Mater Thesis
August, 2022

Discovering Activation by Process Mining: A Case of B2B SaaS Customers

Wan-Hsuan Chiang

Interaction Technology
Faculty of Electrical Engineering, Mathematics and Computer Science

Examination Committee

Dr. F.A. Bukhsh (Faiza)
Dr. S. Wang (Shenghui)

UNIVERSITY OF TWENTE.

CONTENTS

Abstract **12**

1 Introduction **13**

1.1 Aha! Moment 13

1.2 Research Questions 14

1.3 Techniques 14

1.3.1 Why Process Mining 14

1.3.2 Process Discovery Algorithm 15

1.3.3 Tools 15

1.4 Research Plan 15

1.4.1 CRISP-DM Framework 15

1.4.2 Research Plan 17

1.5 Outline 18

2 Process Mining Foundation Related to the Thesis **19**

2.1 Discovery, Conformance, and Enhancement 19

2.2 Perspectives 20

2.3 Process Discovery Algorithm 20

2.3.1 Heuristic Miner 20

2.3.2 Fuzzy Miner 24

2.4 Process Mining in Action 24

3 Background and Related Works **26**

3.1 Commercial Analytics Tools 26

3.2 Web Mining 27

3.3 Related Work of Process Mining 27

3.3.1 Process Mining in Navigation Behavior 27

3.3.2 Process Mining in Learning Behavior 28

3.3.3 Process Mining v.s. Customer Journey 28

3.3.4 Process Mining v.s. Human-Computer Interaction 28

3.4 Summary 29

4 Business and Data Understanding **30**

4.1 Business Overview 30

4.1.1 Software-as-a-Service 30

4.1.2 Zendesk 30

4.1.3 Kaizo App as a Zendesk Extension 31

4.2 Stakeholder Identification 31

4.2.1 External Stakeholders 31

4.2.2 Internal Stakeholder 33

4.3 Product Solutions and Research Pillars 34

4.3.1 Transform Solutions to Pillars 34

4.3.2 Domain Object and Milestone Events 34

4.3.3 Challenge Identification 36

4.4 Weekly Based Activities 36

4.5 Interface Identification 37

4.6 Data Description 39

4.7 Summary 40

5 Data Preparation **41**

5.1 Data Tables 41

5.1.1 Events 41

5.1.2 Users 42

5.1.3 Tables Specific for Pillars 44

5.2 Data Collection, Cleaning, and Construction 46

5.2.1 Data Collection 46

5.2.2 Trace Identification 47

5.2.3 Trial Cleaning 47

5.2.4 Trace Re-identification 49

5.3 Summary 51

6 Retention Clustering **52**

6.1 Retention Identification 52

6.2 Data Construction 53

6.2.1 Domain Object 53

6.2.2 Navigation Behavior in Quality Assurance 54

6.2.3 User Status 54

6.3 Event Classification and Data Cleaning 55

6.3.1 The Focus on Domain Objects 56

6.3.2 Specific Flows 56

6.3.3 Derive New Event Node 56

6.3.4 Duplicated Events 57

6.4 Data Preparation for Modelling 57

6.5 Pillar: Quality Assurance 58

6.5.1 Manager 58

6.5.2 Agent 60

6.5.3 Domain Object Lifecycle 61
6.6 Pillar: Missions 61
6.6.1 Manager 61
6.6.2 Agent 63
6.7 Pillar: Performance Coaching 66
6.7.1 Manager 66
6.7.2 Agent 69
6.7.3 Domain Object Lifecycle 71
6.8 Summary 73

7 Discover the Aha! Moment **74**

7.1 Data Preparation 74
7.1.1 Data Construction 74
7.1.2 Account Selection 75
7.1.3 Test Modelling 75
7.1.4 Trials 75
7.1.5 Data Re-cleaning 78
7.2 Activation Behavior Modelling 78
7.2.1 Data Preparation and Model Configuration 78
7.2.2 Pillar: Quality Assurance 79
7.2.3 Pillar: Missions 81
7.2.4 Pillar: Performance Coaching 82
7.3 Summary and Feedback 83

8 Conclusion **84**

8.1 Result 84
8.2 Contribution 84

8.2.1 Academics 84
8.2.2 Business 85
8.3 Discussion and Future Work 86
8.3.1 Pillar Interaction 86
8.3.2 Process Cube 86
8.3.3 Discovering Other Customer Phases 86

References **87**

A Event Classification **93**

B Test Fuzzy Models for Data Cleaning **99**

C Retention Pattern Identification **103**

D Agents with Missions **107**

E Models in Discovering Aha Moment **108**

E.1 First Trial 108
E.2 Second Trial 109
E.3 Fourth Trial 112
E.4 Pure Pillar Retention Accounts 115
E.5 Last Trial 116

List of Figures

1.1 Customer Life cycle applying AARRR [42] 13

1.2 CRISP-DM Standard Proposed by Wirth and Hipp [69] 16

1.3 Research Plan 17

2.1 Example output of Heuristic Miner from Direct-follows Graph to Causal Net (from Kaizo dataset) 21

2.2 Dependency Graph Example 22

2.3 C-Net Example from Aalst et al. [1] 23

2.4 ProM Screenshots 25

3.1 The User Flow Reporting in GA4 26

4.1 Interaction between each instance 31

4.2 Default Roles 32

4.3 Stakeholders and Client Lifecycle 33

4.4 Product Solutions and the Identified Pillars 34

4.5 Interaction Instrument [3] 35

4.6 Kaizo App's Landing Page 37

4.7 The Screenshot of Each Pillar 38

5.1 The Rule of Assigning Role Schemes in Single User 43

5.2 **Quality Assurance** Related Data Tables 44

5.3 **Performance Coaching** Related Data Tables 45

5.4 How to Select Installation Time 46
5.5 Example of the Broken Model 47
5.6 Assign `module` to Each Event 49
5.7 The Percentage of Expected Start Events and the Module Time Difference . . . 50

