': 42244, 'reject': 4836, 'accept': 5084}}, 'Wil': {'Wil': {'decide': 20000}}, 'Carol': {'Carol': {'get review 3': 686, 'get review X': 3028, 'get review 1': 730, 'get review 2': 702}}, 'Pete': {'Pete': {'get review 2': 722, 'get review X': 3061, 'get review 3': 722, 'get review 1': 728}}, 'Sara': {'Sara': {'get review X': 3049, 'get review 2': 737, 'get review 1': 727, 'get review 3': 682}}, 'John': {'John': {'get review 3': 728, 'get review X': 3031, 'get review 1': 703, 'get review 2': 695}}, 'Sam': {'Sam': {'get review 2': 729, 'get review 3': 699, 'get review X': 3036, 'get review 1': 679}}}

Now, let's unpack this mess of a result by using the definition provided beforehand to interpret each line. We will break down by each category and focus on the example highlighted in red to give a clear interpretation of a single output and you can apply the same or similar logic for each of the outputs within a category.

Using the definition of group_relative_focus, the first line is interpreted that the resource Mike contributed 49.75% in the invite reviewers activity, 50.28% in the collect reviews activity, invite additional reviewer: 49.85%, reject: 50.97, accept: 49.84 relative to all the other resource participant. He contributed 0% to all the other activities

The definition of group_relative_stake can be summed up as finding the percentage of workload a certain activity takes up when compared to the entire organization process. These are estimates from looking at the dataset, but it gives a good idea of which activities can be a bottleneck or be looked at as they comprised the most percentage.

For example, the activity invites reviewers performed by both Mike and Anne comprising 14% of the business process. The variants of getting review 1,2,3 activities would count as one activity as they are performed in parallels as seen in the models and they also are all around the same percentage of 14%. Decide contributes to 0.01%, Accept contributes to 7%, Reject is too low and has 0%, and collect reviews contributes to 13%.

As seen in the previous visual models, we see that most of the time spent in the entire process in either inviting additional reviewers or getting review X contributed to 58% of the entire A suggestion for the business is to investigate and improve the activities contributing the most in this observation. This step identified a bottleneck or an activity that needs to be investigated and improved. For this specific dataset, the activities are to invite additional reviewers or get review X. An interesting observation is that the "Invalid" resource only had activities that ended in time-out. On the other hand, none of the other resources had that problem. This makes sense as an "invalid" resource might refer to events that were discarded within the process and NOT an error in capturing the data when logging events.

Group_coverage is pretty much self- explanatory; this is used to show the proportion of resources involved in performing the activities. Since each resource is involved at least once, then they are all considered participants of the process. Group_member_contribution: specifies the number of times a resource contributed to the action. Mike contributed to 9950 invite reviews, 41996 additional invite reviews, rejected 5028, collected reviews 10,056, and accepted 5052. The same intuition can be applied to all the other resources.

CPM

- Number of events: 106155
- Number of unique cases: 10000
- The number of resources is: 11
- Total duration: 261865 days

Figure 26 - Review Example Large Dataset CPM EDA Activities

Figure 27 - Review Example Large Dataset CPM EDA Resource

Figure 28 - Review Example Large Dataset CPM DFG

Figure 29 - Review Example Large Dataset CPM BPMN Model

Figure 30 - Review Example Large Dataset CPM heuristics

Custom GA

Figure 31 - Review Example Large Dataset Custom GA Implementation

```
#Activities = namedtuple( 'Activities', ['name','aggregated_max_relative_day', 'weight'])
Activities = namedtuple( 'Activities', ['name','days_to_finish', 'percentage'])

max_days = 261865

things = [

          Activities('accept',85800 , 7),
          Activities('collect reviews',250 , 14),
          Activities('decide',251 , 1),
          Activities('get review 1',17 , 14),
          Activities('get review 2',17 , 14),
          Activities('get review 3',17 , 14),
          Activities('get review X',87109 , 59),
          Activities('invite additional reviewer',87108 , 59),
          Activities('invite reviewers',12 , 14),
          Activities('reject' ,87113 , 0),
          Activities('time-out 1',22 , 14),
          Activities('time-out 2',22 , 14),
          Activities('time-out 3',22 , 14),
          Activities('time-out X',86967 , 58),
        ]
```

Figure 32 - Review Example Large Dataset Custom GA Implementation result

```
number of generations: 0
time: 0.00040149688720703125s
best solution: ['accept', 'decide', 'get review 1', 'get review 2', 'get review 3', 'invite reviewers', 'reject']
```

- Number of events: 74969
- Number of unique cases: 10000
- The number of resources is: 10

Figure 33 - Review Example Large Dataset Custom GA EDA Activities

Figure 34 - Review Example Large Dataset Custom GA EDA Resources

Figure 35 - Review Example Large Dataset Custom GA BPMN Model

Figure 36 - Review Example Large Dataset Custom GA DFG

Figure 37 - Review Example Large Dataset Custom GA Heuristics

EasyGA

- Number of events: 60804
- Number of unique cases: 10000
- The number of resources is: 10

Figure 38 - Review Example Large Dataset EasyGA Implementation

```
Current Generation      : 10000
Chromosome - 0 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 1 [1][0][0][0][1][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 2 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 3 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 4 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 5 [0][1][1][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 6 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 7 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 8 [1][1][0][0][0][0][1][0][0][0][0][1][0][0] / Fitness = 100
Chromosome - 9 [1][1][0][0][0][0][1][0][0][0][0][1][0][0] / Fitness = 100
Chromosome - 10 [1][1][0][0][0][0][1][0][0][0][0][1][0][0] / Fitness = 100
Chromosome - 11 [1][0][0][0][1][0][1][0][0][0][0][1][0][0] / Fitness = 100
Chromosome - 12 [1][1][0][0][0][0][1][0][0][0][0][1][1][0] / Fitness = 100
Chromosome - 13 [1][1][0][0][0][0][0][0][0][0][0][1][0][0] / Fitness = 86
```

Figure 39 - Review Example Large Dataset EasyGA EDA Activities

Figure 40 - Review Example Large Dataset EasyGA EDA Resources

Figure 41 - Review Example Large Dataset EasyGA BPMN Model

Figure 42 - Review Example Large Dataset EasyGA DFG

Figure 43 - Review Example Large Dataset EasyGA Heuristics

EasyGA w/ BPMN Modelling

- Number of events: 55080
- Number of unique cases: 10000
- The number of resources is: 9

Figure 44 - Review Example Large Dataset EasyGA w/ BPMN Modelling implementation

Figure 45 - Review Example Large Dataset EasyGA w/ BPMN Modelling BPMN Model

Figure 46 - Review Example Large Dataset EasyGA w/ BPMN Modelling DFG

- Heuristics

Figure 47 - Review Example Large Dataset EasyGA w/ BPMN Modelling Heuristics

Helpdesk Dataset

Original

Figure 48 - Helpdesk Dataset EDA Activities

Figure 49 - Helpdesk Dataset EDA Resources

Figure 50 - Helpdesk Dataset BPMN Model

Figure 51 - Helpdesk Dataset DFG

Figure 52 - Helpdesk Dataset Heuristics

Social Network Analysis

Original Dataset 5 - Handover of Work

Figure 53 - Helpdesk Dataset Handover of Work

Every resource is connected and at some point, follows each other in sequence at some points in the event log. This is an assumption as it is hard to understand this complicated to understand all of resource relationships.

Original Dataset 6 - Subcontracting

Figure 54 - Helpdesk Dataset Subcontracting

Some people's work is only followed by others

Original Dataset 7 - Working Together

Figure 55 - Helpdesk Dataset Working Together

Original Dataset 8 - Similar Activity

Figure 56 - Helpdesk Dataset Similar Activity

Clustering

Clustered similar activities metric: {'1': ['Alex Haley', 'Ricardo Parker', 'James Rogers', 'Wendy Thompson', 'Richard Perry', 'Angela Kirk', 'Angelica Smith', 'John Nichols', 'Michelle Glenn', 'Kelly Williams', 'Martin Burton', 'Nicholas Fowler', 'Donna Pearson', 'Linda Brooks', 'James Cole', 'Mark Valencia'], '0': ['Bradley Acosta', 'Kristin Burton', 'Omar Collins', 'Jennifer Young', 'Olivia Henry', 'Megan Ortiz']}

Working together clustered metric: {'0': ['Alex Haley', 'Kristin Burton', 'Angelica Smith', 'Kelly Williams', 'Linda Brooks', 'Megan Ortiz'], '3': ['Ricardo Parker', 'Donna Pearson'], '1': ['James Rogers', 'Michelle Glenn', 'Omar Collins', 'Nicholas Fowler'], '2': ['Wendy Thompson', 'Angela Kirk', 'John Nichols', 'Jennifer Young', 'Martin Burton'], '4': ['Bradley Acosta', 'Richard Perry', 'Olivia Henry', 'Mark Valencia'], '5': ['James Cole']}

Organizational Mining

group_relative_focus

{'Kelly Williams': {'Assign seriousness': 0.04657756176589713, 'Take in charge ticket': 0.046442687747035576, 'Resolve ticket': 0.04435079269516356, 'Closed': 0.043506777437691296, 'Insert ticket': 0.025423728813559324, 'Wait': 0.04647983595352016, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.08403361344537816, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Angelica Smith': {'Assign seriousness': 0.04637505062778453, 'Take in charge ticket': 0.0458498023715415, 'Resolve ticket': 0.047561709813365445, 'Closed': 0.04459991254919108, 'Insert ticket': 0.06779661016949153, 'Wait': 0.0430622009569378, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.058823529411764705, 'VERIFIED': 0.0, 'DUPLICATE': 1.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'James Rogers': {'Assign seriousness': 0.04678007290400972, 'Take in charge ticket': 0.04426877470355731, 'Resolve ticket': 0.04314669877583785, 'Closed': 0.047879317883690424, 'Insert ticket': 0.025423728813559324, 'Wait': 0.047163362952836636, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.058823529411764705, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Martin Burton': {'Assign seriousness': 0.04374240583232078, 'Take in charge ticket': 0.04683794466403162, 'Resolve ticket': 0.045354204294601644, 'Closed': 0.04241364232619152, 'Insert ticket': 0.0423728813559322, 'Wait': 0.0430622009569378, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.025210084033613446, 'VERIFIED': 0.3333333333333333, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Bradley Acosta': {'Assign seriousness': 0.0479951397326853, 'Take in charge ticket': 0.04110671936758893, 'Resolve ticket': 0.04916716837246639, 'Closed': 0.045474420638390904, 'Insert ticket': 0.06779661016949153, 'Wait': 0.048530416951469584, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.025210084033613446, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.15384615384615385, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Richard Perry': {'Assign seriousness': 0.04637505062778453, 'Take in charge ticket': 0.0458498023715415, 'Resolve ticket': 0.0459562512542645, 'Closed': 0.047879317883690424, 'Insert ticket': 0.059322033898305086, 'Wait': 0.039644565960355434, 'Create SW anomaly': 0.07462686567164178, 'Require upgrade': 0.04201680672268908, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Ricardo Parker': {'Assign seriousness': 0.046172539489671933, 'Take in charge ticket': 0.04762845849802372, 'Resolve ticket': 0.048163756773028296, 'Closed': 0.045693047660690864, 'Insert ticket': 0.0423728813559322, 'Wait': 0.04032809295967191, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.04201680672268908, 'VERIFIED': 0.3333333333333333, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.2, 'RESOLVED': 0.0, 'INVALID': 0.5}, 'Megan Ortiz': {'Assign seriousness': 0.04171729445119481, 'Take in charge ticket': 0.04426877470355731, 'Resolve ticket': 0.04415011037527594, 'Closed': 0.04919108001749016, 'Insert ticket': 0.0423728813559322, 'Wait': 0.04647983595352016, 'Create SW anomaly': 0.04477611940298507, 'Require upgrade': 0.03361344537815126, 'VERIFIED': 0.0,

'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Mark Valencia': {'Assign seriousness': 0.043337383556095586, 'Take in charge ticket': 0.045454545454545456, 'Resolve ticket': 0.04475215733493879, 'Closed': 0.04022737210319195, 'Insert ticket': 0.03389830508474576, 'Wait': 0.04921394395078606, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.07563025210084033, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.5, 'INVALID': 0.0}, 'Nicholas Fowler': {'Assign seriousness': 0.04151478331308222, 'Take in charge ticket': 0.041304347826086954, 'Resolve ticket': 0.04896648605257877, 'Closed': 0.05115872321818977, 'Insert ticket': 0.03389830508474576, 'Wait': 0.050580997949419004, 'Create SW anomaly': 0.014925373134328358, 'Require upgrade': 0.01680672268907563, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.2, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Angela Kirk': {'Assign seriousness': 0.04374240583232078, 'Take in charge ticket': 0.04802371541501976, 'Resolve ticket': 0.04956853301224162, 'Closed': 0.043069523393091384, 'Insert ticket': 0.03389830508474576, 'Wait': 0.04784688995215311, 'Create SW anomaly': 0.04477611940298507, 'Require upgrade': 0.03361344537815126, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.2, 'RESOLVED': 0.0, 'INVALID': 0.5}, 'Linda Brooks': {'Assign seriousness': 0.04414742810854597, 'Take in charge ticket': 0.046442687747035576, 'Resolve ticket': 0.044551475015051176, 'Closed': 0.04481853957149104, 'Insert ticket': 0.059322033898305086, 'Wait': 0.03759398496240601, 'Create SW anomaly': 0.04477611940298507, 'Require upgrade': 0.025210084033613446, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.2, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Kristin Burton': {'Assign seriousness': 0.04495747266099635, 'Take in charge ticket': 0.049209486166007906, 'Resolve ticket': 0.040939193257074055, 'Closed': 0.04241364232619152, 'Insert ticket': 0.05084745762711865, 'Wait': 0.04237867395762133, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.07563025210084033, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.5, 'INVALID': 0.0}, 'Wendy Thompson': {'Assign seriousness': 0.04414742810854597, 'Take in charge ticket': 0.043478260869565216, 'Resolve ticket': 0.041541240216736906, 'Closed': 0.04525579361609095, 'Insert ticket': 0.01694915254237288, 'Wait': 0.03896103896103896, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.058823529411764705, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.2, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Olivia Henry': {'Assign seriousness': 0.0479951397326853, 'Take in charge ticket': 0.03873517786561265, 'Resolve ticket': 0.04776239213325306, 'Closed': 0.04525579361609095, 'Insert ticket': 0.03389830508474576, 'Wait': 0.043745727956254275, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.0, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'James Cole': {'Assign seriousness': 0.05062778452814905, 'Take in charge ticket': 0.05237154150197629, 'Resolve ticket': 0.0459562512542645, 'Closed': 0.04634892872759073, 'Insert ticket': 0.06779661016949153, 'Wait': 0.04921394395078606, 'Create SW anomaly': 0.07462686567164178, 'Require upgrade':

0.05042016806722689, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'John Nichols': {'Assign seriousness': 0.04759011745646011, 'Take in charge ticket': 0.04110671936758893, 'Resolve ticket': 0.041541240216736906, 'Closed': 0.04438128552689112, 'Insert ticket': 0.07627118644067797, 'Wait': 0.053998632946001365, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.008403361344537815, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Jennifer Young': {'Assign seriousness': 0.0479951397326853, 'Take in charge ticket': 0.042292490118577074, 'Resolve ticket': 0.04515352197471403, 'Closed': 0.04722343681679055, 'Insert ticket': 0.0423728813559322, 'Wait': 0.04032809295967191, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.058823529411764705, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Omar Collins': {'Assign seriousness': 0.044349939246658567, 'Take in charge ticket': 0.04387351778656127, 'Resolve ticket': 0.04776239213325306, 'Closed': 0.04438128552689112, 'Insert ticket': 0.05084745762711865, 'Wait': 0.04784688995215311, 'Create SW anomaly': 0.05970149253731343, 'Require upgrade': 0.08403361344537816, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Alex Haley': {'Assign seriousness': 0.04597002835155933, 'Take in charge ticket': 0.04940711462450593, 'Resolve ticket': 0.04394942805538832, 'Closed': 0.045474420638390904, 'Insert ticket': 0.03389830508474576, 'Wait': 0.045112781954887216, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.05042016806722689, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Donna Pearson': {'Assign seriousness': 0.043337383556095586, 'Take in charge ticket': 0.04881422924901186, 'Resolve ticket': 0.04856512141280353, 'Closed': 0.045474420638390904, 'Insert ticket': 0.06779661016949153, 'Wait': 0.045112781954887216, 'Create SW anomaly': 0.029850746268656716, 'Require upgrade': 0.025210084033613446, 'VERIFIED': 0.3333333333333333, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Michelle Glenn': {'Assign seriousness': 0.04455245038477116, 'Take in charge ticket': 0.047233201581027666, 'Resolve ticket': 0.04113987557696167, 'Closed': 0.047879317883690424, 'Insert ticket': 0.025423728813559324, 'Wait': 0.05331510594668489, 'Create SW anomaly': 0.04477611940298507, 'Require upgrade': 0.06722689075630252, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.07692307692307693, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}}

group_relative_stake

{'Kelly Williams': {'Assign seriousness': 0.23760330578512398, 'Take in charge ticket': 0.2427685950413223, 'Resolve ticket': 0.22830578512396693, 'Closed': 0.20557851239669422, 'Insert ticket': 0.0030991735537190084, 'Wait': 0.07024793388429752, 'Create SW anomaly':