6.1 **Quality Assurance**: Manager's Behavior 59
6.2 **Quality Assurance**: Agent's Behavior 60
6.3 Tickets with Rating 61
6.4 The Example of Repeated Pattern 62
6.5 **Missions**: Manager's Behavior 62
6.6 **Missions**: Agent's Behavior 65
6.7 **Missions**: Agent's Behavior After Activation (Part of Directly-follows Graph) . . . 66
6.8 The Loops between Nodes 67
6.9 **Performance Coaching**: Manager's Behavior 69
6.10 **Performance Coaching**: Agents' Behavior 70
6.11 **Performance Coaching**: Domain Object Lifecycle 72

7.1 Weekly Domain Activity, by Role 75
7.2 **Quality Assurance** Retention Accounts in Activation Phase 80
7.3 **Missions** Retention Accounts in Activation Phase 81
7.4 **Performance Coaching** Retention Accounts in Activation Phase 82

B.1 Test Model before Data Cleaning (User as Trace) 99
B.2 Test Model before Data Cleaning (Session as Trace) 100
B.3 Test Model after Removing Invalid Sessions 101
B.4 Test Model Applying `eventCategory` as Event 102

C.1 Sessions contains Managers' Milestone Events in **Quality Assurance** 103

C.2 Sessions contains Managers' Milestone Events in **Missions** 104
C.3 Sessions contains Agents' Milestone Events in **Missions** 105
C.4 Sessions contains Managers' Milestone Events in **Performance Coaching** . . . 106

D.1 The Process Model of Agents who Activated the Missions 107

E.1 The Example Dependency Graph of the First Trial Modelling 108
E.2 The Activation Phase of the Account with High **Quality Assurance** Retention . . 109
E.3 The Activation Phase of the Account with High **Missions** Retention 110
E.4 The Activation Phase of the Account with High **Performance Coaching** Retention 111
E.5 Activation Phase of Accounts with **Quality Assurance** Retention 112
E.6 Activation Phase of Accounts with **Missions** Retention 113
E.7 Activation Phase of Accounts with **Performance Coaching** Retention 114
E.8 First Week Behavior of Accounts without Retention 114
E.9 Mission 115
E.10 Coaching 115
E.11 **Quality Assurance**-Retention Account Activities 116
E.12 **Missions**-Retention Account Activities 117
E.13 **Performance Coaching**-Retention Account Activities 118

List of Tables

1.1 Tools Adopted in This Thesis 15

1.2 Research Tasks 18

2.1 Example of Calculating Significance in Dependency Graph 21

4.1 Terms Used in This Paper 32

4.2 Domain Objects 35

4.3 Milestone Action on Domain Object 35

4.4 UI Modules 39

5.1 The Fields Used in Data Table `events` Table 41

5.2 Example of User Table 42

5.3 Example of Categorize User Role Scheme 43

5.4 The Fields Used in Data Table `latest_missions` Table 45

5.5 Event Attributes 51

6.1 Milestone Events and Test Models 52

6.2 The ID of domain objects 53

6.3 Example of An User's Activity on Ticket 53

6.4 Example of Deriving `has_qa_info` 54

6.5 Example of Folding Duplicated Events 57

6.6 Bindings of Manager's Milestone Event (**Quality Assurance**) 58

6.7 Bindings of Focused Nodes in Manager's Behavior Model (**Missions**) 63

6.8 Example of Trace Selection 63

6.9 Bindings of Focused Nodes in Agent's Behavior Model (**Missions**) 64

6.10 Bindings of Focused Nodes in Manager's Behavior Model (**Performance Coaching**) 68

6.11 Bindings of Focused Events in Agent's Behavior Model (**Performance Coaching**) 71

6.12 Process Duration of Domain Object Life Cycle 72

6.13 Retention Models 73

6.14 Milestone Actions after Retention Analysis 73

7.1 Example of Account Event Sequence Selection 78

A.1 Event Classification 93

ABSTRACT

This research aims to find customers' "aha moment" through process mining. Aha moment is when users realize the value of using the software product, which is a key to driving revenue, especially for B2B SaaS vendors. According to the AARRR model, the aha moment can refer to activation, and the following customer phase is retention. Since the customers in retention are obligated to experience activation before, the research first identifies milestone actions in terms of product features and user roles to cluster the retention customers. After that, the data of the clustered customers from the time before they move to the retention phase can be used to discover the aha moment. The event log analysis is discussed based on multiple dimensions: product solution, time, and user roles. The research mainly applies the process mining technique, heuristic miner, to discover the customer's behavior patterns. Apart from marketing funnels, this project also involves the concept of human-computer interaction for event classification and data cleaning, which is practical for cleaning UI logs. The discovered processes and aha moment can guide future product development and value proposition re-proposal.

1 INTRODUCTION

This research is conducted for Kaizo[1], the software company improving the performance of the customer relationship management (CRM) team in other businesses. It is a B2B company with Software-as-a-Service (SaaS) deployment model [44]. The product developed and owned by Kaizo shares the same name as the company.

1.1 Aha! Moment

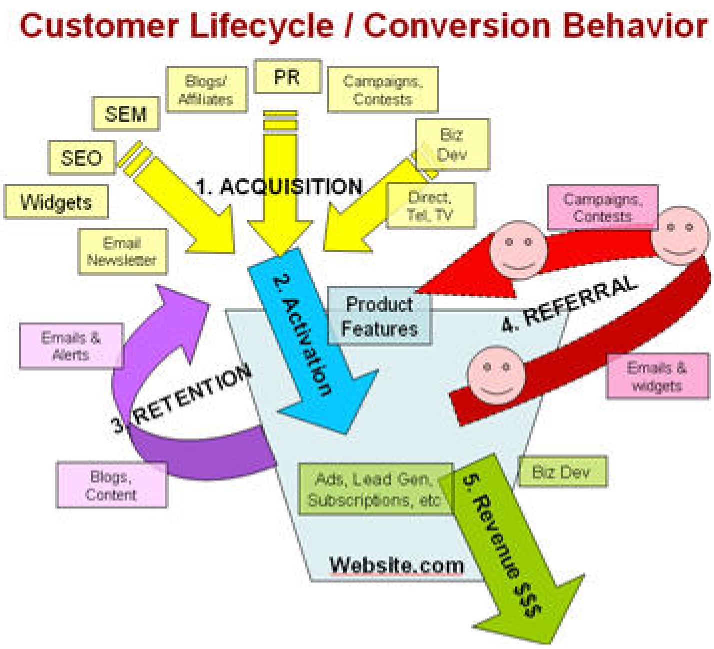


Figure 1.1: Customer Life cycle applying AARRR [42]

AARRR model is a widely spread framework in sales and marketing, which is first proposed by venture capitalist Dave McClure [11]. The model describes the users into five steps: acquisition, activation, retention, revenue, and referral (See Figure 1.1). The activation starts from a happy first visit [42], in other words, the onboarding step of the funnel. It is also known as "Aha! moment:" initially, the term is defined as an insight that represents a sudden cognitive change [7, 56], and was adapted to product development, indicating the time when the users realize the value of the product.

The following customer phase of activation is retention. In this phase, businesses consider

[1] https://kaizo.com/

whether users continuously engage with the product. Therefore, the frequency concept is involved in order to calculate repeated behaviors.

Meanwhile, B2B SaaS vendor needs to show the value to the clients. B2B users has a strong focus on value-in-use [32]. The value-in-use defines the activation or onboarding experience of B2B clients. Considering the customer journey as a funnel, the activation is a fatal step leading to retention and revenue. To be more specific, most of the SaaS providers involve such concepts to create price strategies. For example, Google Cloud provides limited free credit and free monthly usage [20]; the basic plan of Zoom is free with time, and attendee limits [74]. The free version allows the users to adopt and test the service. Users who experience the value or with higher needs will further be converted to paid customers. In general, a B2B SaaS vendor's strategy, especially for a startup, is clear: showing the value of the product to new users in order to push users forward to retention and payment.

1.2 Research Questions

However, according to Lemke et al. [32]'s interviews, the value-in-use is construed by customers, which is not equivalent to service quality. In other words, the value proposition presented by the business and the value customers feel can be different. Therefore, the best strategy is to extract value propositions from users.

This research aims to find users' "Aha! moment" by applying Kaizo's business context. As companies are used to collecting data from the software product, the event logs generated by Kaizo's clients are used. To find the "Aha! moment," the customer phase needs to be defined first. The research, inspired by the AARRR model, assumes that all the clients in the retention phase had experienced "Aha! moment" before. Only if the clients walk through the activation phase will they retain. Therefore, the research questions can be defined as follows:

RQ1 - What is the behavior patterns of the clients in the retention phase in Kaizo?

RQ2 - What is the "Aha! moment" of Kaizo clients, leading to retention?

The questions can be divided into multiple dimensions, such as user permissions, functions in the Kaizo application, and time. The dimension will be further discussed in the following chapters.

1.3 Techniques

1.3.1 Why Process Mining

Many techniques to analyze user behaviors can be roughly categorized into the quantitative and qualitative analysis. Qualitative research, such as user interviews or focus groups, is costly, which is not ideal for resource-limited startups. Quantitative analysis is widely used in today's business world. Many commercial analytic tools are ready to use and easy to apply. However, tools such as Google Analytics require an entire understanding of the product structure and funnel in advance, which is too simplified [58] for exploring user behaviors (See Section 3.1).

Process mining is a discipline bridging the gap between data science and process science van der Aalst [61]. Such techniques mine event logs from process perspectives, and the most powerful and commonly used type of process mining is process discovery, which is ideal for addressing the research questions in an evidence base. Meanwhile, since companies nowadays are used to collecting data to support decision-making and product development, the proposed method is practical for businesses to apply in their business contexts. A more detailed discussion can be seen in Chapter 3.

1.3.2 Process Discovery Algorithm

The real world is full of noise, which can be seen in Appendix B. According to van der Aalst [60], there are three notable process discovery approaches dealing with noisiness and incompleteness: heuristic mining [67], fuzzy mining [24], and genetic process mining [12]. This project considers using heuristic mining and fuzzy mining of the wide application among previous research papers [73, 10].