0.002066115702479339, 'Require upgrade': 0.010330578512396695, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Angelica Smith': {'Assign seriousness': 0.23248730964467004, 'Take in charge ticket': 0.23553299492385787, 'Resolve ticket': 0.24060913705583756, 'Closed': 0.20710659898477157, 'Insert ticket': 0.008121827411167513, 'Wait': 0.06395939086294417, 'Create SW anomaly': 0.0040609137055837565, 'Require upgrade': 0.007106598984771574, 'VERIFIED': 0.0, 'DUPLICATE': 0.0010152284263959391, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'James Rogers': {'Assign seriousness': 0.23789907312049433, 'Take in charge ticket': 0.23069001029866118, 'Resolve ticket': 0.22142121524201855, 'Closed': 0.2255406797116375, 'Insert ticket': 0.003089598352214212, 'Wait': 0.07106076210092688, 'Create SW anomaly': 0.0020597322348094747, 'Require upgrade': 0.007209062821833162, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010298661174047373, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Martin Burton': {'Assign seriousness': 0.22736842105263158, 'Take in charge ticket': 0.24947368421052632, 'Resolve ticket': 0.23789473684210527, 'Closed': 0.20421052631578948, 'Insert ticket': 0.005263157894736842, 'Wait': 0.06631578947368422, 'Create SW anomaly': 0.004210526315789474, 'Require upgrade': 0.003157894736842105, 'VERIFIED': 0.0010526315789473684, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010526315789473684, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Bradley Acosta': {'Assign seriousness': 0.24036511156186613, 'Take in charge ticket': 0.21095334685598377, 'Resolve ticket': 0.2484787018255578, 'Closed': 0.21095334685598377, 'Insert ticket': 0.008113590263691683, 'Wait': 0.0720081135902637, 'Create SW anomaly': 0.004056795131845842, 'Require upgrade': 0.0030425963488843813, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.002028397565922921, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Richard Perry': {'Assign seriousness': 0.23272357723577236, 'Take in charge ticket': 0.23577235772357724, 'Resolve ticket': 0.23272357723577236, 'Closed': 0.2225609756097561, 'Insert ticket': 0.007113821138211382, 'Wait': 0.05894308943089431, 'Create SW anomaly': 0.00508130081300813, 'Require upgrade': 0.00508130081300813, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Ricardo Parker': {'Assign seriousness': 0.229607250755287, 'Take in charge ticket': 0.24269889224572003, 'Resolve ticket': 0.24169184290030213, 'Closed': 0.2104733131923464, 'Insert ticket': 0.005035246727089627, 'Wait': 0.059415911379657606, 'Create SW anomaly': 0.002014098690835851, 'Require upgrade': 0.005035246727089627, 'VERIFIED': 0.0010070493454179255, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010070493454179255, 'Schedule intervention': 0.0010070493454179255, 'RESOLVED': 0.0, 'INVALID': 0.0010070493454179255}, 'Megan Ortiz': {'Assign seriousness': 0.2157068062827225, 'Take in charge ticket': 0.23455497382198953, 'Resolve ticket': 0.23036649214659685, 'Closed': 0.2356020942408377, 'Insert ticket': 0.005235602094240838, 'Wait': 0.0712041884816754, 'Create SW anomaly': 0.0031413612565445027, 'Require upgrade': 0.004188481675392671, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Mark Valencia': {'Assign seriousness': 0.22790202342917998, 'Take in charge ticket': 0.24494142705005326, 'Resolve

ticket': 0.2374866879659212, 'Closed': 0.1959531416400426, 'Insert ticket': 0.004259850905218318, 'Wait': 0.07667731629392971, 'Create SW anomaly': 0.002129925452609159, 'Require upgrade': 0.009584664536741214, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0010649627263045794, 'INVALID': 0.0}, 'Nicholas Fowler': {'Assign seriousness': 0.2104722792607803, 'Take in charge ticket': 0.21457905544147843, 'Resolve ticket': 0.25051334702258726, 'Closed': 0.2402464065708419, 'Insert ticket': 0.004106776180698152, 'Wait': 0.07597535934291581, 'Create SW anomaly': 0.001026694045174538, 'Require upgrade': 0.002053388090349076, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.001026694045174538, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Angela Kirk': {'Assign seriousness': 0.2188449848024316, 'Take in charge ticket': 0.24620060790273557, 'Resolve ticket': 0.2502532928064843, 'Closed': 0.19959473150962512, 'Insert ticket': 0.004052684903748734, 'Wait': 0.07092198581560284, 'Create SW anomaly': 0.00303951367781155, 'Require upgrade': 0.004052684903748734, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010131712259371835, 'Schedule intervention': 0.0010131712259371835, 'RESOLVED': 0.0, 'INVALID': 0.0010131712259371835}, 'Linda Brooks': {'Assign seriousness': 0.2297154899894626, 'Take in charge ticket': 0.2476290832455216, 'Resolve ticket': 0.2339304531085353, 'Closed': 0.2160168598524763, 'Insert ticket': 0.007376185458377239, 'Wait': 0.05795574288724974, 'Create SW anomaly': 0.003161222339304531, 'Require upgrade': 0.003161222339304531, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.001053740779768177, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Kristin Burton': {'Assign seriousness': 0.2339304531085353, 'Take in charge ticket': 0.2623814541622761, 'Resolve ticket': 0.21496311907270813, 'Closed': 0.20442571127502634, 'Insert ticket': 0.006322444678609062, 'Wait': 0.06533192834562697, 'Create SW anomaly': 0.002107481559536354, 'Require upgrade': 0.009483667017913594, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.001053740779768177, 'INVALID': 0.0}, 'Wendy Thompson': {'Assign seriousness': 0.23593073593073594, 'Take in charge ticket': 0.23809523809523808, 'Resolve ticket': 0.22402597402597402, 'Closed': 0.22402597402597402, 'Insert ticket': 0.0021645021645021645, 'Wait': 0.06168831168831169, 'Create SW anomaly': 0.004329004329004329, 'Require upgrade': 0.007575757575757576, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010822510822510823, 'Schedule intervention': 0.0010822510822510823, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Olivia Henry': {'Assign seriousness': 0.24921135646687698, 'Take in charge ticket': 0.20609884332281808, 'Resolve ticket': 0.2502628811777077, 'Closed': 0.21766561514195584, 'Insert ticket': 0.004206098843322818, 'Wait': 0.06729758149316509, 'Create SW anomaly': 0.004206098843322818, 'Require upgrade': 0.0, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010515247108307045, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'James Cole': {'Assign seriousness': 0.2385496183206107, 'Take in charge ticket': 0.25286259541984735, 'Resolve ticket': 0.21851145038167938, 'Closed': 0.20229007633587787, 'Insert ticket': 0.007633587786259542, 'Wait': 0.06870229007633588, 'Create SW anomaly': 0.004770992366412214, 'Require upgrade': 0.0057251908396946565,

'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0009541984732824427, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'John Nichols': {'Assign seriousness': 0.2489406779661017, 'Take in charge ticket': 0.22033898305084745, 'Resolve ticket': 0.21927966101694915, 'Closed': 0.21504237288135594, 'Insert ticket': 0.009533898305084746, 'Wait': 0.0836864406779661, 'Create SW anomaly': 0.00211864406779661, 'Require upgrade': 0.001059322033898305, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Jennifer Young': {'Assign seriousness': 0.2450879007238883, 'Take in charge ticket': 0.22130299896587383, 'Resolve ticket': 0.2326783867631851, 'Closed': 0.2233712512926577, 'Insert ticket': 0.005170630816959669, 'Wait': 0.06101344364012409, 'Create SW anomaly': 0.004136504653567736, 'Require upgrade': 0.007238883143743537, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Omar Collins': {'Assign seriousness': 0.2250770811921891, 'Take in charge ticket': 0.22816032887975335, 'Resolve ticket': 0.2446043165467626, 'Closed': 0.20863309352517986, 'Insert ticket': 0.006166495375128468, 'Wait': 0.07194244604316546, 'Create SW anomaly': 0.0041109969167523125, 'Require upgrade': 0.010277492291880781, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010277492291880781, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Alex Haley': {'Assign seriousness': 0.23092573753814852, 'Take in charge ticket': 0.254323499491353, 'Resolve ticket': 0.22278738555442523, 'Closed': 0.2115971515768057, 'Insert ticket': 0.004069175991861648, 'Wait': 0.0671414038657172, 'Create SW anomaly': 0.002034587995930824, 'Require upgrade': 0.006103763987792472, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.001017293997965412, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Donna Pearson': {'Assign seriousness': 0.2157258064516129, 'Take in charge ticket': 0.24899193548387097, 'Resolve ticket': 0.2439516129032258, 'Closed': 0.20967741935483872, 'Insert ticket': 0.008064516129032258, 'Wait': 0.06653225806451613, 'Create SW anomaly': 0.0020161290322580645, 'Require upgrade': 0.0030241935483870967, 'VERIFIED': 0.0010080645161290322, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010080645161290322, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}, 'Michelle Glenn': {'Assign seriousness': 0.22540983606557377, 'Take in charge ticket': 0.24487704918032788, 'Resolve ticket': 0.21004098360655737, 'Closed': 0.22438524590163936, 'Insert ticket': 0.0030737704918032786, 'Wait': 0.07991803278688525, 'Create SW anomaly': 0.0030737704918032786, 'Require upgrade': 0.00819672131147541, 'VERIFIED': 0.0, 'DUPLICATE': 0.0, 'Resolve SW anomaly': 0.0010245901639344263, 'Schedule intervention': 0.0, 'RESOLVED': 0.0, 'INVALID': 0.0}}

group_coverage

{'Kelly Williams': {'Kelly Williams': 1.0}, 'Angelica Smith': {'Angelica Smith': 1.0}, 'James Rogers': {'James Rogers': 1.0}, 'Martin Burton': {'Martin Burton': 1.0}, 'Bradley Acosta': {'Bradley Acosta': 1.0}, 'Richard Perry': {'Richard Perry': 1.0}, 'Ricardo Parker': {'Ricardo Parker': 1.0}, 'Megan Ortiz': {'Megan Ortiz': 1.0}, 'Mark Valencia': {'Mark Valencia': 1.0},

'Nicholas Fowler': {'Nicholas Fowler': 1.0}, 'Angela Kirk': {'Angela Kirk': 1.0}, 'Linda Brooks': {'Linda Brooks': 1.0}, 'Kristin Burton': {'Kristin Burton': 1.0}, 'Wendy Thompson': {'Wendy Thompson': 1.0}, 'Olivia Henry': {'Olivia Henry': 1.0}, 'James Cole': {'James Cole': 1.0}, 'John Nichols': {'John Nichols': 1.0}, 'Jennifer Young': {'Jennifer Young': 1.0}, 'Omar Collins': {'Omar Collins': 1.0}, 'Alex Haley': {'Alex Haley': 1.0}, 'Donna Pearson': {'Donna Pearson': 1.0}, 'Michelle Glenn': {'Michelle Glenn': 1.0}}

group_member_contribution

{'Kelly Williams': {'Kelly Williams': {'Assign seriousness': 230, 'Closed': 199, 'Take in charge ticket': 235, 'Resolve ticket': 221, 'Wait': 68, 'Require upgrade': 10, 'Create SW anomaly': 2, 'Insert ticket': 3}}, 'Angelica Smith': {'Angelica Smith': {'Take in charge ticket': 232, 'Assign seriousness': 229, 'Closed': 204, 'Create SW anomaly': 4, 'Wait': 63, 'Resolve ticket': 237, 'Insert ticket': 8, 'Require upgrade': 7, 'DUPLICATE': 1}}, 'James Rogers': {'James Rogers': {'Take in charge ticket': 224, 'Closed': 219, 'Resolve ticket': 215, 'Assign seriousness': 231, 'Require upgrade': 7, 'Wait': 69, 'Resolve SW anomaly': 1, 'Insert ticket': 3, 'Create SW anomaly': 2}}, 'Martin Burton': {'Martin Burton': {'Resolve ticket': 226, 'Take in charge ticket': 237, 'Closed': 194, 'Assign seriousness': 216, 'Wait': 63, 'VERIFIED': 1, 'Insert ticket': 5, 'Create SW anomaly': 4, 'Require upgrade': 3, 'Resolve SW anomaly': 1}}, 'Bradley Acosta': {'Bradley Acosta': {'Assign seriousness': 237, 'Take in charge ticket': 208, 'Closed': 208, 'Resolve ticket': 245, 'Insert ticket': 8, 'Wait': 71, 'Create SW anomaly': 4, 'Require upgrade': 3, 'Resolve SW anomaly': 2}}, 'Richard Perry': {'Richard Perry': {'Resolve ticket': 229, 'Assign seriousness': 229, 'Wait': 58, 'Require upgrade': 5, 'Take in charge ticket': 232, 'Closed': 219, 'Create SW anomaly': 5, 'Insert ticket': 7}}, 'Ricardo Parker': {'Ricardo Parker': {'Closed': 209, 'Resolve ticket': 240, 'Assign seriousness': 228, 'Take in charge ticket': 241, 'Wait': 59, 'Insert ticket': 5, 'Require upgrade': 5, 'INVALID': 1, 'VERIFIED': 1, 'Create SW anomaly': 2, 'Resolve SW anomaly': 1, 'Schedule intervention': 1}}, 'Megan Ortiz': {'Megan Ortiz': {'Take in charge ticket': 224, 'Resolve ticket': 220, 'Closed': 225, 'Assign seriousness': 206, 'Create SW anomaly': 3, 'Wait': 68, 'Insert ticket': 5, 'Require upgrade': 4}}, 'Mark Valencia': {'Mark Valencia': {'Resolve ticket': 223, 'Assign seriousness': 214, 'Take in charge ticket': 230, 'Wait': 72, 'Closed': 184, 'Require upgrade': 9, 'Insert ticket': 4, 'Create SW anomaly': 2, 'RESOLVED': 1}}, 'Nicholas Fowler': {'Nicholas Fowler': {'Closed': 234, 'Assign seriousness': 205, 'Take in charge ticket': 209, 'Resolve ticket': 244, 'Insert ticket': 4, 'Wait': 74, 'Create SW anomaly': 1, 'Schedule intervention': 1, 'Require upgrade': 2}}, 'Angela Kirk': {'Angela Kirk': {'Resolve ticket': 247, 'Closed': 197, 'Assign seriousness': 216, 'Take in charge ticket': 243, 'Wait': 70, 'Require upgrade': 4, 'Create SW anomaly': 3, 'Resolve SW anomaly': 1, 'Insert ticket': 4, 'Schedule intervention': 1, 'INVALID': 1}}, 'Linda Brooks': {'Linda Brooks': {'Assign seriousness': 218, 'Take in charge ticket': 235, 'Resolve ticket': 222, 'Closed': 205, 'Wait': 55, 'Create SW anomaly': 3, 'Require upgrade': 3, 'Insert ticket': 7, 'Schedule intervention': 1}}, 'Kristin Burton': {'Kristin Burton': {'Take in charge ticket': 249, 'Closed': 194, 'Resolve ticket': 204, 'Require upgrade': 9, 'Assign seriousness': 222, 'Insert ticket': 6, 'Wait': 62, 'Create SW anomaly': 2, 'RESOLVED': 1}}, 'Wendy Thompson': {'Wendy Thompson': {'Take in charge ticket': 220, 'Closed': 207, 'Assign seriousness': 218,

'Resolve ticket': 207, 'Require upgrade': 7, 'Wait': 57, 'Schedule intervention': 1, 'Insert ticket': 2, 'Create SW anomaly': 4, 'Resolve SW anomaly': 1}}, 'Olivia Henry': {'Olivia Henry': {'Resolve ticket': 238, 'Take in charge ticket': 196, 'Assign seriousness': 237, 'Closed': 207, 'Wait': 64, 'Insert ticket': 4, 'Create SW anomaly': 4, 'Resolve SW anomaly': 1}}, 'James Cole': {'James Cole': {'Take in charge ticket': 265, 'Insert ticket': 8, 'Resolve ticket': 229, 'Wait': 72, 'Assign seriousness': 250, 'Closed': 212, 'Create SW anomaly': 5, 'Require upgrade': 6, 'Resolve SW anomaly': 1}}, 'John Nichols': {'John Nichols': {'Assign seriousness': 235, 'Take in charge ticket': 208, 'Resolve ticket': 207, 'Closed': 203, 'Wait': 79, 'Insert ticket': 9, 'Require upgrade': 1, 'Create SW anomaly': 2}}, 'Jennifer Young': {'Jennifer Young': {'Closed': 216, 'Take in charge ticket': 214, 'Assign seriousness': 237, 'Resolve ticket': 225, 'Wait': 59, 'Insert ticket': 5, 'Require upgrade': 7, 'Create SW anomaly': 4}}, 'Omar Collins': {'Omar Collins': {'Assign seriousness': 219, 'Take in charge ticket': 222, 'Resolve ticket': 238, 'Create SW anomaly': 4, 'Closed': 203, 'Wait': 70, 'Insert ticket': 6, 'Require upgrade': 10, 'Resolve SW anomaly': 1}}, 'Alex Haley': {'Alex Haley': {'Take in charge ticket': 250, 'Closed': 208, 'Assign seriousness': 227, 'Resolve ticket': 219, 'Require upgrade': 6, 'Wait': 66, 'Resolve SW anomaly': 1, 'Create SW anomaly': 2, 'Insert ticket': 4}}, 'Donna Pearson': {'Donna Pearson': {'Resolve ticket': 242, 'Take in charge ticket': 247, 'Assign seriousness': 214, 'Closed': 208, 'Wait': 66, 'Insert ticket': 8, 'Create SW anomaly': 2, 'Require upgrade': 3, 'Resolve SW anomaly': 1, 'VERIFIED': 1}}, 'Michelle Glenn': {'Michelle Glenn': {'Take in charge ticket': 239, 'Closed': 219, 'Resolve ticket': 205, 'Assign seriousness': 220, 'Wait': 78, 'Insert ticket': 3, 'Require upgrade': 8, 'Create SW anomaly': 3, 'Resolve SW anomaly': 1}}}

We can determine the percentage that each activity occupies in the process. In total, the activities that have a close to 0% are Insert Tickets, Schedule Intervention, Create SW Anomaly, Resolve Anomaly, Resolved, Invalid, Verified, and Duplicate tasks. The Require Upgrade task contributes to 1%. The Closed task contributes to 21%. The Resolve Ticket contributes 23 %. Finally, the Assign Seriousness and Take in Charge Ticket contributes 24%.

CPM

An interesting observation is that the activities chosen by the CPM are the ones that have a 0% in their contribution to the process. This proves the useful and trustworthy of CPM staying true to its purpose of outputting the most critical path. However, it might not make sense in there what the activities are.