1.3.3 Tools

Tools	Usage		
	Data Understanding	Data Preparation	Modelling
Google BigQuery	✓	✓	
Google Data Studio	✓	✓	
Python		✓	
ProM Lite 1.3	✓	✓	✓

Table 1.1: Tools Adopted in This Thesis

Kaizo stores the data on Google BigQuery, so this research mainly applies Google BigQuery to collect, explore, clean, and construct data. In some cases, Python `pandas` library is used to assist the data preparation; `matplotlib` is used to generate graphs. To apply process mining techniques, ProM Lite 1.3 is used. Apart from modeling, some of the filtering and exploration work is done by creating test models, creating the dotted chart, and applying filter plug-ins. More details can be seen in Section 2.4.

1.4 Research Plan

1.4.1 CRISP-DM Framework

This paper follows CRISP-DM (CRoss Industry Standard Process for Data Mining) framework, which define the steps required in data mining projects [69]. It is domain-independent and can apply to multiple industries. According to Schröer et al. [51]'s literature review, CRISP is applied in nine domains from 24 papers, including information technology. A latest study [40] has retrospected the use of CRISP-DM twenty years after the Wirth and Hipp [69]'s proposal. The framework still plays an important role in data science field, but the author also mention the limitation of applying CRISP-DM in the exploratory projects. However, since the goal of this project is pre-identified, CRISP-DM is an ideal framework to apply.

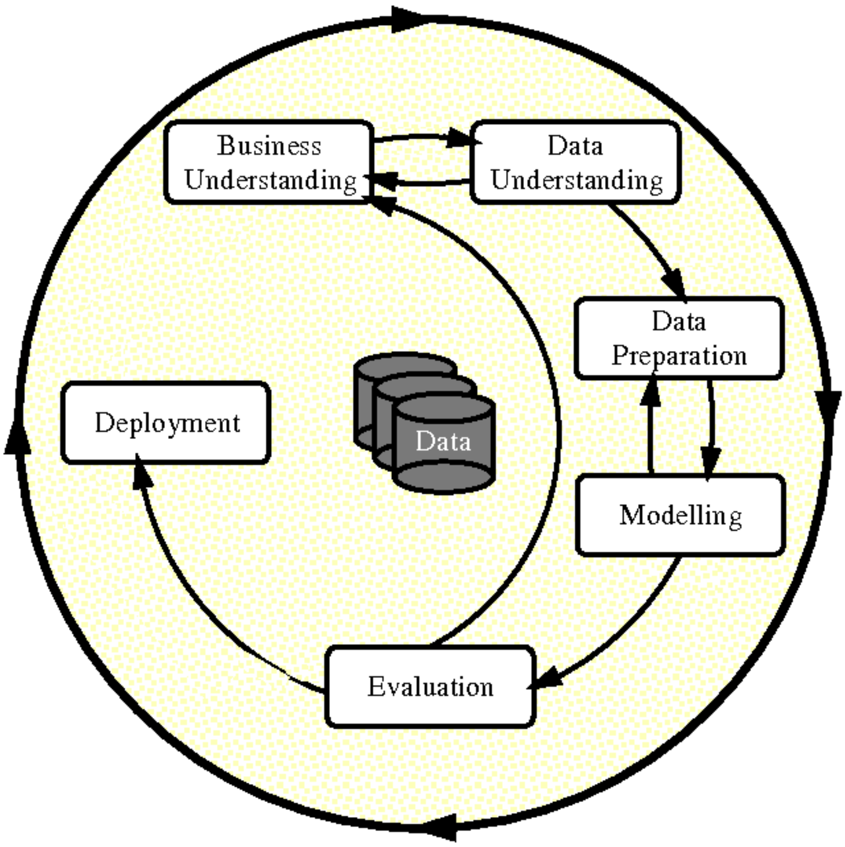


Figure 1.2: CRISP-DM Standard Proposed by Wirth and Hipp [69]

Figure 1.2 shows the steps of the framework: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. The framework starts from understanding the business by organizational structure, business model, business goals, and research goals if cooperating with businesses. In other word, the purpose of the initial phase is to identify research goal and accumulate domain knowledge.

This project takes advantage of involving domain knowledge because the author worked in the company. Researchers can take advantage of using domain knowledge to define event attributes for discovering data patterns, as well as define the project scales [2]. Schuster et al. [52] also point out the importance of utilizing domain knowledge in process discovery in order to improve model quality.

Data understanding phase aims to know how the data looks like. Understanding the schema can further improve the understanding of business, so it is a loop of business understanding and data understanding in Figure 1.2.

Data preparation and modelling are iterations. Data preparation is the phase to collect, clean, and merge data. In the real world, data is not always store in the ideal way, so it is necessary to prepare data needed for modelling. The insight from the result model can further improve the data preparation of the next iteration.

After that, the evaluation is required to see the accuracy of the data models and whether the models achieves the business goals; the final phase is deployment. It could be a report or a application for the business to apply the research framework.

1.4.2 Research Plan

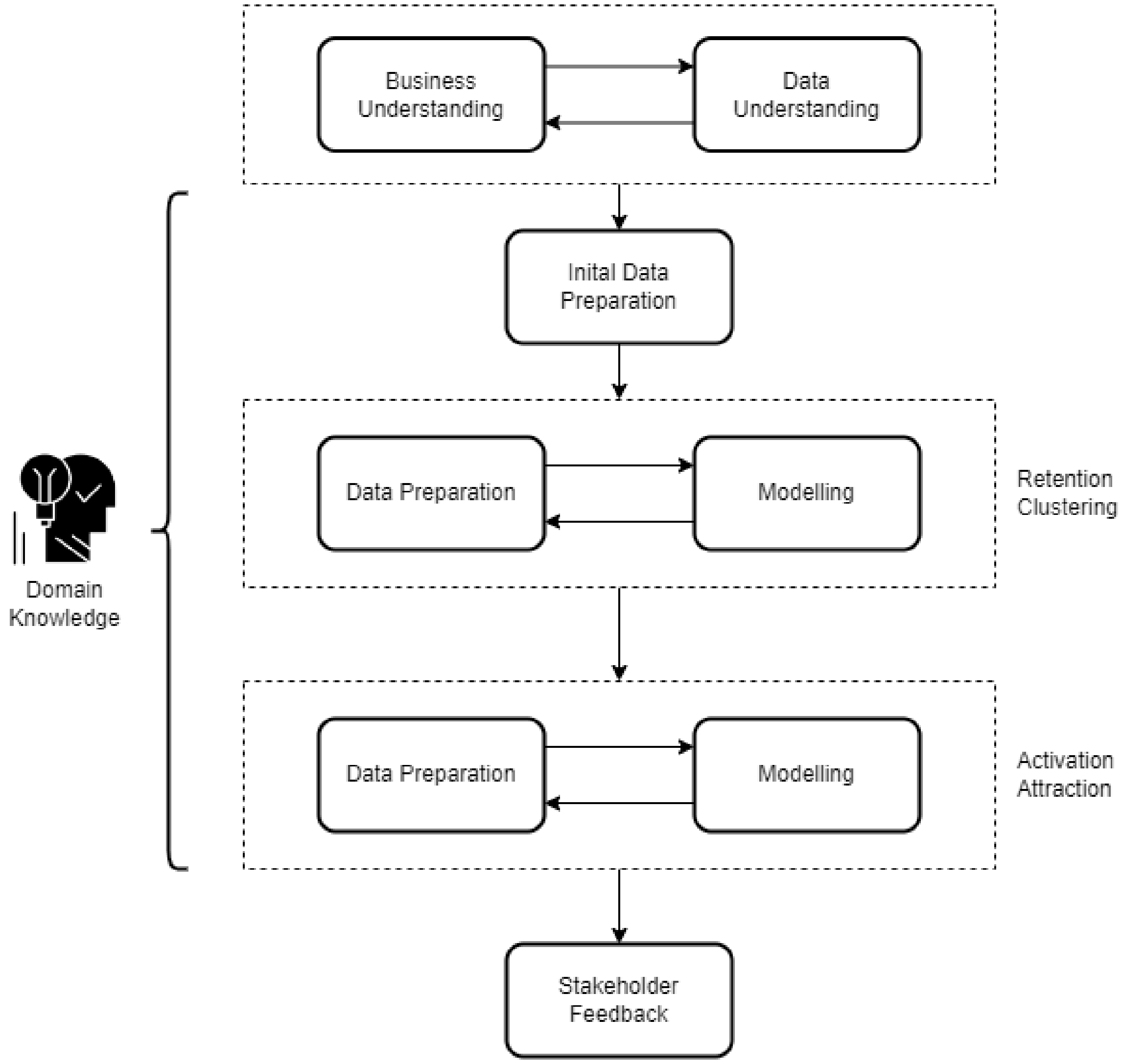


Figure 1.3: Research Plan

Adapted from CRISP-DM, Figure 1.3 shows the research plan of this paper. There are five steps:

Step 1 - Understanding

Step 2 - Initial Data Collection

Step 3 - Retention Clustering

Step 4 - Activation Identification

Step 5 - Feedback

Compared with original CRISP-DM framework, **Step 3**, the retention clustering is added in order to answer **RQ1**. After that, the insight is added to the next phase for activation identification (**Step 4**) in order to answer **RQ2**. Retention clustering and activation identification contains loops between data preparation and modelling. Each step contains more than 10 small and big iterations. The following discussions only write down the big iterations.

1.5 Outline

Table 1.2 shows the tasks that should be down according to the CRISP-DM framework with the cross-reference. Except business and data understanding, data preparation and modelling contain in several iterations, which include in multiple sections.

Phase	Tasks	(Sub)sections
Business Understanding	Identify business model and business goal	4.1
	Identify company's requirement of the project	1.1
	Identify resources of the company	4.2
	Identify challenges and limitation	4.1, 4.6
	Define project goal	1.2
	Produce project plan	1.5
Data Understanding	Collect initial data	5.2.1
	Describe data	4.6
	Explore data	4.6
	Verify data quality	4.6
Data Preparation	Describe dataset	5.1
	Select data	5.2.1, 7.1.2, 6.4, 7.2.1.1
	Clean data	5.2.3, 6.3, 7.1, 6.3, 7.2.1.2
	Construct and merge data	5.1, 6.2, 5.2.4, 6.3, 7.1.1
Modelling	Select Modelling Technique	1.3.2
	Build Model	6.5, 6.6, 6.7 7.2
Evaluation	Stakeholder Feedback	7.3

Table 1.2: Research Tasks

2 PROCESS MINING FOUNDATION RELATED TO THE THESIS

According to van der Aalst [61], process mining is a discipline bridging the gap between data science and process science. Data science and process science are both umbrella terms that contain multiple disciplines. As an interdisciplinary field, data science builds on informatics, data, statistics, computing, communication, data management, sociology, and the environment containing various domains and contexts [6]. The most well-known data science fields include machine learning, data mining, and data visualization. Process science is the study of continuous change. The series of changes can be uncovered in terms of time and other dimensions [64], which includes all process-oriented approaches, including business process management, workflow management, business process reengineering, and operations research [**?**]. However, although process mining is the umbrella term bridging two umbrella terms, the essentials remain the same that process mining leverage event logs to analyze the process.