- Number of events: 4718
- Number of unique cases: 4565
- The number of resources is: 22
- Total duration: 257 days

Figure 57 - Helpdesk Dataset CPM EDA Activities

Figure 58 - Helpdesk Dataset CPM EDA Resources

Figure 59 - Helpdesk Dataset CPM BPMN Model

Figure 60 - Helpdesk Dataset CPM DFG

Figure 61 - Helpdesk Dataset CPM Heuristic

Custom GA

Figure 62 - Helpdesk Dataset Custom GA Implementation

```
[ ] #Activities = namedtuple( 'Activities', ['name','aggregated_max_relative_day', 'weight'])
    Helpdesk_activities = namedtuple( 'Helpdesk_activities', ['name','days_to_finish', 'percentage'])

    max_days = 257

    things = [


              Helpdesk_activities('Insert Ticket',0 , 0),
              Helpdesk_activities('Assign Seriousness',44 , 24),
              Helpdesk_activities('Take in charge ticket',48 , 24),
              Helpdesk_activities('Schedule intervention',11 , 0),
              Helpdesk_activities('Create SW anomaly',55 , 0),
              Helpdesk_activities('Resolve SW anomaly',37 , 0),
              Helpdesk_activities('Wait',48 , 7),
              Helpdesk_activities('Require upgrade',55 , 1),
              Helpdesk_activities('Resolve ticket',58 , 23),
              Helpdesk_activities('RESOLVED' ,34 , 0),
              Helpdesk_activities('INVALID',34 , 0),
              Helpdesk_activities('VERIFIED',56 , 0),
              Helpdesk_activities('DUPLICATE',7 , 0),
              Helpdesk_activities('Closed',59 , 21),
            ]
```

Figure 63 Helpdesk Dataset Custom GA Implementation result

```
number of generations: 0
time: 0.0004467964172363281s
best solution: ['Assign Seriousness', 'Take in charge ticket', 'Create SW anomaly', 'Wait', 'Require upgrade', 'RESOLVED', 'INVALID', 'VERIFIED', 'Closed']
```

- Number of events: 11290
- Number of unique cases: 4580
- The number of resources is: 22

Figure 64 - Helpdesk Dataset Custom GA EDA Activities

Figure 65 - Helpdesk Dataset Custom GA EDA Resources

Figure 66 - Helpdesk Dataset Custom GA BPMN Model

Figure 67 - Helpdesk Dataset Custom GA DFG

Figure 68 - Helpdesk Dataset Custom GA Heuristic

EasyGA

- Number of events: 14941
- Number of unique cases: 4580
- The number of resources is: 22

Figure 69 - Helpdesk Dataset EasyGA Implementation

```
Current Generation      : 10000
Chromosome - 0 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 1 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 2 [0][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 3 [0][1][1][Create SW anomaly][0][Wait][1][INVALID][1] / Fitness = 99
Chromosome - 4 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 5 [0][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 6 [1][1][1][Create SW anomaly][1][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 7 [0][1][1][Create SW anomaly][1][Wait][1][INVALID][0] / Fitness = 78
Chromosome - 8 [0][1][1][Create SW anomaly][0][Wait][0][VERIFIED][1] / Fitness = 76
```

Figure 70 - Helpdesk Dataset EasyGA EDA Activities

Figure 71 - Helpdesk Dataset EasyGA EDA Resources

Figure 72 - Helpdesk Dataset EasyGA BPMN Model

Insert ticket
Closed
Resolve SW anomaly
Schedule intervention
Require upgrade
Resolve ticket
Take in charge ticket
Create SW anomaly

Figure 73 - Helpdesk Dataset EasyGA DFG

Figure 74 - Helpdesk Dataset EasyGA Heuristics

EasyGA w/ BPMN Modelling

- Number of events: 14810
- Number of unique cases: 4580
- The number of resources is: 22

Figure 75 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model Implementation

```
Current Generation      : 10000
Chromosome - 0 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 1 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 2 [0][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 3 [0][1][1][Create SW anomaly][0][Wait][1][INVALID][1] / Fitness = 99
Chromosome - 4 [1][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 5 [0][1][1][Create SW anomaly][0][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 6 [1][1][1][Create SW anomaly][1][Wait][1][VERIFIED][1] / Fitness = 99
Chromosome - 7 [0][1][1][Create SW anomaly][1][Wait][1][INVALID][0] / Fitness = 78
Chromosome - 8 [0][1][1][Create SW anomaly][0][Wait][0][VERIFIED][1] / Fitness = 76
```

Figure 76 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model EDA Activities

Figure 77 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model EDA Resources

Figure 78 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model BPMN Model

Figure 79 - Helpdesk Dataset EasyGA w/ BPMN Modelling DFG

Figure 80 - Helpdesk Dataset EasyGA w/ BPMN Modelling A BPMN Model EDA Activities

BPI 2012 Dataset

Original

Figure 81 – BPI 2012 Dataset EDA Activities

Figure 82 – BPI 2012 Dataset EDA Resources

Figure 83 – BPI 2012 Dataset BPMN Model

Figure 84 – BPI 2012 Dataset DFG

Figure 85 – BPI 2012 Dataset Heuristics

Social Network Analysis

Original Dataset 9 - Handover of Work

Figure 86 – BPI 2012 Dataset EDA Handover of Work

Original Dataset 10 - Subcontracting

Figure 87 – BPI 2012 Dataset EDA Subcontracting

Original Dataset 11 - Working Together

Figure 88 – BPI 2012 Dataset EDA Working Together

Original Dataset 12 - Similar Activity

Figure 89 – BPI 2012 Dataset EDA Similar Activity

Clustering

clustered similar activities metric: {'1': ['Derrick Watts', 'Christina Barnes', 'Jennifer Wilson', 'Heather Patterson', 'Brianna Bell', 'Joshua Phelps', 'Justin Thompson', 'Travis Edwards', 'Scott Little', 'Steven Jones', 'Jacob Wilson', 'Jane Turner'], '2': ['Patrick Cole', 'Lisa Henson', 'Donald Wilson', 'Rebecca Guerrero', 'Matthew Walker', 'William Arroyo', 'Nancy Clayton', 'Robert Peters', 'Kathy Ingram', 'Brianna Martinez', 'Anna Schroeder', 'Jennifer Collier', 'Nathan Thomas', 'Vanessa Shepherd', 'Luis Jackson', 'Michelle Brennan', 'Ruth Davies', 'Donald White', 'Emily Rojas', 'Aaron Ray', 'Kevin Schultz', 'Shawn Barnes', 'Paul Alvarado', 'Michael Anderson MD', 'Sherry Klein', 'Patricia Lloyd', 'Miguel Reeves', 'Austin Jones'], '0': ['Kayla Woods', 'Martha Costa', 'Laura Morris', 'Brandon Powell', 'Ashley Daniel', 'David Doyle', 'Jennifer Thomas MD', 'Jesse Moore', 'Brittany Valencia', 'Pam Martinez', 'Sarah Hogan', 'Wesley Mills', 'David Adams', 'Emily Rice', 'Rhonda Brady', 'Mary Thompson']}

working together clustered metric: {'3': ['Derrick Watts', 'Lisa Henson', 'Martha Costa', 'Justin Thompson', 'Brittany Valencia', 'Jennifer Collier', 'Shawn Barnes'], '6': ['Patrick Cole', 'Kayla Woods', 'David Doyle', 'Sarah Hogan', 'Ruth Davies', 'Aaron Ray'], '8': ['Christina Barnes', 'Brianna Bell', 'Jesse Moore', 'Anna Schroeder', 'Luis Jackson', 'Michelle Brennan', 'Donald White', 'Kevin Schultz', 'Rhonda Brady', 'Mary Thompson', 'Michael Anderson MD', 'Patricia Lloyd'], '1': ['Donald Wilson', 'Nancy Clayton', 'Travis Edwards', 'Miguel Reeves'], '0': ['Rebecca Guerrero',

'Matthew Walker', 'William Arroyo', 'Heather Patterson', 'Jennifer Thomas MD', 'Pam Martinez'], '5': ['Laura Morris', 'Kathy Ingram', 'David Adams', 'Sherry Klein', 'Jane Turner'], '7': ['Brandon Powell', 'Emily Rojas', 'Austin Jones'], '2': ['Jennifer Wilson', 'Robert Peters', 'Joshua Phelps', 'Scott Little', 'Vanessa Shepherd', 'Paul Alvarado', 'Steven Jones'], '4': ['Ashley Daniel', 'Brianna Martinez', 'Nathan Thomas', 'Wesley Mills', 'Emily Rice', 'Jacob Wilson']}

Organizational Mining

group_relative_focus

{'Jennifer Thomas MD': {'W_Completeren aanvraag': 0.018608920599157173, 'W_Nabellen offertes': 0.01841051532033426, 'W_Valideren aanvraag': 0.019379354021532614, 'W_Afhandelen leads': 0.01865038996269922, 'W_Nabellen incomplete dossiers': 0.018848075742964845, 'W_Beoordelen fraude': 0.011111111111111112}, 'Robert Peters': {'W_Completeren aanvraag': 0.01568823799390829, 'W_Nabellen offertes': 0.01745299442896936, 'W_Valideren aanvraag': 0.01773274224192527, 'W_Afhandelen leads': 0.016615801966768397, 'W_Nabellen incomplete dossiers': 0.017182431840098187, 'W_Beoordelen fraude': 0.022222222222222223}, 'Christina Barnes': {'W_Completeren aanvraag': 0.018567196561939333, 'W_Nabellen offertes': 0.017104805013927575, 'W_Valideren aanvraag': 0.018872704243191894, 'W_Afhandelen leads': 0.018480840963038318, 'W_Nabellen incomplete dossiers': 0.01630577715437889, 'W_Beoordelen fraude': 0.014814814814814815}, 'David Adams': {'W_Completeren aanvraag': 0.017524095631493302, 'W_Nabellen offertes': 0.01749651810584958, 'W_Valideren aanvraag': 0.016592780240658644, 'W_Afhandelen leads': 0.020176330959647337, 'W_Nabellen incomplete dossiers': 0.018059086525817482, 'W_Beoordelen fraude': 0.025925925925925925}, 'William Arroyo': {'W_Completeren aanvraag': 0.016814786998790004, 'W_Nabellen offertes': 0.018149373259052925, 'W_Valideren aanvraag': 0.016339455351488282, 'W_Afhandelen leads': 0.018480840963038318, 'W_Nabellen incomplete dossiers': 0.015253791531515735, 'W_Beoordelen fraude': 0.025925925925925925}, 'Anna Schroeder': {'W_Completeren aanvraag': 0.016981683147661367, 'W_Nabellen offertes': 0.018932799442896935, 'W_Valideren aanvraag': 0.01823939202026599, 'W_Afhandelen leads': 0.01441166497117667, 'W_Nabellen incomplete dossiers': 0.017796090120101693, 'W_Beoordelen fraude': 0.037037037037037035}, 'Brianna Bell': {'W_Completeren aanvraag': 0.017941336003671714, 'W_Nabellen offertes': 0.01688718662952646, 'W_Valideren aanvraag': 0.015579480683977201, 'W_Afhandelen leads': 0.01593760596812479, 'W_Nabellen incomplete dossiers': 0.01665643902866661, 'W_Beoordelen fraude': 0.022222222222222223}, 'Brandon Powell': {'W_Completeren aanvraag': 0.019276505194642635, 'W_Nabellen offertes': 0.017278899721448467, 'W_Valideren aanvraag': 0.01735275490816973, 'W_Afhandelen leads': 0.014750762970498474, 'W_Nabellen incomplete dossiers': 0.019023406680108705, 'W_Beoordelen fraude': 0.018518518518518517}, 'Aaron Ray': {'W_Completeren aanvraag': 0.01764926774314683, 'W_Nabellen offertes': 0.017844707520891366, 'W_Valideren aanvraag': 0.01823939202026599, 'W_Afhandelen leads': 0.015768056968463885, 'W_Nabellen incomplete dossiers': 0.019023406680108705, 'W_Beoordelen fraude': 0.007407407407407408}, 'Scott Little': {'W_Completeren aanvraag': 0.01915133308298911, 'W_Nabellen offertes': 0.018758704735376046, 'W_Valideren aanvraag':

0.019886003799873337, 'W_Afhandelen leads': 0.0167853509664293, 'W_Nabellen incomplete dossiers': 0.016042780748663103, 'W_Beoordelen fraude': 0.014814814814814815}, 'Ashley Daniel': {'W_Completeren aanvraag': 0.018191680226978764, 'W_Nabellen offertes': 0.017757660167130918, 'W_Valideren aanvraag': 0.01773274224192527, 'W_Afhandelen leads': 0.0171244489657511, 'W_Nabellen incomplete dossiers': 0.019111072148680634, 'W_Beoordelen fraude': 0.014814814814814815}, 'Lisa Henson': {'W_Completeren aanvraag': 0.01677306296157216, 'W_Nabellen offertes': 0.018628133704735376, 'W_Valideren aanvraag': 0.018746041798606713, 'W_Afhandelen leads': 0.015089860969820278, 'W_Nabellen incomplete dossiers': 0.017007100902954327, 'W_Beoordelen fraude': 0.014814814814814815}, 'Jennifer Wilson': {'W_Completeren aanvraag': 0.01677306296157216, 'W_Nabellen offertes': 0.019106894150417827, 'W_Valideren aanvraag': 0.01773274224192527, 'W_Afhandelen leads': 0.018141742963716515, 'W_Nabellen incomplete dossiers': 0.019198737617252565, 'W_Beoordelen fraude': 0.022222222222222223}, 'Jane Turner': {'W_Completeren aanvraag': 0.01906788500855343, 'W_Nabellen offertes': 0.018105849582172703, 'W_Valideren aanvraag': 0.018746041798606713, 'W_Afhandelen leads': 0.015259409969481181, 'W_Nabellen incomplete dossiers': 0.01472779872008416, 'W_Beoordelen fraude': 0.007407407407407408}, 'Brianna Martinez': {'W_Completeren aanvraag': 0.01677306296157216, 'W_Nabellen offertes': 0.018584610027855154, 'W_Valideren aanvraag': 0.017859404686510448, 'W_Afhandelen leads': 0.018819938962360123, 'W_Nabellen incomplete dossiers': 0.018935741211536777, 'W_Beoordelen fraude': 0.003703703703703704}, 'Laura Morris': {'W_Completeren aanvraag': 0.01702340718487921, 'W_Nabellen offertes': 0.019368036211699163, 'W_Valideren aanvraag': 0.016719442685243825, 'W_Afhandelen leads': 0.013394370973211257, 'W_Nabellen incomplete dossiers': 0.01595511528009117, 'W_Beoordelen fraude': 0.014814814814814815}, 'Wesley Mills': {'W_Completeren aanvraag': 0.01810823215254308, 'W_Nabellen offertes': 0.017714136490250696, 'W_Valideren aanvraag': 0.01811272957568081, 'W_Afhandelen leads': 0.019328585961342827, 'W_Nabellen incomplete dossiers': 0.017445428245813972, 'W_Beoordelen fraude': 0.011111111111111112}, 'Brittany Valencia': {'W_Completeren aanvraag': 0.01769099178036467, 'W_Nabellen offertes': 0.016582520891364902, 'W_Valideren aanvraag': 0.018492716909436352, 'W_Afhandelen leads': 0.01797219396405561, 'W_Nabellen incomplete dossiers': 0.01823441746296134, 'W_Beoordelen fraude': 0.014814814814814815}, 'Rhonda Brady': {'W_Completeren aanvraag': 0.01739892351983978, 'W_Nabellen offertes': 0.016626044568245124, 'W_Valideren aanvraag': 0.01823939202026599, 'W_Afhandelen leads': 0.02102407595795185, 'W_Nabellen incomplete dossiers': 0.01595511528009117, 'W_Beoordelen fraude': 0.025925925925925925}, 'Emily Rice': {'W_Completeren aanvraag': 0.0187340927108107, 'W_Nabellen offertes': 0.017278899721448467, 'W_Valideren aanvraag': 0.019632678910702975, 'W_Afhandelen leads': 0.018311291963377416, 'W_Nabellen incomplete dossiers': 0.02033838870868765, 'W_Beoordelen fraude': 0.018518518518518517}, 'Patrick Cole': {'W_Completeren aanvraag': 0.01777443985480035, 'W_Nabellen offertes': 0.018149373259052925, 'W_Valideren aanvraag': 0.017226092463584548, 'W_Afhandelen leads': 0.019498134961003732, 'W_Nabellen incomplete dossiers': 0.018497413868677128, 'W_Beoordelen fraude': 0.022222222222222223}, 'Paul Alvarado': {'W_Completeren aanvraag':

0.01814995618976092, 'W_Nabellen offertes': 0.01854108635097493, 'W_Valideren aanvraag': 0.018872704243191894, 'W_Afhandelen leads': 0.019667683960664633, 'W_Nabellen incomplete dossiers': 0.016831769965810466, 'W_Beoordelen fraude': 0.014814814814814815}, 'Justin Thompson': {'W_Completeren aanvraag': 0.019610297492385364, 'W_Nabellen offertes': 0.017365947075208915, 'W_Valideren aanvraag': 0.016972767574414186, 'W_Afhandelen leads': 0.0167853509664293, 'W_Nabellen incomplete dossiers': 0.01595511528009117, 'W_Beoordelen fraude': 0.014814814814814815}, 'Vanessa Shepherd': {'W_Completeren aanvraag': 0.017941336003671714, 'W_Nabellen offertes': 0.018323467966573817, 'W_Valideren aanvraag': 0.017099430018999367, 'W_Afhandelen leads': 0.019667683960664633, 'W_Nabellen incomplete dossiers': 0.01814675199438941, 'W_Beoordelen fraude': 0.003703703703703704}, 'Matthew Walker': {'W_Completeren aanvraag': 0.016480994701047275, 'W_Nabellen offertes': 0.01780118384401114, 'W_Valideren aanvraag': 0.01811272957568081, 'W_Afhandelen leads': 0.018989487962021025, 'W_Nabellen incomplete dossiers': 0.01797142105724555, 'W_Beoordelen fraude': 0.03333333333333333}, 'Miguel Reeves': {'W_Completeren aanvraag': 0.01840030041306797, 'W_Nabellen offertes': 0.018584610027855154, 'W_Valideren aanvraag': 0.016846105129829005, 'W_Afhandelen leads': 0.0167853509664293, 'W_Nabellen incomplete dossiers': 0.01770842465152976, 'W_Beoordelen fraude': 0.007407407407407408}, 'Kathy Ingram': {'W_Completeren aanvraag': 0.017732715817582508, 'W_Nabellen offertes': 0.018454038997214484, 'W_Valideren aanvraag': 0.01735275490816973, 'W_Afhandelen leads': 0.019837232960325534, 'W_Nabellen incomplete dossiers': 0.01981239589725607, 'W_Beoordelen fraude': 0.022222222222222223}, 'Derrick Watts': {'W_Completeren aanvraag': 0.01877581674802854, 'W_Nabellen offertes': 0.017540041782729804, 'W_Valideren aanvraag': 0.016592780240658644, 'W_Afhandelen leads': 0.018141742963716515, 'W_Nabellen incomplete dossiers': 0.015779784342947314, 'W_Beoordelen fraude': 0.018518518518518517}, 'Sarah Hogan': {'W_Completeren aanvraag': 0.017983060040889558, 'W_Nabellen offertes': 0.01697423398328691, 'W_Valideren aanvraag': 0.01811272957568081, 'W_Afhandelen leads': 0.016615801966768397, 'W_Nabellen incomplete dossiers': 0.019900061365828, 'W_Beoordelen fraude': 0.007407407407407408}, 'Martha Costa': {'W_Completeren aanvraag': 0.01810823215254308, 'W_Nabellen offertes': 0.017844707520891366, 'W_Valideren aanvraag': 0.016339455351488282, 'W_Afhandelen leads': 0.01627670396744659, 'W_Nabellen incomplete dossiers': 0.018935741211536777, 'W_Beoordelen fraude': 0.025925925925925925}, 'Emily Rojas': {'W_Completeren aanvraag': 0.01781616389201819, 'W_Nabellen offertes': 0.017540041782729804, 'W_Valideren aanvraag': 0.021532615579480684, 'W_Afhandelen leads': 0.015259409969481181, 'W_Nabellen incomplete dossiers': 0.018760410274392916, 'W_Beoordelen fraude': 0.025925925925925925}, 'Steven Jones': {'W_Completeren aanvraag': 0.018275128301414444, 'W_Nabellen offertes': 0.01762708913649025, 'W_Valideren aanvraag': 0.018872704243191894, 'W_Afhandelen leads': 0.017633095964733808, 'W_Nabellen incomplete dossiers': 0.016919435434382398, 'W_Beoordelen fraude': 0.025925925925925925}, 'Rebecca Guerrero': {'W_Completeren aanvraag': 0.016564442775482954, 'W_Nabellen offertes': 0.018279944289693595, 'W_Valideren aanvraag': 0.016719442685243825, 'W_Afhandelen leads': 0.0171244489657511, 'W_Nabellen

incomplete dossiers': 0.01797142105724555, 'W_Beoordelen fraude': 0.018518518518518517},
'David Doyle': {'W_Completeren aanvraag': 0.01781616389201819, 'W_Nabellen offertes':
0.016451949860724235, 'W_Valideren aanvraag': 0.01836605446485117, 'W_Afhandelen
leads': 0.018141742963716515, 'W_Nabellen incomplete dossiers': 0.019900061365828,
'W_Beoordelen fraude': 0.018518518518518517}, 'Michelle Brennan': {'W_Completeren
aanvraag': 0.017565819668711145, 'W_Nabellen offertes': 0.018105849582172703,
'W_Valideren aanvraag': 0.014312856238125397, 'W_Afhandelen leads':
0.019667683960664633, 'W_Nabellen incomplete dossiers': 0.018672744805820988,
'W_Beoordelen fraude': 0.014814814814814815}, 'Donald Wilson': {'W_Completeren
aanvraag': 0.017983060040889558, 'W_Nabellen offertes': 0.018105849582172703,
'W_Valideren aanvraag': 0.017099430018999367, 'W_Afhandelen leads':
0.019837232960325534, 'W_Nabellen incomplete dossiers': 0.01814675199438941,
'W_Beoordelen fraude': 0.02962962962962963}, 'Austin Jones': {'W_Completeren aanvraag':
0.01735719948262194, 'W_Nabellen offertes': 0.01854108635097493, 'W_Valideren aanvraag':
0.019126029132362256, 'W_Afhandelen leads': 0.019159036961681926, 'W_Nabellen
incomplete dossiers': 0.020776716051547295, 'W_Beoordelen fraude':
0.011111111111111112}, 'Joshua Phelps': {'W_Completeren aanvraag': 0.01840030041306797,
'W_Nabellen offertes': 0.01749651810584958, 'W_Valideren aanvraag':
0.015959468017732743, 'W_Afhandelen leads': 0.01780264496439471, 'W_Nabellen
incomplete dossiers': 0.01630577715437889, 'W_Beoordelen fraude': 0.011111111111111112},
'Nathan Thomas': {'W_Completeren aanvraag': 0.017482371594275462, 'W_Nabellen offertes':
0.018628133704735376, 'W_Valideren aanvraag': 0.016212792906903105, 'W_Afhandelen
leads': 0.01780264496439471, 'W_Nabellen incomplete dossiers': 0.019461734022968354,
'W_Beoordelen fraude': 0.007407407407407408}, 'Mary Thompson': {'W_Completeren
aanvraag': 0.01840030041306797, 'W_Nabellen offertes': 0.016626044568245124,
'W_Valideren aanvraag': 0.01747941735275491, 'W_Afhandelen leads':
0.018141742963716515, 'W_Nabellen incomplete dossiers': 0.01770842465152976,
'W_Beoordelen fraude': 0.018518518518518517}, 'Patricia Lloyd': {'W_Completeren aanvraag':
0.018984436934117745, 'W_Nabellen offertes': 0.016016713091922007, 'W_Valideren
aanvraag': 0.016086130462317924, 'W_Afhandelen leads': 0.018311291963377416,
'W_Nabellen incomplete dossiers': 0.017094766371526255, 'W_Beoordelen fraude':
0.02962962962962963}, 'Kayla Woods': {'W_Completeren aanvraag': 0.018567196561939333,
'W_Nabellen offertes': 0.018628133704735376, 'W_Valideren aanvraag':
0.016212792906903105, 'W_Afhandelen leads': 0.019328585961342827, 'W_Nabellen
incomplete dossiers': 0.018760410274392916, 'W_Beoordelen fraude':
0.014814814814814815}, 'Luis Jackson': {'W_Completeren aanvraag': 0.01810823215254308,
'W_Nabellen offertes': 0.019933844011142062, 'W_Valideren aanvraag':
0.018492716909436352, 'W_Afhandelen leads': 0.019837232960325534, 'W_Nabellen
incomplete dossiers': 0.018672744805820988, 'W_Beoordelen fraude':
0.014814814814814815}, 'Donald White': {'W_Completeren aanvraag':
0.016522718738265114, 'W_Nabellen offertes': 0.017409470752089137, 'W_Valideren
aanvraag': 0.018619379354021533, 'W_Afhandelen leads': 0.018819938962360123,
'W_Nabellen incomplete dossiers': 0.018497413868677128, 'W_Beoordelen fraude':

0.025925925925925925}, 'Shawn Barnes': {'W_Completeren aanvraag': 0.01764926774314683, 'W_Nabellen offertes': 0.01788823119777159, 'W_Valideren aanvraag': 0.01760607979734009, 'W_Afhandelen leads': 0.018141742963716515, 'W_Nabellen incomplete dossiers': 0.018848075742964845, 'W_Beoordelen fraude': 0.025925925925925925}, 'Heather Patterson': {'W_Completeren aanvraag': 0.01915133308298911, 'W_Nabellen offertes': 0.017540041782729804, 'W_Valideren aanvraag': 0.017226092463584548, 'W_Afhandelen leads': 0.016615801966768397, 'W_Nabellen incomplete dossiers': 0.017094766371526255, 'W_Beoordelen fraude': 0.011111111111111112}, 'Pam Martinez': {'W_Completeren aanvraag': 0.017482371594275462, 'W_Nabellen offertes': 0.017540041782729804, 'W_Valideren aanvraag': 0.015199493350221659, 'W_Afhandelen leads': 0.018480840963038318, 'W_Nabellen incomplete dossiers': 0.017094766371526255, 'W_Beoordelen fraude': 0.014814814814814815}, 'Jesse Moore': {'W_Completeren aanvraag': 0.01781616389201819, 'W_Nabellen offertes': 0.016451949860724235, 'W_Valideren aanvraag': 0.019252691576947437, 'W_Afhandelen leads': 0.015259409969481181, 'W_Nabellen incomplete dossiers': 0.017620759182957833, 'W_Beoordelen fraude': 0.022222222222222223}, 'Travis Edwards': {'W_Completeren aanvraag': 0.01952684941794968, 'W_Nabellen offertes': 0.01732242339832869, 'W_Valideren aanvraag': 0.02051931602279924, 'W_Afhandelen leads': 0.01797219396405561, 'W_Nabellen incomplete dossiers': 0.01788375558867362, 'W_Beoordelen fraude': 0.011111111111111112}, 'Ruth Davies': {'W_Completeren aanvraag': 0.01673133892435432, 'W_Nabellen offertes': 0.01701775766016713, 'W_Valideren aanvraag': 0.016466117796073463, 'W_Afhandelen leads': 0.017633095964733808, 'W_Nabellen incomplete dossiers': 0.019461734022968354, 'W_Beoordelen fraude': 0.014814814814814815}, 'Nancy Clayton': {'W_Completeren aanvraag': 0.01848374848750365, 'W_Nabellen offertes': 0.019063370473537605, 'W_Valideren aanvraag': 0.019126029132362256, 'W_Afhandelen leads': 0.017293997965412006, 'W_Nabellen incomplete dossiers': 0.018760410274392916, 'W_Beoordelen fraude': 0.018518518518518517}, 'Sherry Klein': {'W_Completeren aanvraag': 0.01769099178036467, 'W_Nabellen offertes': 0.018802228412256268, 'W_Valideren aanvraag': 0.019126029132362256, 'W_Afhandelen leads': 0.0171244489657511, 'W_Nabellen incomplete dossiers': 0.01814675199438941, 'W_Beoordelen fraude': 0.022222222222222223}, 'Jacob Wilson': {'W_Completeren aanvraag': 0.018650644636375016, 'W_Nabellen offertes': 0.018584610027855154, 'W_Valideren aanvraag': 0.018619379354021533, 'W_Afhandelen leads': 0.019837232960325534, 'W_Nabellen incomplete dossiers': 0.016744104497238538, 'W_Beoordelen fraude': 0.022222222222222223}, 'Michael Anderson MD': {'W_Completeren aanvraag': 0.016647890849918637, 'W_Nabellen offertes': 0.01688718662952646, 'W_Valideren aanvraag': 0.018746041798606713, 'W_Afhandelen leads': 0.020854526958290945, 'W_Nabellen incomplete dossiers': 0.015867449811519243, 'W_Beoordelen fraude': 0.022222222222222223}, 'Jennifer Collier': {'W_Completeren aanvraag': 0.01702340718487921, 'W_Nabellen offertes': 0.01745299442896936, 'W_Valideren aanvraag': 0.019632678910702975, 'W_Afhandelen leads': 0.0171244489657511, 'W_Nabellen incomplete dossiers': 0.019023406680108705, 'W_Beoordelen fraude': 0.011111111111111112}, 'Kevin Schultz': {'W_Completeren aanvraag': 0.017607543705928985, 'W_Nabellen offertes': 0.01841051532033426, 'W_Valideren

aanvraag': 0.019632678910702975, 'W_Afhandelen leads': 0.020006781959986435, 'W_Nabellen incomplete dossiers': 0.015166126062943807, 'W_Beoordelen fraude': 0.018518518518518517}}

group_relative_stake

{'Jennifer Thomas MD': {'W_Completeren aanvraag': 0.33037037037037037, 'W_Nabellen offertes': 0.31333333333333335, 'W_Valideren aanvraag': 0.11333333333333333, 'W_Afhandelen leads': 0.08148148148148149, 'W_Nabellen incomplete dossiers': 0.15925925925925927, 'W_Beoordelen fraude': 0.0022222222222222222}, 'Robert Peters': {'W_Completeren aanvraag': 0.30895645028759244, 'W_Nabellen offertes': 0.3294987674609696, 'W_Valideren aanvraag': 0.11503697617091208, 'W_Afhandelen leads': 0.08052588331963846, 'W_Nabellen incomplete dossiers': 0.16105176663927692, 'W_Beoordelen fraude': 0.0049301561216105174}, 'Christina Barnes': {'W_Completeren aanvraag': 0.34603421461897355, 'W_Nabellen offertes': 0.3055987558320373, 'W_Valideren aanvraag': 0.11586314152410575, 'W_Afhandelen leads': 0.08475894245723173, 'W_Nabellen incomplete dossiers': 0.14463452566096424, 'W_Beoordelen fraude': 0.003110419906687403}, 'David Adams': {'W_Completeren aanvraag': 0.32684824902723736, 'W_Nabellen offertes': 0.31284046692607004, 'W_Valideren aanvraag': 0.10194552529182879, 'W_Afhandelen leads': 0.09260700389105059, 'W_Nabellen incomplete dossiers': 0.16031128404669262, 'W_Beoordelen fraude': 0.005447470817120622}, 'William Arroyo': {'W_Completeren aanvraag': 0.3252623083131558, 'W_Nabellen offertes': 0.3365617433414044, 'W_Valideren aanvraag': 0.10411622276029056, 'W_Afhandelen leads': 0.08797417271993543, 'W_Nabellen incomplete dossiers': 0.14043583535108958, 'W_Beoordelen fraude': 0.005649717514124294}, 'Anna Schroeder': {'W_Completeren aanvraag': 0.31697819314641745, 'W_Nabellen offertes': 0.338785046728972, 'W_Valideren aanvraag': 0.11214953271028037, 'W_Afhandelen leads': 0.06619937694704049, 'W_Nabellen incomplete dossiers': 0.15809968847352024, 'W_Beoordelen fraude': 0.00778816199376947}, 'Brianna Bell': {'W_Completeren aanvraag': 0.34930950446791226, 'W_Nabellen offertes': 0.3151909017059301, 'W_Valideren aanvraag': 0.09991876523151909, 'W_Afhandelen leads': 0.07636068237205523, 'W_Nabellen incomplete dossiers': 0.15434606011372867, 'W_Beoordelen fraude': 0.00487408610885459}, 'Brandon Powell': {'W_Completeren aanvraag': 0.35402298850574715, 'W_Nabellen offertes': 0.30421455938697317, 'W_Valideren aanvraag': 0.1049808429118774, 'W_Afhandelen leads': 0.06666666666666667, 'W_Nabellen incomplete dossiers': 0.16628352490421455, 'W_Beoordelen fraude': 0.0038314176245210726}, 'Aaron Ray': {'W_Completeren aanvraag': 0.3281613653995345, 'W_Nabellen offertes': 0.3180760279286268, 'W_Valideren aanvraag': 0.11171450737005431, 'W_Afhandelen leads': 0.0721489526764934, 'W_Nabellen incomplete dossiers': 0.16834755624515127, 'W_Beoordelen fraude': 0.0015515903801396431}, 'Scott Little': {'W_Completeren aanvraag': 0.34433608402100524, 'W_Nabellen offertes': 0.32333083270817703, 'W_Valideren aanvraag': 0.1177794448612153, 'W_Afhandelen leads': 0.07426856714178545, 'W_Nabellen incomplete dossiers': 0.13728432108027006, 'W_Beoordelen fraude': 0.003000750187546887}, 'Ashley Daniel': {'W_Completeren aanvraag':

0.3335883703136955, 'W_Nabellen offertes': 0.31216526396327465, 'W_Valideren aanvraag': 0.10711553175210406, 'W_Afhandelen leads': 0.0772762050497322, 'W_Nabellen incomplete dossiers': 0.16679418515684774, 'W_Beoordelen fraude': 0.00306044376434583}, 'Lisa Henson': {'W_Completeren aanvraag': 0.31778656126482213, 'W_Nabellen offertes': 0.3383399209486166, 'W_Valideren aanvraag': 0.11699604743083004, 'W_Afhandelen leads': 0.07035573122529644, 'W_Nabellen incomplete dossiers': 0.1533596837944664, 'W_Beoordelen fraude': 0.0031620553359683794}, 'Jennifer Wilson': {'W_Completeren aanvraag': 0.30616907844630614, 'W_Nabellen offertes': 0.33434881949733436, 'W_Valideren aanvraag': 0.10662604722010663, 'W_Afhandelen leads': 0.08149276466108149, 'W_Nabellen incomplete dossiers': 0.1667936024371668, 'W_Beoordelen fraude': 0.00456968773800457}, 'Jane Turner': {'W_Completeren aanvraag': 0.35675253708040594, 'W_Nabellen offertes': 0.32474629195940674, 'W_Valideren aanvraag': 0.11553473848555816, 'W_Afhandelen leads': 0.0702576112412178, 'W_Nabellen incomplete dossiers': 0.13114754098360656, 'W_Beoordelen fraude': 0.00156128024980484}, 'Brianna Martinez': {'W_Completeren aanvraag': 0.30970724191063176, 'W_Nabellen offertes': 0.3289676425269646, 'W_Valideren aanvraag': 0.1086286594761171, 'W_Afhandelen leads': 0.08551617873651772, 'W_Nabellen incomplete dossiers': 0.16640986132511557, 'W_Beoordelen fraude': 0.0007704160246533128}, 'Laura Morris': {'W_Completeren aanvraag': 0.3264, 'W_Nabellen offertes': 0.356, 'W_Valideren aanvraag': 0.1056, 'W_Afhandelen leads': 0.0632, 'W_Nabellen incomplete dossiers': 0.1456, 'W_Beoordelen fraude': 0.0032}, 'Wesley Mills': {'W_Completeren aanvraag': 0.33384615384615385, 'W_Nabellen offertes': 0.3130769230769231, 'W_Valideren aanvraag': 0.11, 'W_Afhandelen leads': 0.0876923076923077, 'W_Nabellen incomplete dossiers': 0.15307692307692308, 'W_Beoordelen fraude': 0.002307692307692308}, 'Brittany Valencia': {'W_Completeren aanvraag': 0.3341213553979511, 'W_Nabellen offertes': 0.30023640661938533, 'W_Valideren aanvraag': 0.11505122143420016, 'W_Afhandelen leads': 0.08353033884948778, 'W_Nabellen incomplete dossiers': 0.16390858944050432, 'W_Beoordelen fraude': 0.0031520882584712374}, 'Rhonda Brady': {'W_Completeren aanvraag': 0.3320063694267516, 'W_Nabellen offertes': 0.304140127388535, 'W_Valideren aanvraag': 0.11464968152866242, 'W_Afhandelen leads': 0.09872611464968153, 'W_Nabellen incomplete dossiers': 0.1449044585987261, 'W_Beoordelen fraude': 0.005573248407643312}, 'Emily Rice': {'W_Completeren aanvraag': 0.33358098068350667, 'W_Nabellen offertes': 0.2949479940564636, 'W_Valideren aanvraag': 0.1151560178306092, 'W_Afhandelen leads': 0.08023774145616643, 'W_Nabellen incomplete dossiers': 0.1723625557206538, 'W_Beoordelen fraude': 0.003714710252600297}, 'Patrick Cole': {'W_Completeren aanvraag': 0.32494279176201374, 'W_Nabellen offertes': 0.3180778032036613, 'W_Valideren aanvraag': 0.10373760488176964, 'W_Afhandelen leads': 0.08771929824561403, 'W_Nabellen incomplete dossiers': 0.16094584286803967, 'W_Beoordelen fraude': 0.004576659038901602}, 'Paul Alvarado': {'W_Completeren aanvraag': 0.3290468986384266, 'W_Nabellen offertes': 0.32223903177004537, 'W_Valideren aanvraag': 0.11270801815431165, 'W_Afhandelen leads': 0.08774583963691376, 'W_Nabellen incomplete dossiers': 0.14523449319213314, 'W_Beoordelen fraude': 0.0030257186081694403}, 'Justin Thompson': {'W_Completeren aanvraag': 0.36490683229813664, 'W_Nabellen offertes': 0.30978260869565216, 'W_Valideren aanvraag': 0.10403726708074534, 'W_Afhandelen leads': 0.07686335403726709, 'W_Nabellen