According to the characteristics of process mining, three components are mandatory: event representing activity, timestamp to determine order and process, and trace. The trace is the process's unit or case as mentioned in **?? ??**. Manipulating and assigning different traces provide different perspectives of the process models.

2.1 Discovery, Conformance, and Enhancement

Process mining can be categorized into three types: discovery, conformance, and enhancement [61]. Process discovery can be viewed as the initial phase of process mining because conformance checking and process enhancement require a given process model, which can be built by process discovery [10]. Therefore, process discovery is the most commonly used type in process mining research [73, 23]. The process model can be built without a priori knowledge, making it a significant advantage of process mining. As mentioned in Section 3.1, using existing analytic tools requires an understanding of the process, making process discovery especially valuable during the exploration phase.

Conformance checking aims to find the deviation between the given model and event logs. In other words, it is used to compare the difference between expected behavior and actual behavior [43]. It is often used in auditing in business operation management [14].

After identifying the gap between expected and real behavior, process mining moves forward to enhancement to improve the workflow. Process enhancement ”extends or improves an existing process model using information about the actual process recorded in some event log.” [61]. Researchers may enhance the process model by repair or extension, for example, by adding a new dimension to the model. Extending process model for enhancement is more commonly used than repairing [70].

This thesis focuses on process discovery instead of conformance checking and process enhancement.

2.2 Perspectives

According to van der Aalst [61], the process model can be separated into four perspectives: control-flow, organizational, case, and time perspectives. Control-flow perspective can be viewed as process perspective as well [73]. It is somehow the mandatory perspective because it considers the ordering of the activities. The organizational perspective considers the resources, such as roles and teams. Since the resources and players are considered, the relations between each other are also part of the organizational perspective. Case perspective considers the case, called trace, in process mining. A case is an unit of analyzing data, for example, a user or session. In this perspective, the attributes of cases are considered. The time perspective focuses on time and frequency, applying the timestamp of the event logs.

2.3 Process Discovery Algorithm

Researchers have proposed multiple algorithms to discover process models. In this research, *fuzzy miner* and *heuristic miner* are used because of their capacity to deal with noisy data and unstructured processes.

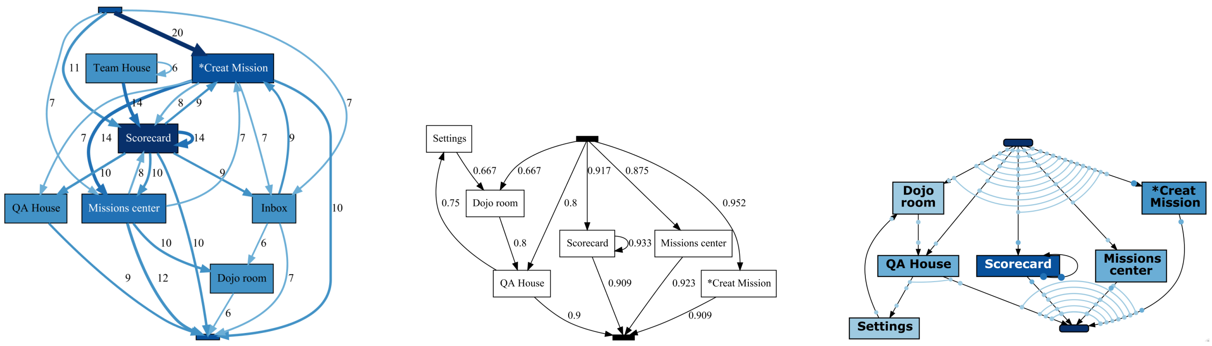
2.3.1 Heuristic Miner

Heuristic miner is another popular process discovery algorithm [10]. This algorithm considers the frequencies of events and sequences to avoid noisiness [61], which is also suitable for applying real-life contexts. On the other hand, the output includes more semantic meanings than *Fuzzy Miner* [38] because the models show dependency and causation by AND- and XOR-semantics [24].

2.3.1.1. Dependency Graph

The expected outcome of *heuristic miner* is casual net (C-Nets), but the direct-follows graph and dependency graph are built before [67]. Figure 2.1 represents the order in how Heuristic miner creates the process graphs. The sample graphs are generated from the *interactive data-aware heuristic miner* [38].

The initial concept of causal dependency is that if another activity always follows an activity, the two activities are dependent. Therefore, the frequency of occurrence of the relations needs to be calculated.



(a) Direct-follows graph (b) Dependency Graph (c) Causal Net

Figure 2.1: Example output of Heuristic Miner from Direct-follows Graph to Causal Net (from Kaizo dataset)

```
Trace 1: <A,B,C,D>
Trace 2: <A,B,C,D>
Trace 3: <A,B,C,D>
Trace 4: <A,C,B,D>
Trace 5: <A,B,D>
Trace 6: <A,C,D>
Trace 7: <A,C,D>
Trace 8: <A,D>
```

Listing 2.1: Event Sequence Example

	$* \to \mathtt{A}$	$* \to \mathtt{B}$	$* \to \mathtt{C}$	$* \to \mathtt{D}$
$\mathtt{A} \to *$		4	3	1
$\mathtt{B} \to *$			3	2
$\mathtt{C} \to *$		1		5
$\mathtt{D} \to *$				

(a) Frequency Matrix

	$* \to \mathtt{A}$	$* \to \mathtt{B}$	$* \to \mathtt{C}$	$* \to \mathtt{D}$
$\mathtt{A} \to *$		$\frac{4-0}{4+0+1}$	$\frac{3-0}{3+0+1}$	$\frac{1-0}{1+0+1}$
$\mathtt{B} \to *$	$\frac{0-4}{0+4+1}$		$\frac{3-1}{3+1+1}$	$\frac{2-0}{2+0+1}$
$\mathtt{C} \to *$	$\frac{0-3}{0+3+1}$	$\frac{1-3}{1+3+1}$		$\frac{5-0}{5+0+1}$
$\mathtt{D} \to *$	$\frac{0-1}{0+1+1}$	$\frac{0-2}{0+2+1}$	$\frac{0-5}{0+5+1}$	

(b) Causal Dependency Calculation

	$* \to \mathtt{A}$	$* \to \mathtt{B}$	$* \to \mathtt{C}$	$* \to \mathtt{D}$
$\mathtt{A} \to *$		0.80	0.75	0.50
$\mathtt{B} \to *$	-0.80		0.40	0.67
$\mathtt{C} \to *$	-0.75	-0.40		0.83
$\mathtt{D} \to *$	-0.50	-0.67	-0.83	

(c) Dependency Measure (Significance)

	$* \to \mathtt{A}$	$* \to \mathtt{B}$	$* \to \mathtt{C}$	$* \to \mathtt{D}$
$\mathtt{A} \to *$		0.80	0.75	
$\mathtt{B} \to *$				0.67
$\mathtt{C} \to *$				0.83
$\mathtt{D} \to *$				

(d) Apply $\textit{Dependency Threshold} = 0.6$

Table 2.1: Example of Calculating Significance in Dependency Graph

The first graph, directly-follows graph (Figure 2.1a), is intuitive and represents the direct-follow relations and frequencies between two events within a trace. Only the event sequences and event orders within the same trace are considered. The frequency can be counted and presented as a frequency matrix (See Table 2.1a as an example.)

The significance of the dependency, take $\mathtt{A} \to \mathtt{B}$ as an example, can be calculated by

$$\text{Significance of } \mathtt{A} \to \mathtt{B} = \frac{|\mathtt{A} \to \mathtt{B}| - |\mathtt{B} \to \mathtt{A}|}{|\mathtt{A} \to \mathtt{B}| + |\mathtt{B} \to \mathtt{A}| + 1}$$

where the calculation case can be seen in Table 2.1b.

The higher the number, the stronger relations between the two activities. That is, relations between A and B has relatively high dependency than B and C as the former's significance is 0.8 and the later is 0.4.

After applying the pre-defined thresholds, the dependency graph can be created (See Figure 2.1b). According to [67], there are three thresholds:

Dependency	The minimum significance.
Positive Observation	The minimum occurrence of the events. In ProM [37], the threshold is named as **Frequency**, with the use of relative value rather than absolute value.
Relative to Best	The difference of the significance and the "best" significance is lower than the threshold, which is used to capture infrequent behaviors.

Involving thresholds in the modeling allows heuristic mining to deal with noisy data. The dependency and positive observation thresholds help filter out the noisy data, while relative to the best threshold, add back infrequent patterns back to the model.

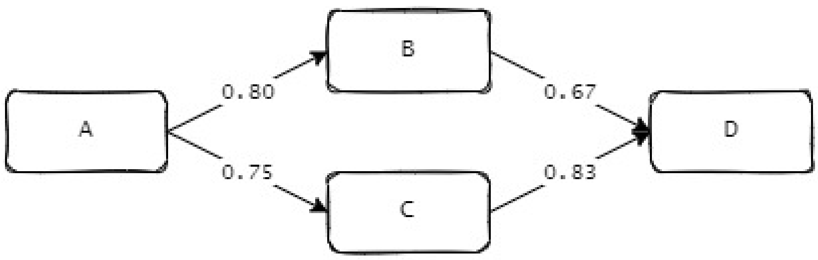


Figure 2.2: Dependency Graph Example

Table 2.1d shows the example of applying $\textit{Dependency Threshold} = 0.6$, $\textit{Positive Observation} = 1$, and $\textit{Relative to Best} = 0$. As as result, four relations are remained on the graph: $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow D$, and $C \rightarrow D$ (See Figure 2.2).