incomplete dossiers': 0.14130434782608695, 'W_Beoordelen fraude': 0.003105590062111801}, 'Vanessa Shepherd': {'W_Completeren aanvraag': 0.3282442748091603, 'W_Nabellen offertes': 0.32137404580152673, 'W_Valideren aanvraag': 0.10305343511450382, 'W_Afhandelen leads': 0.08854961832061069, 'W_Nabellen incomplete dossiers': 0.15801526717557252, 'W_Beoordelen fraude': 0.0007633587786259542}, 'Matthew Walker': {'W_Completeren aanvraag': 0.3102906520031422, 'W_Nabellen offertes': 0.32128829536527886, 'W_Valideren aanvraag': 0.11233307148468186, 'W_Afhandelen leads': 0.08798114689709348, 'W_Nabellen incomplete dossiers': 0.1610369206598586, 'W_Beoordelen fraude': 0.007069913589945012}, 'Miguel Reeves': {'W_Completeren aanvraag': 0.33819018404907975, 'W_Nabellen offertes': 0.32745398773006135, 'W_Valideren aanvraag': 0.10199386503067484, 'W_Afhandelen leads': 0.07592024539877301, 'W_Nabellen incomplete dossiers': 0.1549079754601227, 'W_Beoordelen fraude': 0.0015337423312883436}, 'Kathy Ingram': {'W_Completeren aanvraag': 0.31835205992509363, 'W_Nabellen offertes': 0.31760299625468164, 'W_Valideren aanvraag': 0.10262172284644194, 'W_Afhandelen leads': 0.08764044943820225, 'W_Nabellen incomplete dossiers': 0.16928838951310862, 'W_Beoordelen fraude': 0.0044943820224719105}, 'Derrick Watts': {'W_Completeren aanvraag': 0.35266457680250785, 'W_Nabellen offertes': 0.3158307210031348, 'W_Valideren aanvraag': 0.10266457680250783, 'W_Afhandelen leads': 0.08385579937304075, 'W_Nabellen incomplete dossiers': 0.14106583072100312, 'W_Beoordelen fraude': 0.003918495297805642}, 'Sarah Hogan': {'W_Completeren aanvraag': 0.3338497288923315, 'W_Nabellen offertes': 0.3020914020139427, 'W_Valideren aanvraag': 0.11076684740511232, 'W_Afhandelen leads': 0.07591014717273431, 'W_Nabellen incomplete dossiers': 0.1758326878388846, 'W_Beoordelen fraude': 0.001549186676994578}, 'Martha Costa': {'W_Completeren aanvraag': 0.33591331269349844, 'W_Nabellen offertes': 0.3173374613003096, 'W_Valideren aanvraag': 0.0998452012383901, 'W_Afhandelen leads': 0.07430340557275542, 'W_Nabellen incomplete dossiers': 0.16718266253869968, 'W_Beoordelen fraude': 0.005417956656346749}, 'Emily Rojas': {'W_Completeren aanvraag': 0.32570556826849734, 'W_Nabellen offertes': 0.30739893211289093, 'W_Valideren aanvraag': 0.12967200610221205, 'W_Afhandelen leads': 0.06864988558352403, 'W_Nabellen incomplete dossiers': 0.16323417238749047, 'W_Beoordelen fraude': 0.005339435545385202}, 'Steven Jones': {'W_Completeren aanvraag': 0.33796296296296297, 'W_Nabellen offertes': 0.3125, 'W_Valideren aanvraag': 0.11496913580246913, 'W_Afhandelen leads': 0.08024691358024691, 'W_Nabellen incomplete dossiers': 0.14891975308641975, 'W_Beoordelen fraude': 0.005401234567901234}, 'Rebecca Guerrero': {'W_Completeren aanvraag': 0.31507936507936507, 'W_Nabellen offertes': 0.3333333333333333, 'W_Valideren aanvraag': 0.10476190476190476, 'W_Afhandelen leads': 0.08015873015873017, 'W_Nabellen incomplete dossiers': 0.1626984126984127, 'W_Beoordelen fraude': 0.003968253968253968}, 'David Doyle': {'W_Completeren aanvraag': 0.3312645461598138, 'W_Nabellen offertes': 0.2932505818463926, 'W_Valideren aanvraag': 0.11249030256012413, 'W_Afhandelen leads': 0.08301008533747091, 'W_Nabellen incomplete dossiers': 0.1761055081458495, 'W_Beoordelen fraude': 0.003878975950349108}, 'Michelle Brennan': {'W_Completeren aanvraag': 0.32813717848791896, 'W_Nabellen offertes': 0.32424006235385816, 'W_Valideren aanvraag': 0.08807482462977397, 'W_Afhandelen leads': 0.09041309431021044, 'W_Nabellen incomplete dossiers': 0.16601714731098988,

'W_Beoordelen fraude': 0.003117692907248636}, 'Donald Wilson': {'W_Completeren aanvraag': 0.3280060882800609, 'W_Nabellen offertes': 0.3165905631659056, 'W_Valideren aanvraag': 0.10273972602739725, 'W_Afhandelen leads': 0.08904109589041095, 'W_Nabellen incomplete dossiers': 0.15753424657534246, 'W_Beoordelen fraude': 0.0060882800608828}, 'Austin Jones': {'W_Completeren aanvraag': 0.30906389301634474, 'W_Nabellen offertes': 0.3164933135215453, 'W_Valideren aanvraag': 0.11218424962852898, 'W_Afhandelen leads': 0.08395245170876671, 'W_Nabellen incomplete dossiers': 0.1760772659732541, 'W_Beoordelen fraude': 0.002228826151560178}, 'Joshua Phelps': {'W_Completeren aanvraag': 0.3491686460807601, 'W_Nabellen offertes': 0.3182897862232779, 'W_Valideren aanvraag': 0.0997624703087886, 'W_Afhandelen leads': 0.0831353919239905, 'W_Nabellen incomplete dossiers': 0.14726840855106887, 'W_Beoordelen fraude': 0.0023752969121140144}, 'Nathan Thomas': {'W_Completeren aanvraag': 0.32131901840490795, 'W_Nabellen offertes': 0.3282208588957055, 'W_Valideren aanvraag': 0.09815950920245399, 'W_Afhandelen leads': 0.08052147239263803, 'W_Nabellen incomplete dossiers': 0.17024539877300612, 'W_Beoordelen fraude': 0.0015337423312883436}, 'Mary Thompson': {'W_Completeren aanvraag': 0.3458823529411765, 'W_Nabellen offertes': 0.2996078431372549, 'W_Valideren aanvraag': 0.10823529411764705, 'W_Afhandelen leads': 0.08392156862745098, 'W_Nabellen incomplete dossiers': 0.1584313725490196, 'W_Beoordelen fraude': 0.00392156862745098}, 'Patricia Lloyd': {'W_Completeren aanvraag': 0.36082474226804123, 'W_Nabellen offertes': 0.29183187946074546, 'W_Valideren aanvraag': 0.1007137192704203, 'W_Afhandelen leads': 0.08564631245043616, 'W_Nabellen incomplete dossiers': 0.15463917525773196, 'W_Beoordelen fraude': 0.006344171292624901}, 'Kayla Woods': {'W_Completeren aanvraag': 0.33383345836459116, 'W_Nabellen offertes': 0.3210802700675169, 'W_Valideren aanvraag': 0.09602400600150038, 'W_Afhandelen leads': 0.08552138034508627, 'W_Nabellen incomplete dossiers': 0.16054013503375844, 'W_Beoordelen fraude': 0.003000750187546887}, 'Luis Jackson': {'W_Completeren aanvraag': 0.3163265306122449, 'W_Nabellen offertes': 0.3338192419825073, 'W_Valideren aanvraag': 0.10641399416909621, 'W_Afhandelen leads': 0.08527696793002916, 'W_Nabellen incomplete dossiers': 0.1552478134110787, 'W_Beoordelen fraude': 0.0029154518950437317}, 'Donald White': {'W_Completeren aanvraag': 0.3113207547169811, 'W_Nabellen offertes': 0.31446540880503143, 'W_Valideren aanvraag': 0.11556603773584906, 'W_Afhandelen leads': 0.08726415094339622, 'W_Nabellen incomplete dossiers': 0.16588050314465408, 'W_Beoordelen fraude': 0.00550314465408805}, 'Shawn Barnes': {'W_Completeren aanvraag': 0.3248847926267281, 'W_Nabellen offertes': 0.315668202764977, 'W_Valideren aanvraag': 0.10675883256528418, 'W_Afhandelen leads': 0.08218125960061444, 'W_Nabellen incomplete dossiers': 0.1651305683563748, 'W_Beoordelen fraude': 0.005376344086021506}, 'Heather Patterson': {'W_Completeren aanvraag': 0.3547140649149923, 'W_Nabellen offertes': 0.3114374034003091, 'W_Valideren aanvraag': 0.10510046367851623, 'W_Afhandelen leads': 0.07573415765069552, 'W_Nabellen incomplete dossiers': 0.15069551777434312, 'W_Beoordelen fraude': 0.00231839258114374}, 'Pam Martinez': {'W_Completeren aanvraag': 0.3352, 'W_Nabellen offertes': 0.3224, 'W_Valideren aanvraag': 0.096, 'W_Afhandelen leads': 0.0872, 'W_Nabellen incomplete dossiers': 0.156, 'W_Beoordelen fraude': 0.0032}, 'Jesse Moore': {'W_Completeren aanvraag': 0.3405103668261563, 'W_Nabellen offertes': 0.3014354066985646, 'W_Valideren aanvraag':

0.12121212121212122, 'W_Afhandelen leads': 0.07177033492822966, 'W_Nabellen incomplete dossiers': 0.16028708133971292, 'W_Beoordelen fraude': 0.004784688995215311}, 'Travis Edwards': {'W_Completeren aanvraag': 0.348993288590604, 'W_Nabellen offertes': 0.29679343773303507, 'W_Valideren aanvraag': 0.12080536912751678, 'W_Afhandelen leads': 0.0790454884414616, 'W_Nabellen incomplete dossiers': 0.15212527964205816, 'W_Beoordelen fraude': 0.0022371364653243847}, 'Ruth Davies': {'W_Completeren aanvraag': 0.32028753993610226, 'W_Nabellen offertes': 0.31230031948881787, 'W_Valideren aanvraag': 0.10383386581469649, 'W_Afhandelen leads': 0.08306709265175719, 'W_Nabellen incomplete dossiers': 0.17731629392971246, 'W_Beoordelen fraude': 0.003194888178913738}, 'Nancy Clayton': {'W_Completeren aanvraag': 0.32742054693274203, 'W_Nabellen offertes': 0.3237250554323725, 'W_Valideren aanvraag': 0.11160384331116038, 'W_Afhandelen leads': 0.07538802660753881, 'W_Nabellen incomplete dossiers': 0.1581670362158167, 'W_Beoordelen fraude': 0.003695491500369549}, 'Sherry Klein': {'W_Completeren aanvraag': 0.3209689629068887, 'W_Nabellen offertes': 0.3270249810749432, 'W_Valideren aanvraag': 0.11430734292202877, 'W_Afhandelen leads': 0.07645722937168811, 'W_Nabellen incomplete dossiers': 0.1566994700984103, 'W_Beoordelen fraude': 0.004542013626040878}, 'Jacob Wilson': {'W_Completeren aanvraag': 0.3348314606741573, 'W_Nabellen offertes': 0.31985018726591763, 'W_Valideren aanvraag': 0.1101123595505618, 'W_Afhandelen leads': 0.08764044943820225, 'W_Nabellen incomplete dossiers': 0.14307116104868914, 'W_Beoordelen fraude': 0.0044943820224719105}, 'Michael Anderson MD': {'W_Completeren aanvraag': 0.3204819277108434, 'W_Nabellen offertes': 0.3116465863453815, 'W_Valideren aanvraag': 0.11887550200803212, 'W_Afhandelen leads': 0.09879518072289156, 'W_Nabellen incomplete dossiers': 0.14538152610441768, 'W_Beoordelen fraude': 0.004819277108433735}, 'Jennifer Collier': {'W_Completeren aanvraag': 0.3175097276264591, 'W_Nabellen offertes': 0.3120622568093385, 'W_Valideren aanvraag': 0.12062256809338522, 'W_Afhandelen leads': 0.07859922178988327, 'W_Nabellen incomplete dossiers': 0.1688715953307393, 'W_Beoordelen fraude': 0.0023346303501945525}, 'Kevin Schultz': {'W_Completeren aanvraag': 0.3256172839506173, 'W_Nabellen offertes': 0.3263888888888889, 'W_Valideren aanvraag': 0.11959876543209877, 'W_Afhandelen leads': 0.09104938271604938, 'W_Nabellen incomplete dossiers': 0.13348765432098766, 'W_Beoordelen fraude': 0.0038580246913580245}}

group_coverage

{'Jennifer Thomas MD': {'Jennifer Thomas MD': 1.0}, 'Robert Peters': {'Robert Peters': 1.0}, 'Christina Barnes': {'Christina Barnes': 1.0}, 'David Adams': {'David Adams': 1.0}, 'William Arroyo': {'William Arroyo': 1.0}, 'Anna Schroeder': {'Anna Schroeder': 1.0}, 'Brianna Bell': {'Brianna Bell': 1.0}, 'Brandon Powell': {'Brandon Powell': 1.0}, 'Aaron Ray': {'Aaron Ray': 1.0}, 'Scott Little': {'Scott Little': 1.0}, 'Ashley Daniel': {'Ashley Daniel': 1.0}, 'Lisa Henson': {'Lisa Henson': 1.0}, 'Jennifer Wilson': {'Jennifer Wilson': 1.0}, 'Jane Turner': {'Jane Turner': 1.0}, 'Brianna Martinez': {'Brianna Martinez': 1.0}, 'Laura Morris': {'Laura Morris': 1.0}, 'Wesley Mills': {'Wesley Mills': 1.0}, 'Brittany Valencia': {'Brittany Valencia': 1.0},

'Rhonda Brady': {'Rhonda Brady': 1.0}, 'Emily Rice': {'Emily Rice': 1.0}, 'Patrick Cole': {'Patrick Cole': 1.0}, 'Paul Alvarado': {'Paul Alvarado': 1.0}, 'Justin Thompson': {'Justin Thompson': 1.0}, 'Vanessa Shepherd': {'Vanessa Shepherd': 1.0}, 'Matthew Walker': {'Matthew Walker': 1.0}, 'Miguel Reeves': {'Miguel Reeves': 1.0}, 'Kathy Ingram': {'Kathy Ingram': 1.0}, 'Derrick Watts': {'Derrick Watts': 1.0}, 'Sarah Hogan': {'Sarah Hogan': 1.0}, 'Martha Costa': {'Martha Costa': 1.0}, 'Emily Rojas': {'Emily Rojas': 1.0}, 'Steven Jones': {'Steven Jones': 1.0}, 'Rebecca Guerrero': {'Rebecca Guerrero': 1.0}, 'David Doyle': {'David Doyle': 1.0}, 'Michelle Brennan': {'Michelle Brennan': 1.0}, 'Donald Wilson': {'Donald Wilson': 1.0}, 'Austin Jones': {'Austin Jones': 1.0}, 'Joshua Phelps': {'Joshua Phelps': 1.0}, 'Nathan Thomas': {'Nathan Thomas': 1.0}, 'Mary Thompson': {'Mary Thompson': 1.0}, 'Patricia Lloyd': {'Patricia Lloyd': 1.0}, 'Kayla Woods': {'Kayla Woods': 1.0}, 'Luis Jackson': {'Luis Jackson': 1.0}, 'Donald White': {'Donald White': 1.0}, 'Shawn Barnes': {'Shawn Barnes': 1.0}, 'Heather Patterson': {'Heather Patterson': 1.0}, 'Pam Martinez': {'Pam Martinez': 1.0}, 'Jesse Moore': {'Jesse Moore': 1.0}, 'Travis Edwards': {'Travis Edwards': 1.0}, 'Ruth Davies': {'Ruth Davies': 1.0}, 'Nancy Clayton': {'Nancy Clayton': 1.0}, 'Sherry Klein': {'Sherry Klein': 1.0}, 'Jacob Wilson': {'Jacob Wilson': 1.0}, 'Michael Anderson MD': {'Michael Anderson MD': 1.0}, 'Jennifer Collier': {'Jennifer Collier': 1.0}, 'Kevin Schultz': {'Kevin Schultz': 1.0}}

group_member_contribution

{'Jennifer Thomas MD': {'Jennifer Thomas MD': {'W_Completeren aanvraag': 446, 'W_Nabellen incomplete dossiers': 215, 'W_Nabellen offertes': 423, 'W_Valideren aanvraag': 153, 'W_Afhandelen leads': 110, 'W_Beoordelen fraude': 3}}, 'Robert Peters': {'Robert Peters': {'W_Nabellen offertes': 401, 'W_Completeren aanvraag': 376, 'W_Valideren aanvraag': 140, 'W_Nabellen incomplete dossiers': 196, 'W_Beoordelen fraude': 6, 'W_Afhandelen leads': 98}}, 'Christina Barnes': {'Christina Barnes': {'W_Nabellen offertes': 393, 'W_Completeren aanvraag': 445, 'W_Valideren aanvraag': 149, 'W_Nabellen incomplete dossiers': 186, 'W_Afhandelen leads': 109, 'W_Beoordelen fraude': 4}}, 'David Adams': {'David Adams': {'W_Nabellen offertes': 402, 'W_Nabellen incomplete dossiers': 206, 'W_Completeren aanvraag': 420, 'W_Valideren aanvraag': 131, 'W_Afhandelen leads': 119, 'W_Beoordelen fraude': 7}}, 'William Arroyo': {'William Arroyo': {'W_Valideren aanvraag': 129, 'W_Nabellen incomplete dossiers': 174, 'W_Nabellen offertes': 417, 'W_Completeren aanvraag': 403, 'W_Afhandelen leads': 109, 'W_Beoordelen fraude': 7}}, 'Anna Schroeder': {'Anna Schroeder': {'W_Completeren aanvraag': 407, 'W_Nabellen offertes': 435, 'W_Nabellen incomplete dossiers': 203, 'W_Valideren aanvraag': 144, 'W_Afhandelen leads': 85, 'W_Beoordelen fraude': 10}}, 'Brianna Bell': {'Brianna Bell': {'W_Completeren aanvraag': 430, 'W_Nabellen incomplete dossiers': 190, 'W_Nabellen offertes': 388, 'W_Afhandelen leads': 94, 'W_Valideren aanvraag': 123, 'W_Beoordelen fraude': 6}}, 'Brandon Powell': {'Brandon Powell': {'W_Nabellen offertes': 397, 'W_Completeren aanvraag': 462, 'W_Nabellen incomplete dossiers': 217, 'W_Afhandelen leads': 87, 'W_Valideren aanvraag': 137, 'W_Beoordelen fraude': 5}}, 'Aaron Ray': {'Aaron Ray': {'W_Nabellen offertes': 410, 'W_Completeren aanvraag': 423, 'W_Valideren aanvraag': 144, 'W_Nabellen incomplete dossiers': 217, 'W_Afhandelen leads': 93, 'W_Beoordelen fraude': 2}},

'Scott Little': {'Scott Little': {'W_Nabellen offertes': 431, 'W_Nabellen incomplete dossiers': 183, 'W_Completeren aanvraag': 459, 'W_Afhandelen leads': 99, 'W_Valideren aanvraag': 157, 'W_Beoordelen fraude': 4}}, 'Ashley Daniel': {'Ashley Daniel': {'W_Valideren aanvraag': 140, 'W_Nabellen offertes': 408, 'W_Nabellen incomplete dossiers': 218, 'W_Completeren aanvraag': 436, 'W_Afhandelen leads': 101, 'W_Beoordelen fraude': 4}}, 'Lisa Henson': {'Lisa Henson': {'W_Valideren aanvraag': 148, 'W_Nabellen offertes': 428, 'W_Completeren aanvraag': 402, 'W_Nabellen incomplete dossiers': 194, 'W_Afhandelen leads': 89, 'W_Beoordelen fraude': 4}}, 'Jennifer Wilson': {'Jennifer Wilson': {'W_Valideren aanvraag': 140, 'W_Completeren aanvraag': 402, 'W_Afhandelen leads': 107, 'W_Beoordelen fraude': 6, 'W_Nabellen offertes': 439, 'W_Nabellen incomplete dossiers': 219}}, 'Jane Turner': {'Jane Turner': {'W_Completeren aanvraag': 457, 'W_Valideren aanvraag': 148, 'W_Nabellen offertes': 416, 'W_Afhandelen leads': 90, 'W_Nabellen incomplete dossiers': 168, 'W_Beoordelen fraude': 2}}, 'Brianna Martinez': {'Brianna Martinez': {'W_Nabellen offertes': 427, 'W_Completeren aanvraag': 402, 'W_Valideren aanvraag': 141, 'W_Nabellen incomplete dossiers': 216, 'W_Afhandelen leads': 111, 'W_Beoordelen fraude': 1}}, 'Laura Morris': {'Laura Morris': {'W_Nabellen offertes': 445, 'W_Completeren aanvraag': 408, 'W_Valideren aanvraag': 132, 'W_Nabellen incomplete dossiers': 182, 'W_Afhandelen leads': 79, 'W_Beoordelen fraude': 4}}, 'Wesley Mills': {'Wesley Mills': {'W_Nabellen offertes': 407, 'W_Completeren aanvraag': 434, 'W_Afhandelen leads': 114, 'W_Nabellen incomplete dossiers': 199, 'W_Valideren aanvraag': 143, 'W_Beoordelen fraude': 3}}, 'Brittany Valencia': {'Brittany Valencia': {'W_Nabellen offertes': 381, 'W_Nabellen incomplete dossiers': 208, 'W_Completeren aanvraag': 424, 'W_Valideren aanvraag': 146, 'W_Afhandelen leads': 106, 'W_Beoordelen fraude': 4}}, 'Rhonda Brady': {'Rhonda Brady': {'W_Nabellen offertes': 382, 'W_Completeren aanvraag': 417, 'W_Afhandelen leads': 124, 'W_Nabellen incomplete dossiers': 182, 'W_Valideren aanvraag': 144, 'W_Beoordelen fraude': 7}}, 'Emily Rice': {'Emily Rice': {'W_Nabellen offertes': 397, 'W_Nabellen incomplete dossiers': 232, 'W_Completeren aanvraag': 449, 'W_Afhandelen leads': 108, 'W_Valideren aanvraag': 155, 'W_Beoordelen fraude': 5}}, 'Patrick Cole': {'Patrick Cole': {'W_Completeren aanvraag': 426, 'W_Nabellen incomplete dossiers': 211, 'W_Nabellen offertes': 417, 'W_Valideren aanvraag': 136, 'W_Afhandelen leads': 115, 'W_Beoordelen fraude': 6}}, 'Paul Alvarado': {'Paul Alvarado': {'W_Completeren aanvraag': 435, 'W_Valideren aanvraag': 149, 'W_Nabellen offertes': 426, 'W_Nabellen incomplete dossiers': 192, 'W_Afhandelen leads': 116, 'W_Beoordelen fraude': 4}}, 'Justin Thompson': {'Justin Thompson': {'W_Afhandelen leads': 99, 'W_Completeren aanvraag': 470, 'W_Nabellen incomplete dossiers': 182, 'W_Nabellen offertes': 399, 'W_Valideren aanvraag': 134, 'W_Beoordelen fraude': 4}}, 'Vanessa Shepherd': {'Vanessa Shepherd': {'W_Completeren aanvraag': 430, 'W_Nabellen incomplete dossiers': 207, 'W_Valideren aanvraag': 135, 'W_Nabellen offertes': 421, 'W_Afhandelen leads': 116, 'W_Beoordelen fraude': 1}}, 'Matthew Walker': {'Matthew Walker': {'W_Completeren aanvraag': 395, 'W_Nabellen incomplete dossiers': 205, 'W_Valideren aanvraag': 143, 'W_Nabellen offertes': 409, 'W_Afhandelen leads': 112, 'W_Beoordelen fraude': 9}}, 'Miguel Reeves': {'Miguel Reeves': {'W_Completeren aanvraag': 441, 'W_Afhandelen leads': 99, 'W_Nabellen offertes': 427, 'W_Nabellen incomplete dossiers': 202, 'W_Valideren aanvraag': 133, 'W_Beoordelen fraude': 2}}, 'Kathy Ingram': {'Kathy Ingram': {'W_Completeren aanvraag': 425, 'W_Valideren aanvraag': 137, 'W_Afhandelen leads': 117, 'W_Nabellen offertes': 424, 'W_Nabellen incomplete

dossiers': 226, 'W_Beoordelen fraude': 6}}, 'Derrick Watts': {'Derrick Watts': {'W_Afhandelen leads': 107, 'W_Valideren aanvraag': 131, 'W_Completeren aanvraag': 450, 'W_Nabellen incomplete dossiers': 180, 'W_Nabellen offertes': 403, 'W_Beoordelen fraude': 5}}, 'Sarah Hogan': {'Sarah Hogan': {'W_Completeren aanvraag': 431, 'W_Nabellen incomplete dossiers': 227, 'W_Nabellen offertes': 390, 'W_Afhandelen leads': 98, 'W_Valideren aanvraag': 143, 'W_Beoordelen fraude': 2}}, 'Martha Costa': {'Martha Costa': {'W_Completeren aanvraag': 434, 'W_Nabellen incomplete dossiers': 216, 'W_Valideren aanvraag': 129, 'W_Nabellen offertes': 410, 'W_Afhandelen leads': 96, 'W_Beoordelen fraude': 7}}, 'Emily Rojas': {'Emily Rojas': {'W_Completeren aanvraag': 427, 'W_Nabellen offertes': 403, 'W_Valideren aanvraag': 170, 'W_Nabellen incomplete dossiers': 214, 'W_Afhandelen leads': 90, 'W_Beoordelen fraude': 7}}, 'Steven Jones': {'Steven Jones': {'W_Completeren aanvraag': 438, 'W_Nabellen incomplete dossiers': 193, 'W_Nabellen offertes': 405, 'W_Afhandelen leads': 104, 'W_Valideren aanvraag': 149, 'W_Beoordelen fraude': 7}}, 'Rebecca Guerrero': {'Rebecca Guerrero': {'W_Nabellen offertes': 420, 'W_Completeren aanvraag': 397, 'W_Nabellen incomplete dossiers': 205, 'W_Afhandelen leads': 101, 'W_Valideren aanvraag': 132, 'W_Beoordelen fraude': 5}}, 'David Doyle': {'David Doyle': {'W_Valideren aanvraag': 145, 'W_Nabellen offertes': 378, 'W_Afhandelen leads': 107, 'W_Completeren aanvraag': 427, 'W_Nabellen incomplete dossiers': 227, 'W_Beoordelen fraude': 5}}, 'Michelle Brennan': {'Michelle Brennan': {'W_Completeren aanvraag': 421, 'W_Nabellen incomplete dossiers': 213, 'W_Nabellen offertes': 416, 'W_Valideren aanvraag': 113, 'W_Afhandelen leads': 116, 'W_Beoordelen fraude': 4}}, 'Donald Wilson': {'Donald Wilson': {'W_Nabellen offertes': 416, 'W_Completeren aanvraag': 431, 'W_Nabellen incomplete dossiers': 207, 'W_Valideren aanvraag': 135, 'W_Afhandelen leads': 117, 'W_Beoordelen fraude': 8}}, 'Austin Jones': {'Austin Jones': {'W_Nabellen offertes': 426, 'W_Nabellen incomplete dossiers': 237, 'W_Completeren aanvraag': 416, 'W_Afhandelen leads': 113, 'W_Valideren aanvraag': 151, 'W_Beoordelen fraude': 3}}, 'Joshua Phelps': {'Joshua Phelps': {'W_Nabellen offertes': 402, 'W_Completeren aanvraag': 441, 'W_Nabellen incomplete dossiers': 186, 'W_Afhandelen leads': 105, 'W_Valideren aanvraag': 126, 'W_Beoordelen fraude': 3}}, 'Nathan Thomas': {'Nathan Thomas': {'W_Nabellen incomplete dossiers': 222, 'W_Completeren aanvraag': 419, 'W_Nabellen offertes': 428, 'W_Valideren aanvraag': 128, 'W_Afhandelen leads': 105, 'W_Beoordelen fraude': 2}}, 'Mary Thompson': {'Mary Thompson': {'W_Nabellen incomplete dossiers': 202, 'W_Nabellen offertes': 382, 'W_Completeren aanvraag': 441, 'W_Afhandelen leads': 107, 'W_Valideren aanvraag': 138, 'W_Beoordelen fraude': 5}}, 'Patricia Lloyd': {'Patricia Lloyd': {'W_Nabellen incomplete dossiers': 195, 'W_Completeren aanvraag': 455, 'W_Afhandelen leads': 108, 'W_Nabellen offertes': 368, 'W_Valideren aanvraag': 127, 'W_Beoordelen fraude': 8}}, 'Kayla Woods': {'Kayla Woods': {'W_Nabellen incomplete dossiers': 214, 'W_Nabellen offertes': 428, 'W_Completeren aanvraag': 445, 'W_Valideren aanvraag': 128, 'W_Afhandelen leads': 114, 'W_Beoordelen fraude': 4}}, 'Luis Jackson': {'Luis Jackson': {'W_Nabellen incomplete dossiers': 213, 'W_Completeren aanvraag': 434, 'W_Nabellen offertes': 458, 'W_Valideren aanvraag': 146, 'W_Afhandelen leads': 117, 'W_Beoordelen fraude': 4}}, 'Donald White': {'Donald White': {'W_Completeren aanvraag': 396, 'W_Nabellen offertes': 400, 'W_Nabellen incomplete dossiers': 211, 'W_Beoordelen fraude': 7, 'W_Valideren aanvraag': 147, 'W_Afhandelen leads': 111}}, 'Shawn Barnes': {'Shawn Barnes': {'W_Nabellen incomplete dossiers': 215, 'W_Completeren aanvraag': 423, 'W_Valideren

aanvraag': 139, 'W_Nabellen offertes': 411, 'W_Afhandelen leads': 107, 'W_Beoordelen fraude': 7}}, 'Heather Patterson': {'Heather Patterson': {'W_Nabellen incomplete dossiers': 195, 'W_Afhandelen leads': 98, 'W_Completeren aanvraag': 459, 'W_Valideren aanvraag': 136, 'W_Nabellen offertes': 403, 'W_Beoordelen fraude': 3}}, 'Pam Martinez': {'Pam Martinez': {'W_Nabellen incomplete dossiers': 195, 'W_Completeren aanvraag': 419, 'W_Nabellen offertes': 403, 'W_Valideren aanvraag': 120, 'W_Afhandelen leads': 109, 'W_Beoordelen fraude': 4}}, 'Jesse Moore': {'Jesse Moore': {'W_Afhandelen leads': 90, 'W_Valideren aanvraag': 152, 'W_Completeren aanvraag': 427, 'W_Nabellen offertes': 378, 'W_Nabellen incomplete dossiers': 201, 'W_Beoordelen fraude': 6}}, 'Travis Edwards': {'Travis Edwards': {'W_Nabellen offertes': 398, 'W_Nabellen incomplete dossiers': 204, 'W_Completeren aanvraag': 468, 'W_Valideren aanvraag': 162, 'W_Afhandelen leads': 106, 'W_Beoordelen fraude': 3}}, 'Ruth Davies': {'Ruth Davies': {'W_Valideren aanvraag': 130, 'W_Completeren aanvraag': 401, 'W_Nabellen offertes': 391, 'W_Nabellen incomplete dossiers': 222, 'W_Afhandelen leads': 104, 'W_Beoordelen fraude': 4}}, 'Nancy Clayton': {'Nancy Clayton': {'W_Valideren aanvraag': 151, 'W_Nabellen offertes': 438, 'W_Afhandelen leads': 102, 'W_Completeren aanvraag': 443, 'W_Nabellen incomplete dossiers': 214, 'W_Beoordelen fraude': 5}}, 'Sherry Klein': {'Sherry Klein': {'W_Nabellen incomplete dossiers': 207, 'W_Completeren aanvraag': 424, 'W_Afhandelen leads': 101, 'W_Valideren aanvraag': 151, 'W_Nabellen offertes': 432, 'W_Beoordelen fraude': 6}}, 'Jacob Wilson': {'Jacob Wilson': {'W_Nabellen offertes': 427, 'W_Completeren aanvraag': 447, 'W_Valideren aanvraag': 147, 'W_Nabellen incomplete dossiers': 191, 'W_Afhandelen leads': 117, 'W_Beoordelen fraude': 6}}, 'Michael Anderson MD': {'Michael Anderson MD': {'W_Nabellen incomplete dossiers': 181, 'W_Completeren aanvraag': 399, 'W_Nabellen offertes': 388, 'W_Afhandelen leads': 123, 'W_Valideren aanvraag': 148, 'W_Beoordelen fraude': 6}}, 'Jennifer Collier': {'Jennifer Collier': {'W_Completeren aanvraag': 408, 'W_Valideren aanvraag': 155, 'W_Nabellen offertes': 401, 'W_Afhandelen leads': 101, 'W_Nabellen incomplete dossiers': 217, 'W_Beoordelen fraude': 3}}, 'Kevin Schultz': {'Kevin Schultz': {'W_Beoordelen fraude': 5, 'W_Completeren aanvraag': 422, 'W_Nabellen offertes': 423, 'W_Valideren aanvraag': 155, 'W_Nabellen incomplete dossiers': 173, 'W_Afhandelen leads': 118}}}

The contribution of activities according to the organization mining are as follows:

- W_Afhandelen leads = 8%
- W_Completeren aanvraag = 33%
- W_Nabellen offertes = 31%
- W_Valideren aanvraag = 12%
- W_Beoordelen fraude = 1%
- W_Nabellen incomplete dossiers = 16%

CPM

- Number of events: 11677
- Number of unique cases: 1731
- The number of resources is: 56

- Total duration: 296 days

Figure 90 – BPI 2012 CPM Dataset EDA Activities

Figure 91 – BPI 2012 CPM Dataset EDA Resources

Figure 92 - – BPI 2012 CPM Dataset BPMN Model

Figure 93 – BPI 2012 CPM Dataset DFG

Figure 94 – BPI 2012 CPM Dataset Heuristics

Custom GA

Figure 95 – BPI 2012 Custom GA Dataset Implementation

```
BPI2012 = namedtuple( 'BPI2012', ['name','days_to_finish', 'percentage'])

max_days = 296

things = [

          BPI2012('W_Afhandelen leads', 4 , 8),
          BPI2012('W_Completeren aanvraag',32 , 33),
          BPI2012('W_Nabellen offertes',91 , 31),
          BPI2012('W_Valideren aanvraag',85 , 12),
          BPI2012('W_Beoordelen fraude',34 , 1),
          BPI2012('W_Nabellen incomplete dossiers',84 , 16),
        ]
```

Figure 96 – BPI 2012 Custom GA Dataset Implementation result

```
number of generations: 0
time: 0.0003209114074707031s
best solution: ['W_Afhandelen leads', 'W_Nabellen offertes', 'W_Beoordelen fraude', 'W_Nabellen incomplete dossiers']
```

- Number of events: 40551
- Number of unique cases: 8441
- The number of resources is: 56

Figure 97 – BPI 2012 Custom GA Dataset EDA Activities

Figure 98 – BPI 2012 Custom GA Dataset EDA Resources

Figure 99 – BPI 2012 Custom GA Dataset BPMN Model

Figure 100 – BPI 2012 Custom GA Dataset DFG

Figure 101 – BPI 2012 Custom GA Dataset Heuristic

EasyGA

Figure 102 – BPI 2012 EasyGA Dataset Implementation

```
Current Generation      : 10000
Chromosome - 0 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 1 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 2 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 3 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 4 [1][1][1][1][0][0] / Fitness = 84
Chromosome - 5 [1][1][0][1][0][1] / Fitness = 69
```

- Number of events: 53111
- Number of unique cases: 9658
- The number of resources is: 56

Figure 103 – BPI 2012 EasyGA Dataset EDA Activities

Figure 104 – BPI 2012 EasyGA Dataset EDA Resources

Figure 105 – BPI 2012 EasyGA Dataset BPMN Model

Figure 106 – BPI 2012 EasyGA Dataset DFG

Figure 107 – BPI 2012 EasyGA Dataset Heuristic

EasyGA w/ BPMN Modelling

- Number of events: 53111
- Number of unique cases: 9658
- The number of resources is: 56

Figure 108 – BPI 2012 Dataset EasyGA w/ BPMN Modelling Implementation

```
Current Generation      : 10000
Chromosome - 0 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 1 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 2 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 3 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 4 [1][1][1][1][0][1] / Fitness = 100
Chromosome - 5 [1][1][1][1][0][0] / Fitness = 84
```

Figure 109 – BPI 2012 Dataset EasyGA w/ BPMN Modelling EDA Activities

Figure 110 – BPI 2012 Dataset EasyGA w/ BPMN Modelling EDA Resources

Figure 111 – BPI 2012 Dataset EasyGA w/ BPMN Modelling BPMN Model

Figure 112 – BPI 2012 Dataset EasyGA w/ BPMN Modelling DFG

Figure 113 – BPI 2012 Dataset EasyGA w/ BPMN Modelling Heuristic

Result

Conformance checking is a technique used to check process compliance by comparing event logs for a discovered process with the existing reference model (target model) of the same process. This technique is used to determine whether the target process corresponds to the actual process, highlighting deviations between the two (Appian, 2022).
Conformance checking compares the discovered model with a reference or target model, which can be visualized as a superimposition of the two models. While it's ideal to see a high degree of similarity between the basic activities in both models, this isn't always the case. Process workflows can include extra, unnecessary steps, and steps can be missed or performed in the wrong order.

Conformance checking is designed to identify these cases (Appian, 2022). The discovered model for this research is the generated model from each technique and the reference model is the original dataset. The discovered model will be compared to the reference model to see the deviation in the activities.
There are four qualities with conflicting criteria when it comes to conformance checking. The four conflicting criteria are referred to as the *four main quality dimensions.* They are fitness, simplicity, precision, and generalization (van der Aalst, 2016). According to Mlwiki.com, the four criteria are defined as follows:

1. Fitness: the discovered network should allow the behavior seen in the logs
    - tries to replay each sequence from the logs
    - so, for each case, we see if this sequence is allowed
    - if each sequence can be replayed - 100% fitness'
2. Precision: the discovered network should not allow the behavior not seen in the logs
    - too precise → bad generalization
    - play out the model, capture logs
    - see what's generated and compare it to the original logs
    - is it far? (Need to have some distance measure)
    - if produced logs ⊆ original logs --> then we have 100% precision
3. Generalization: the discovered model should generalize the behavior seen in the logs
4. Simplicity: it should be as simple as possible

Conformance checking is used to test the generated models. Conformance checking ensures all process deviations from the target process or reference model are identified, minimizing the risk of audit problems or legal violations. Process deviations often require more resources than the target process or may harm product or service quality. As a result, most process deviations lead to financial loss, so companies have a great interest in identifying unplanned process sequences and introducing appropriate preventive measures (Appian, 2022). There are 7 steps involved.

1. Convert the obtained models to a Petri net
2. Check obtained model's soundness
3. Check obtained models' fitness
    1. Token-based replay
    2. Alignment

4. Check obtained models' precision,
    1. Token-based replay
    2. Alignment
5. Check obtained models' diagnostics
6. Check obtained models' Generalization,
7. Check obtained models' Simplicity

Soundness

Workflow nets (WF-nets) are a subclass of Petri nets intending to model the workflow of process activities. The WF-net transitions are assigned to tasks or activities, and places are assigned to the pre/post conditions. The WF-nets have additional structural and operational requirements, mainly the addition of a single input (source) place with no previous transitions, and an output place (sink) with no following transitions. Accordingly, start and termination markings can be defined that represent the process status. WF-nets have the soundness property indicating that a process with a start marking of k tokens in its source place, can reach the termination state marking with k tokens in its sink place (defined as k-sound WF-net). Additionally, all the transitions in the process could fire (i.e., for each transition there is a reachable state in which the transition is enabled). A general sound (G-sound) WF-net is defined as being a k-sound for every k > 0 (Wikipedia 2022).

Figure 114 – Generated Models Soundness

The soundness for each implementation are as follows:

Critical Path Method
Input is ok
Petri Net is a workflow net.
Every place is covered by s-components
There are no dead tasks
All tasks are live

Custom Genetic Algorithm
Input is ok
Petri Net is a workflow net.
Every place is covered by s-components
There are no dead tasks
All tasks are live

EasyGA
Input is ok
Petri Net is a workflow net.
Every place is covered by s-components
There are no dead tasks
All tasks are live

EasyGA w/ BPMN Modelling
Input is ok
Petri Net is a workflow net.
Every place is covered by s-components
There are no dead tasks
All tasks are live

Here we can see that all the different methodologies' generated models for this dataset are sound. This means that all the generated models are sound.

Fitness

The fitness dimension has two types of tests. The two types of tests are token-based replay and alignment. PM4PY's *fitness_token_based_replay()* function is used to determine the token-based replay. The function returns the percentage of traces that are completely fit is returned, along with a fitness value that is calculated.
Token-based replay matches a trace and a Petri net model, starting from the initial place, to discover which transitions are executed and in which places we have remaining or missing tokens for the given process instance. Token-based replay is useful for Conformance Checking: indeed, a trace is fitting according to the model if, during its execution, the transitions can be fired without the need to insert any missing token. If the reaching of the final marking is imposed, then a trace is fitting if it reaches the final marking without any missing or remaining tokens (Conformance Fitness Token Based Replay Documentation, n.d.).

It can go across hidden transitions (calculating shortest paths between places) and can be used with any Petri net model with unique visible transitions and hidden transitions. When a visible transition needs to be fired and not all places in the preset are provided with the correct number of tokens, starting from the current marking it is checked if for someplace there is a sequence of hidden transitions that could be fired to enable the visible transition. The hidden transitions are then fired and a marking that permits to enable of the visible transition is reached (Conformance Fitness Token Based Replay Documentation, n.d.)
The *fitness_alignments()* function evaluates the fitness between an event log and a Petri net using alignments. The official Conformance Fitness Alignments Documentation states that " Alignment-based replay aims to find one of the best alignments between the trace and the model. For each trace, the output of an alignment is a list of couples where the first element is an event (of the trace) or » and the second element is a transition (of the model) or » ".
For each couple, the following classification could be provided:

- Sync moves: the classification of the event corresponds to the transition label; in this case, both the trace and the model advance in the same way during the replay.
- Move on the log: for couples where the second element is », it corresponds to a replay move in the trace that is not mimicked in the model. This kind of move is unfit and signals a deviation between the trace and the model.
- Move on model: for couples where the first element is », it corresponds to a replay move in the model that is not mimicked in the trace. For moves on the model, we can have the following distinction:
    - Moves on a model involving hidden transitions: in this case, even if it is not a sync move, the move is fit.
    - Moves on a model not involving hidden transitions: in this case, the move is unfit and signals a deviation between the trace and the model.

(Conformance Fitness Alignments Documentation, n.d.)

The calculation of replay fitness aims to calculate how much of the behavior in the log is admitted by the process model. They propose two methods to calculate replay fitness, based on token-based replay and alignments respectively. For alignments, the percentage of traces that are completely fit is returned, along with a fitness value that is calculated as the average of the fitness values of the single traces (Conformance Fitness Alignments Documentation, n.d.).

Figure 115 – Fitness Review Example Large

	Token Based Replay Percentage Fit Traces	Token Based Replay Average Trace Fitness	Token Based Replay Log Fitness	Alignment Based Replay Percentage Fit Traces	Alignment Based Replay Average Trace Fitness	Alignment Based Replay Log Fitness
EasyGA w/ BPMN Modelling	100	1	1	0	0.3296	0.2929
EasyGA	100	1	1	0	0.2953	0.2874
Custom Genetic Algortihm	100	1	1	0	0.4413	0.3941
Critical Path Method	100	1	1	0	0.5499	0.511

Figure 116 - Fitness Helpdesk

	Token Based Replay Percentage Fit Traces	Token Based Replay Average Trace Fitness	Token Based Replay Log Fitness	Alignment Based Replay Percentage Fit Traces	Alignment Based Replay Average Trace Fitness	Alignment Based Replay Log Fitness
EasyGA w/ BPMN Modelling	100	1	1	1.07	0.7031	0.6933
EasyGA	100	1	1	1.09	0.7082	0.6995
Custom Genetic Algortihm	100	1	1	0	0.5172	0.5286
Critical Path Method	100	1	1	0	0.2303	0.2209

Figure 117 - Fitness BPI 2012

	Token Based Replay Percentage Fit Traces	Token Based Replay Average Trace Fitness	Token Based Replay Log Fitness	Alignment Based Replay Percentage Fit Traces	Alignment Based Replay Average Trace Fitness	Alignment Based Replay Log Fitness
EasyGA w/ BPMN Modelling	100	1	1	0.6677	0.8641	0.7332
EasyGA	100	1	1	0.6677	0.8641	0.7332
Custom Genetic Algortihm	100	1	1	0.2372	0.5884	0.5598
Critical Path Method	100	1	1	0.5281	0.067	0.1612

In all datasets, the token-based relay percentages and token-based average trace fitness are all 100 and 1-to-1. All the traces in the generated models completely fit their original dataset. In addition, the average of the single traces in the generated models fits completely with the one in their original datasets.

In the Alignment-Based type, the best-performing methodologies are EasyGA and EasyGA w/ BPMN Modelling. This is followed by the Custom Genetic Algorithm. Lastly, CPM performed the least with the alignment test. However, this is only true for 2 out of 3 datasets. The total inversed is shown in the Review Example Large datasets, where CPM scored the highest. In hindsight looking at the result of CPM for each dataset, the critical path for Review Large Example dataset makes a lot more sense compared to the one found in the two other datasets. This will need to be further investigated to find the cause and root of the difference.

Precision

The *precision_token_based_replay()* function was used to calculate the precision of token-based replay. The precision_alignments() function was used to calculate the alignment way. In this approach, the different prefixes of the log are replayed (whether possible) on the model. At the reached marking, the set of transitions that are enabled in the process model is compared with the set of activities that follow the prefix. The more the sets are different, the more the precision value is low. The more the sets are similar, the more the precision value is high (*Conformance Precision Token Based Replay Documentation*, n.d.).

Figure 118 - Precision for Review Example Large Dataset

Figure 119 - Precision for Helpdesk

Figure 120 - Precision for BPI 2012

Comparing the Token-Based Replay and Alignment-Based Replay, the latter performed better across all the implementations in 2 out of 3 datasets. Now, let's compare the different methodologies.

<u>Review Example Large</u>

By the provided definition of precision, the methodologies that are the most different (lower precision) compared to their original dataset are EasyGA's implementation and the custom GA. CPM had higher precision and differs less from its original dataset.

<u>Helpdesk</u>

The most precise models are the Genetic Algorithms and more specifically EasyGA w/ BPMN Modeling. The more specific and optimized the GA is, the more precise it becomes. The worst performance came from CPM in both Token-Based-Replay and Alignment-Based-Replay

<u>BPI 2012</u>

In this dataset, CPM performed the worst in Token-Based-Replay. However, it is performing on par with the other methodologies in the alignment test.

Diagnostics

The *conformance_diagnostics_alignments()* method applies the alignments algorithm between a log and a process model. The methods return the full alignment diagnostics. With each trace, a dictionary containing among the others the following information is associated:

- alignment: contains the alignment (sync moves, moves on a log, moves on model)
- cost: contains the cost of the alignment according to the provided cost function
- fitness: is equal to 1 if the trace is perfectly fitting.

(*Conformance Diagnostics Alignments Documentation*, n.d.)

Figure 121 - Diagnostics Review Example Large

	Cost	Visited States	Queued States	Traversed Arcs	LP Solved	Fitness	BWC
Critical Path Method	40019	40	111	117	25	0.7647	170010
Custom Genetic Algortihm	60013	31	86	89	9	0.6471	170008
EasyGA	90013	31	74	76	12	0.4	150008
EasyGA w/ BPMN Modelling	90008	26	66	68	15	0.4375	160007

Figure 122 - Diagnostics HelpDesk

	Cost	Visited States	Queued States	Traversed Arcs	LP Solved	Fitness	BWC
Critical Path Method	40008	35	64	83	11	0.199	50005
Custom Genetic Algortihm	20012	90	193	298	90	0.6	50004
EasyGA	90013	31	74	76	12	40	50008
EasyGA w/ BPMN Modelling	10022	293	735	1232	243	0.8	50005

Figure 123 - Diagnostics BPI 2012

	Cost	Visited States	Queued States	Traversed Arcs	LP Solved	Fitness	BWC
Critical Path Method	50005	10	17	17	1	0	50005
Custom Genetic Algortihm	20013	46	107	124	40	0.6	50006
EasyGA	10021	147	360	537	110	0.8	50005
EasyGA w/ BPMN Modelling	10015	62	162	208	76	0.8	50004

There seems to be an inverse relationship between an event log fitness and the cost associated with it. The more fit a process model is, the lower the cost associated with it.

Review Example Large

The fittest model is CPM. The least fit model is the two EasyGA implementations. Inversely, the two EasyGA implementations have the highest cost associated with them.

Helpdesk

The fittest model is EasyGA w/ BPMN modeling. The least fit model is CPM at nearly 0. Inversely, CPM has the highest cost and EasyGA w/ BPMN modeling has the fewest cost.

BPI 2012

In this dataset, CPM performed the worst. The two EasyGA implementations performed the best.

Generalization & Simplicity

Generalization describes the likelihood that future behavior will already be executable in the model. The notion of generalization is in some ways related to the notion of overfitting known from the field of data science. If a model is overfitting, it is likely to also support variations such as noise. Through this, it gets too specific and therefore fails to truly capture the underlying structure of the data. So, a model with good generalization aims to prevent overfitting.

One approach to measure generalization is to use alignment to see how frequently parts of the model are used. If all parts of the model are used frequently, most likely all future behavior is captured and therefore generalization is high. If there are parts of the model that are infrequently used, it is more likely that there is behavior that is not seen yet. This makes generalization low (Pohl).

A model has good simplicity if it lacks complexity and is easy to understand for a human. Measuring simplicity on the other hand is not a trivial matter. In literature, there are many different approaches to measuring this quality criterion. For example, some approaches take the size or the diameter of the model into account. Size can refer to the number of nodes in the model and diameter in this case refers to the length of the shortest path from a start node to an end node (Pohl).

Figure 124 - Generalization & Simplicity for Review Example Large

Figure 125 - Generalization & Simplicity for Helpdesk

Figure 126 - Generalization & Simplicity for BPI 2012

<u>Review Example Large</u>

Most methodologies are around the same generalization percentage but EasyGA w/ BPMN modeling with the lowest. Inversely, EasyGA w/ BPMN modeling has the highest simplicity. This makes sense as activities are simplified.

<u>Helpdesk</u>

The same comment can be made for this dataset as the Review Example Large unless the CPM had the least generalization.

<u>BPI 2012</u>

The dataset performed about the same as the other two datasets in generalization. However, CPM performed better in simplicity by 10%.

Median Case Duration

Figure 127 - Median Case Duration Review Example Large

Figure 128 - Median Case Duration Helpdesk

Figure 129 - Median Case Duration BPI 2012

All Case Duration

Figure 130 - Review Example Large All Case Duration

Figure 131 - Helpdesk All Case Duration

Figure 132 - BPI 2012 All Cases Duration

Review Example Large

In this dataset, CPM has the most case compared to the other methodologies. As the models became more specific to the activities, the fewer cases.

Helpdesk

In this dataset, CPM had the least cases from a mile away. The cases rapidly converge to their end very fast compared to the other method which gradually and steadily increases in days which each new case. The CPM looks like an outlier.

BPI 2012

The same comment can be made for this dataset, but with the caveat that the cases in EasyGA method last much longer than the one in EasyGA w/ BPMN Modelling

Questions

In this Section, I will be answering the questions that this research posed. The first question is that can PM4PY be utilized in an organizational commercial setting effectively? From this research, I do not believe so. Not in a large commercial setting at least. PM4PY is good with simple and non-complex event logs. An example of a complex event log is the BPI 2012 with its 56 resources. As seen before, it becomes impossible to make any sense or interpretation from the Social Network Analysis with that many people. An actual organization with hundreds of employees would not benefit from this feature. In addition, more established process mining has better integration with their tools whereas PM4PY has less integration, as the writing of this research.

Another reasoning developing a process discovery with a complex and not well-defined sequence of activities. The limitations come inherently from PM4PY lack of clarity in model presentations and limited visual functions. A positive to PM4PY is that the team is constantly updating and adding new features to the library.

The second question asked which techniques and implementations tested are the most suitable in a commercial enterprise; which is worst? In this research, the most successful technique has been EasyGA w/ BPMN modeling or EasyGA. Both options also reduced the number of events, resources, and activities. The Custom GA has not been performing any better and stays in the middle of being the worst or the best. CPM has performed greatly in the Review Example of Large datasets in conformance checking but still didn’t manage to reduce the number of events, resources, and activities compared to its competitors. One of its best achievements is to reduce the median case duration the most in all datasets. This is a statement in its ability to reduce the project or activity's total duration.

Conclusion

After extensive research, the collected data and results refuted my hypothesis. CPM did not perform the best. The model that performed the best is EasyGA with the BPMN modeling, followed by EasyGA and my custom GA. I concluded that there is a field of exploration in deploying optimization algorithms for process mining. However, the optimizations' algorithms must be handled with care and scrutiny in building the fitness of the genes. The new researchers must be careful in defining the parents, population, and genes as they can easily influence the output. GA has been performing must better in terms of outputting results that make sense and also finding a satisfactory optimal answer. However, this research is the beginning of the building block of using GA for process mining.

For future work, it is imperative to find a lower and more reasonable generation limit as this implementation had an "overkill" run so there can be overfitting on the dataset. In addition, there needs to be an exploration of the fitness limit and fitness function of GA with different parameters based on duration time, or resources involved. Another future work is to explore the nuances and limitations of building populations for larger datasets and complex relationships among genes.

Regarding CPM, future work needs to study more algorithmics and automated method to do CPM calculations and find dependencies. This research showed a manual way to calculate the time durations and establish activity dependencies. The manual can lead to human error such as wrong calculations or interpreting a model with a relationship that might not be accurate. The generated model from Process Discovery is not 100%, but rather gives a good indication of the relationships and some activities might seem connected due to some deviations in the model but not present in the actual day-to-day operation. Misinterpreting the model might lead to a new model that is not representative of the original. With that said, an exploration of algorithmic CPM is much appreciated.

PM4PY is not fully ready to be adopted in large commercial settings to implement GA or CPM. There are still some lacking features that hold it back. In addition, large and complex datasets take massive computational time to be executed especially when it comes to visualizing results. However, it is not the end for PM4PY. Founder Sebastiaan van Zelst and the PM4PY team are consistently pushing updates and new features for the open-source library. Still, this means that new updates can break someone's code. In my experience, the format and location of some modules were changed, and it caused some unexpected errors. had to find out that the repository was updated, and some features were deprecated or moved to different folders, so important certain functions would not work. Though it has a long way to be used for GA and CPM, the future for PM4PY is looking bright with the love of the team and new features being implemented. On the contrary, this means that sound codes are bound to break along the way.

In conclusion, this paper addressed the problem of inefficiencies in business processes using Process Mining. The paper refuted my hypothesis that CPM would be the best-performing model. CPM did perform well on the synthetic dataset. However, CPM did not perform well in the two real-life datasets. EasyGA and EasyGA w/ BPMN modeling performed quite well in those datasets. In addition, we concluded that fine-tuning the fitness function is necessary to

create more refined and tailored results. The research showed better performance in activity optimization and conformance checking.

Acknowledgments

I would like to acknowledge Zeanique L. Barber for the amazing help she provided and the great support system she has been throughout this research. She provided me with many links and references for me to learn about the different methods that I employed. Her extensive ability to research information in a quick time frame is very amazing. I hope to publish this research and have her as a contributor,

I would like to acknowledge Daniel Wilczak for his amazing help and contribution to teaching me about EasyGA. He took his time to develop a very quick example regarding my specific research. Thank you for taking the time to respond to me over Discord.

Table Tables of Content

Table 1 – Advantages and disadvantages of Critical Path Method 7
Table 2 - Advantages and disadvantages of Genetic Algorithms 9
Table 3- Advantages and disadvantages of BPMN Model 12
Table 4 - Advantages and Disadvantages of Heuristic Miner 12
Table 5 - Advantages and disadvantages of Direct Follow Graph 13

Figures Tables of Content

Figure 1 – Distribution of the Events over the Years for Review Example Large dataset..... 16
Figure 2 – Review Example Large dataset starting and ending activities 16
Figure 3 - Review Example Large dataset number of event and unique cases..... 17
Figure 4 - Review Example Large dataset activities 17
Figure 5 - Review Example Large dataset resources..... 17
Figure 6 - Review Example Large dataset dataframe..... 17
Figure 7 - Helpdesk dataset distribution of the events over the years 18
Figure 8 - Helpdesk dataset starting and ending activities 18
Figure 9 - Helpdesk dataset number of event and unique cases 19
Figure 10 - Helpdesk dataset resources 19
Figure 11 - activities 19
Figure 12 – BPI 2012 dataset distribution of the events over the years 20
Figure 13 - BPI 2012 dataset starting and ending activities 20
Figure 14 - BPI 2012 dataset number of event and unique cases 20
Figure 15 - BPI 2012 dataset resources 21
Figure 16 - BPI 2012 dataset activities..... 21
Figure 17 - Review Example Large Dataset EDA Activities 22
Figure 18 - Review Example Large Dataset EDA Resources 22
Figure 19 - Review Example Large Dataset BPMN Model 23
Figure 20 - Review Example Large Dataset Heuristics Model 23
Figure 21- Review Example Large Dataset DFG Model..... 24
Figure 22 - Review Example Large Dataset Handover of Work..... 25
Figure 23 - Review Example Large Dataset Subcontracting..... 25
Figure 24 - Review Example Large Dataset Working Together 26
Figure 25 - Review Example Large Dataset Similar Activity 26
Figure 26 - Review Example Large Dataset CPM EDA Activities..... 32
Figure 27 - Review Example Large Dataset CPM EDA Resource 32
Figure 28 - Review Example Large Dataset CPM DFG..... 33
Figure 29 - Review Example Large Dataset CPM BPMN Model..... 33
Figure 30 - Review Example Large Dataset CPM heuristics 34
Figure 31 - Review Example Large Dataset Custom GA Implementation 35
Figure 32 - Review Example Large Dataset Custom GA Implementation result..... 35
Figure 33 - Review Example Large Dataset Custom GA EDA Activities 36
Figure 34 - Review Example Large Dataset Custom GA EDA Resources 36
Figure 35 - Review Example Large Dataset Custom GA BPMN Model 37
Figure 36 - Review Example Large Dataset Custom GA DFG..... 37
Figure 37 - Review Example Large Dataset Custom GA Heuristics..... 38
Figure 38 - Review Example Large Dataset EasyGA Implementation 39
Figure 39 - Review Example Large Dataset EasyGA EDA Activities..... 40
Figure 40 - Review Example Large Dataset EasyGA EDA Resources..... 40
Figure 41 - Review Example Large Dataset EasyGA BPMN Model..... 41
Figure 42 - Review Example Large Dataset EasyGA DFG 41

Figure 43 - Review Example Large Dataset EasyGA Heuristics 41
Figure 44 - Review Example Large Dataset EasyGA w/ BPMN Modelling implementation 42
Figure 45 - Review Example Large Dataset EasyGA w/ BPMN Modelling BPMN Model..... 42
Figure 46 - Review Example Large Dataset EasyGA w/ BPMN Modelling DFG 43
Figure 47 - Review Example Large Dataset EasyGA w/ BPMN Modelling Heuristics 43
Figure 48 - Helpdesk Dataset EDA Activities..... 44
Figure 49 - Helpdesk Dataset EDA Resources..... 44
Figure 50 - Helpdesk Dataset BPMN Model..... 45
Figure 51 - Helpdesk Dataset DFG..... 46
Figure 52 - Helpdesk Dataset Heuristics 47
Figure 53 - Helpdesk Dataset Handover of Work 48
Figure 54 - Helpdesk Dataset Subcontracting 48
Figure 55 - Helpdesk Dataset Working Together..... 49
Figure 56 - Helpdesk Dataset Similar Activity..... 49
Figure 57 - Helpdesk Dataset CPM EDA Activities 58
Figure 58 - Helpdesk Dataset CPM EDA Resources..... 58
Figure 59 - Helpdesk Dataset CPM BPMN Model 59
Figure 60 - Helpdesk Dataset CPM DFG 59
Figure 61 - Helpdesk Dataset CPM Heuristic..... 60
Figure 62 - Helpdesk Dataset Custom GA Implementation 60
Figure 63 Helpdesk Dataset Custom GA Implementation result..... 61
Figure 64 - Helpdesk Dataset Custom GA EDA Activities..... 61
Figure 65 - Helpdesk Dataset Custom GA EDA Resources..... 62
Figure 66 - Helpdesk Dataset Custom GA BPMN Model..... 62
Figure 67 - Helpdesk Dataset Custom GA DFG 63
Figure 68 - Helpdesk Dataset Custom GA Heuristic..... 64
Figure 69 - Helpdesk Dataset EasyGA Implementation..... 65
Figure 70 - Helpdesk Dataset EasyGA EDA Activities 65
Figure 71 - Helpdesk Dataset EasyGA EDA Resources 66
Figure 72 - Helpdesk Dataset EasyGA BPMN Model 67
Figure 73 - Helpdesk Dataset EasyGA DFG 69
Figure 74 - Helpdesk Dataset EasyGA Heuristics..... 70
Figure 75 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model Implementation.... 71
Figure 76 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model EDA Activities..... 72
Figure 77 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model EDA Resources.... 72
Figure 78 - Helpdesk Dataset EasyGA w/ BPMN Modelling BPMN Model BPMN Model 73
Figure 79 - Helpdesk Dataset EasyGA w/ BPMN Modelling DFG 73
Figure 80 - Helpdesk Dataset EasyGA w/ BPMN Modelling A BPMN Model EDA Activities. 74
Figure 81 – BPI 2012 Dataset EDA Activities 75
Figure 82 – BPI 2012 Dataset EDA Resources 75
Figure 83 – BPI 2012 Dataset BPMN Model 76
Figure 84 – BPI 2012 Dataset DFG..... 78
Figure 85 – BPI 2012 Dataset Heuristics..... 79

Figure 86 – BPI 2012 Dataset EDA Handover of Work 80
Figure 87 – BPI 2012 Dataset EDA Subcontracting 80
Figure 88 – BPI 2012 Dataset EDA Working Together 81
Figure 89 – BPI 2012 Dataset EDA Similar Activity 81
Figure 90 – BPI 2012 CPM Dataset EDA Activities 96
Figure 91 – BPI 2012 CPM Dataset EDA Resources 96
Figure 92 - – BPI 2012 CPM Dataset BPMN Model 97
Figure 93 – BPI 2012 CPM Dataset DFG 97
Figure 94 – BPI 2012 CPM Dataset Heuristics 98
Figure 95 – BPI 2012 Custom GA Dataset Implementation 99
Figure 96 – BPI 2012 Custom GA Dataset Implementation result 99
Figure 97 – BPI 2012 Custom GA Dataset EDA Activities 99
Figure 98 – BPI 2012 Custom GA Dataset EDA Resources 100
Figure 99 – BPI 2012 Custom GA Dataset BPMN Model 100
Figure 100 – BPI 2012 Custom GA Dataset DFG 101
Figure 101 – BPI 2012 Custom GA Dataset Heuristic 102
Figure 102 – BPI 2012 EasyGA Dataset Implementation 103
Figure 103 – BPI 2012 EasyGA Dataset EDA Activities 104
Figure 104 – BPI 2012 EasyGA Dataset EDA Resources 104
Figure 105 – BPI 2012 EasyGA Dataset BPMN Model 105
Figure 106 – BPI 2012 EasyGA Dataset DFG 105
Figure 107 – BPI 2012 EasyGA Dataset Heuristic 106
Figure 108 – BPI 2012 Dataset EasyGA w/ BPMN Modelling Implementation 107
Figure 109 – BPI 2012 Dataset EasyGA w/ BPMN Modelling EDA Activities 108
Figure 110 – BPI 2012 Dataset EasyGA w/ BPMN Modelling EDA Resources 108
Figure 111 – BPI 2012 Dataset EasyGA w/ BPMN Modelling BPMN Model 109
Figure 112 – BPI 2012 Dataset EasyGA w/ BPMN Modelling DFG 110
Figure 113 – BPI 2012 Dataset EasyGA w/ BPMN Modelling Heuristic 111
Figure 114 – Generated Models Soundness 113
Figure 115 – Fitness Review Example Large 115
Figure 116 - Fitness Helpdesk 115
Figure 117 - Fitness BPI 2012 116
Figure 118 - Precision for Review Example Large Dataset 117
Figure 119 - Precision for Helpdesk 117
Figure 120 - Precision for BPI 2012 118
Figure 121 - Diagnostics Review Example Large 119
Figure 122 - Diagnostics HelpDesk 120
Figure 123 - Diagnostics BPI 2012 120
Figure 124 - Generalization & Simplicity for Review Example Large 122
Figure 125 - Generalization & Simplicity for Helpdesk 122
Figure 126 - Generalization & Simplicity for BPI 2012 123
Figure 127 - Median Case Duration Review Example Large 124
Figure 128 - Median Case Duration Helpdesk 124

Figure 129 - Median Case Duration BPI 2012 125
Figure 130 - Review Example Large All Case Duration 125
Figure 131 - Helpdesk All Case Duration 126
Figure 132 - BPI 2012 All Cases Duration 126

References

4TU.ResearchData. (2020, July 25). *Synthetic event logs - review example large.xes.gz*. Figshare. https://data.4tu.nl/articles/dataset/Synthetic_event_logs_-_review_example_large_xes_gz/12716609

A Systematic Review on Recommender Systems in Process Mining. (2022). *Concurrency and Computation: Practice and Experience*, *34*(26). https://doi.org/10.1002/cpe.7304

Appian. (2022). *Conformance Checking – The Process Mining Glossary*. Retrieved December 17, 2022, from https://appian.com/process-mining/conformance-checking.html

Burattin, A., Polyvyanyy, A., & Van Zelst, S. J. (2018). ICPMD 2019, ICPM Demo Track 2019: proceedings of the ICPM Demo Track 2019, co-located with 1st International Conference on Process Mining (ICPM 2019) : Aachen, Germany, June 24-26, 2019. *1st International Conference on Process Mining*.

Celonis Process Mining Software. (n.d.). Celonis. https://www.celonis.com/process-mining/what-is-process-mining/

Conformance Diagnostics Alignments Documentation. (n.d.). PM4PY. Retrieved December 18, 2022, from https://pm4py.fit.fraunhofer.de/static/assets/api/2.3.0/generated/pm4py.conformance.conformance_diagnostics_alignments.html

Conformance Diagnostics Token Based Replay Documentation. (n.d.). PM4PY. Retrieved December 18, 2022, from https://pm4py.fit.fraunhofer.de/static/assets/api/2.3.0/generated/pm4py.conformance.conformance_diagnostics_token_based_replay.html

Conformance Fitness Alignments Documentation. (n.d.). PM4PY. Retrieved December 18, 2022, from

https://pm4py.fit.fraunhofer.de/static/assets/api/2.3.0/generated/pm4py.conformance.fitness_alignments.html

Conformance Fitness Token Based Replay Documentation. (n.d.). PM4PY. Retrieved December 18, 2022, from https://pm4py.fit.fraunhofer.de/static/assets/api/2.3.0/generated/pm4py.conformance.fitness_token_based_replay.html

Conformance Precision Token Based Replay Documentation. (n.d.). PM4PY. Retrieved December 18, 2022, from https://pm4py.fit.fraunhofer.de/static/assets/api/2.3.0/generated/pm4py.conformance.precision_token_based_replay.html

Corporate Finance Institute. (2022, December 1). *Null Hypothesis*. https://corporatefinanceinstitute.com/resources/data-science/null-hypothesis-2/

Critical Path in Project Management – Definition and Method. (2021, October 13). PeerSpot. https://www.peerspot.com/articles/critical-path-in-project-management-definition-and-method

Daniel Wilczak. (2021, April 6). *Getting Started with EasyGA. Genetic Algorithm made Easy.* [Video]. YouTube. https://www.youtube.com/watch?v=jbuDKwIiYBw

Editor. (2021, June 11). *Process Mining Explained: Techniques, Applications, and Challenges*. AltexSoft. https://www.altexsoft.com/blog/process-mining/

Effendi, Y. A., & Sarno, R. (2017). Non-linear optimization of critical path method. *2017 3rd International Conference on Science in Information Technology (ICSITech)*. https://doi.org/10.1109/icsitech.2017.8257091

Faker's documentation. (n.d.). https://faker.readthedocs.io/en/master/

Figshare. (n.d.). Figshare. Retrieved December 19, 2022, from https://figshare.com/

Fraunhofer Institute for Applied Information Technology. (n.d.). *PM4Py - Process Mining for Python*. https://pm4py.fit.fraunhofer.de/documentation

Gerend, J. G. (2022, August 18). *Windows commands*. Microsoft Learn. https://learn.microsoft.com/en-us/windows-server/administration/windows-commands/windows-commands

Gibbons, S. (2020, December 15). *5 Inefficiencies Hiding In Your Everyday Operations*. Forbes. https://www.forbes.com/sites/serenitygibbons/2020/12/15/5-inefficiencies-hiding-in-your-everyday-operations/?sh=22a2d1d76141

Git. (n.d.). https://git-scm.com/

GitHub. (n.d.). DBpedia. https://dbpedia.org/page/GitHub

Google Colab. (n.d.). https://research.google.com/colaboratory/faq.html

Homepage Wil van der Aalst. (n.d.). Retrieved December 19, 2022, from http://www.padsweb.rwth-aachen.de/wvdaalst/

Hussain, M. S. (2022, August 30). *Process mining: Introducing event log mining*. Data Science Dojo. https://datasciencedojo.com/blog/process-mining-event-log-mining/

Low, W. Z., De Weerdt, J., Wynn, M. T., ter Hofstede, A. H. M., van der Aalst, W. M. P., & vanden Broucke, S. (2014). Perturbing event logs to identify cost reduction opportunities: A genetic algorithm-based approach. *2014 IEEE Congress on Evolutionary Computation (CEC)*. https://doi.org/10.1109/cec.2014.6900465

Mallawaarachchi, V. (2020, March 1). *Introduction to Genetic Algorithms — Including Example Code*. Medium. https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3

Nemerever, M. N. (2019, March 11). *Git, GitHub, & Workflow Fundamentals*. Dev.to. https://dev.to/mollynem/git-github--workflow-fundamentals-5496

Neu, D. A., Lahann, J., & Fettke, P. (2021). A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artificial Intelligence Review*, *55*(2), 801–827. https://doi.org/10.1007/s10462-021-09960-8

Oliveira, H. D., Augusto, V., Jouaneton, B., Lamarsalle, L., Prodel, M., & Xie, X. (2020). An optimization-based process mining approach for explainable classification of timed event logs. *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. https://doi.org/10.1109/case48305.2020.9216841

Process Mining - ML Wiki. (2015, July 5). Retrieved December 18, 2022, from http://mlwiki.org/index.php/Process_Mining

Process Mining: Data Science in Action by Wil van der Aalst (2016-05-21). (n.d.). Springer.

Process Mining for Python. (2020, November 5). *pm4py tutorials - tutorial #1: What is Process Mining?* [Video]. YouTube. https://www.youtube.com/watch?v=XLHtvt36g6U

Process Mining Introduction. (n.d.). ProcessMining.org. Retrieved December 19, 2022, from https://www.processmining.org/introduction.html

Sarojag. (2021, July 10). *An Introduction to Process Mining and Algorithms*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/07/process-mining-and-algorithms-an-introduction/

Seeliger, A., Nolle, T., & Mühlhäuser, M. (2018). ProcessExplorer: An Interactive Visual Recommendation System for Process Mining. *Semantic Scholar*. https://www.semanticscholar.org/paper/ProcessExplorer%3A-An-Interactive-Visual-System-for-Seeliger-

Nolle/eb1719f1dab7f48cfeb3a4ac49ddf0de955468a9?sort=relevance&citationIntent=methodology

Shah, P. (2021, April 17). *Hello World! - GitHub Action*. https://www.linkedin.com/pulse/hello-world-github-action-parthiv-shah

Sharma, A. (2020, December 29). *Free GPUs for Everyone! Get Started with Google Colab for Machine Learning and Deep Learning*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/

Team Asana. (n.d.). *Critical Path Method: How to Use CPM for Project Management* • [Video]. Asana. https://asana.com/resources/critical-path-method

The ABCs of the Critical Path Method. (2014, August 1). Harvard Business Review. https://hbr.org/1963/09/the-abcs-of-the-critical-path-method

Usmani, F. P. (2022, October 1). *Critical Path Method (CPM) in Project Management* |. Fahad Usmani. https://pmstudycircle.com/critical-path-method-cpm-in-project-management/

van der Aalst, W. (2016). Process Mining. *Springer Link*. https://doi.org/10.1007/978-3-662-49851-4

van der Aalst, W. M. P., de Medeiros, A. K. A., & Weijters, A. J. M. M. (2005). Genetic Process Mining. *Applications and Theory of Petri Nets 2005*, 48–69. https://doi.org/10.1007/11494744_5

VISPM. (2022, July 16). PyPI. https://pypi.org/project/vispm/

Wikipedia contributors. (2022, December 17). *Petri net*. Wikipedia. https://en.wikipedia.org/wiki/Petri_net

Wilczak, D. W., [Danielwilczak101], & Nguyen, J. N. (n.d.). *GitHub - danielwilczak101/EasyGA: EasyGA is a python package designed to provide an easy-to-use Genetic Algorithm. The package is designed to work right out of the box, while also allowing the user to customize features as they see fit.* GitHub. https://github.com/danielwilczak101/EasyGA

Yang, Sen, Y., Ivan, M., Yingying, C., Anand, S. D., Hui, X., & Rutgers University: School of Graduate Studies. (2019). *Applied Process Mining, Recommendation, and Visual Analytics* [Ph.D. Dissertation]. Rutgers, The State University of New Jersey.