2.3.1.2. *Causal Nets (C-Nets)*

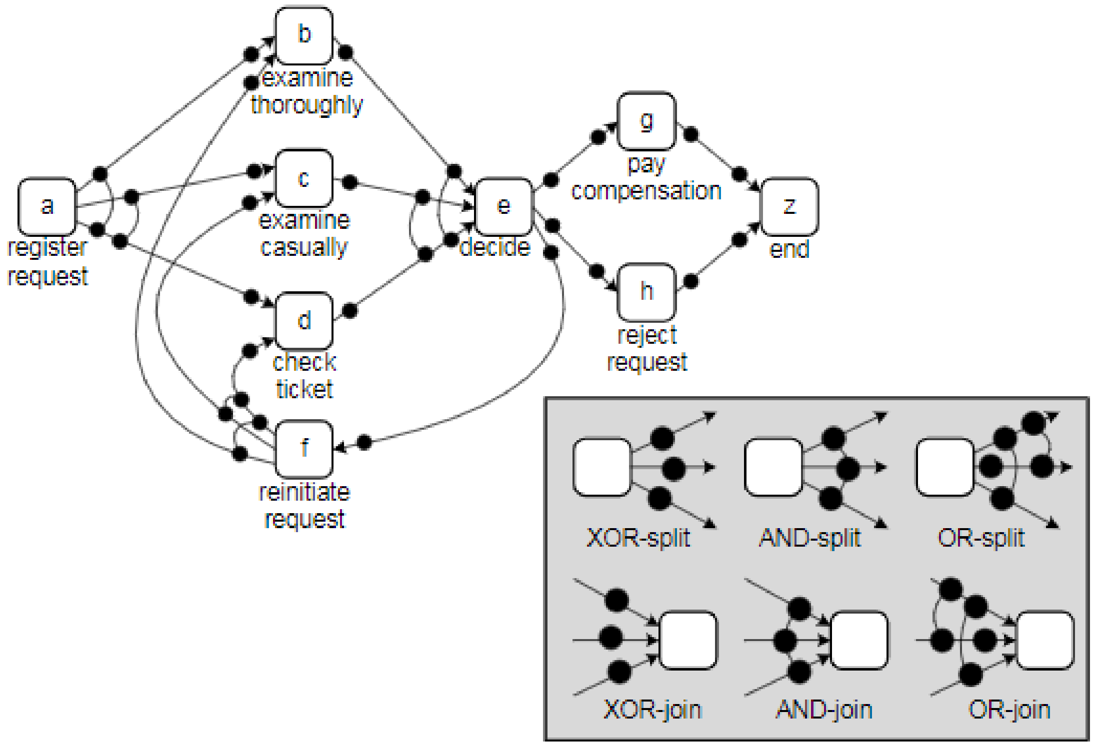


Figure 2.3: C-Net Example from Aalst et al. [1]

Causal nets (C-Nets) are graphs where nodes represent the events, and the edge represents the causal dependencies [1]. Apart from the starting and ending nodes, each node has inbound and outbound edges. Each black dot shows a possible join or split. The link between the dots is the binding that the preceding events must contain the corresponding events in AND-join, or the following events must contain the corresponding events in AND-split (See Figure 2.3). For example, node a, with OR-split, the possible event sequences can be $\{a, b, d\}$, $\{a, d, b\}$, $\{a, c, d\}$, or $\{a, d, c\}$. Meanwhile, the frequency of directly-follow relations is also considered. The low-frequency relations will be filtered out from the process model.

2.3.1.3. *Model Configuration*

The modelling in this research has adjusted the *length-one loops* and *length-two loops* thresholds. *Length-one loops* can be considered as a self-loop. By choosing the *length-one loops* to 1, the models are built without *length-one loops* [65]. *Length-two loops* is a longer loop such as the event sequence $\{a, b, a\}$ and $a \neq b$. $\{a, b\}$ is detected as *length-one loops*. Removing the loops masks the model clean and release more binding information.

All-tasks-connected heuristic is also applied in several models. *all-tasks-connected* heuristic sometimes makes the model more explainable since many dependency relations are hidden without records [65]. However, *all-tasks-connected* makes the model more complicated as it includes all observed activities without the consideration of observation frequency. The infrequent behavior can be a noise in the model.

2.3.2 Fuzzy Miner

The *fuzzy miner* algorithm is designed to cope with the real-life environment with less structured and spaghetti-like models [24]. The output simplified graph can be viewed as user flow, which is intuitive for stakeholders to involve. The events are translated into nodes, while the edge directly shows the preceding and following relations between two events, making *fuzzy miner* the most commonly used algorithm in business process mining [73].

2.4 Process Mining in Action

This section introduces all the practical steps in this process mining project. The process mining analysis is conducted on ProM, an open-source software and framework with multiple plug-ins integrating process mining techniques [1]. This research uses ProM Lite 1.3.

The required data format of ProM is XES (eXtensible Event Stream), a standard especially suitable for storing process data [36] because of the advantage of representing traces. As this paper exports event logs from Google BigQuery in CSV format, the pre-installed plug-in [39] can easily convert the CSV file into XES format after assigning the trace, event, and timestamp. Researchers are allowed to add more event attributes. Take the code snippet as an example, `user_id` is predefined as trace, and `role_type` is the event attribute.

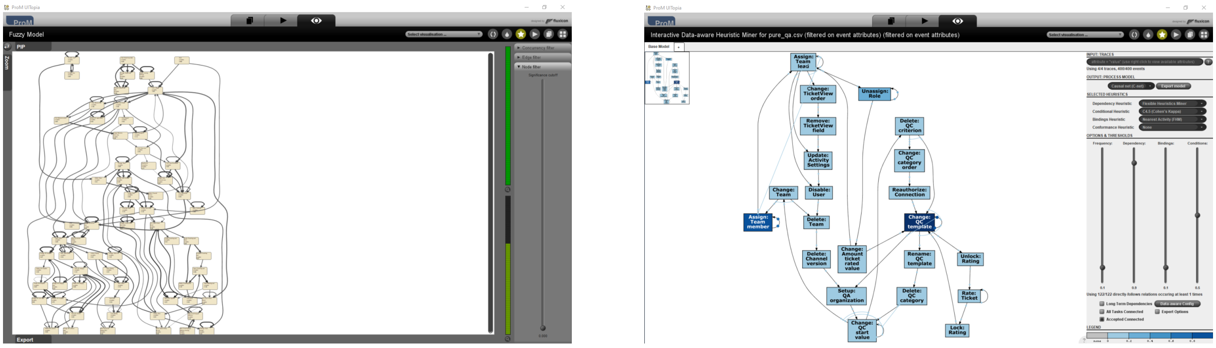
```xml
<?xml version="1.0" encoding="UTF-8" ?>
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">
  <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext"/>
  <extension name="Lifecycle" prefix="lifecycle" uri="http://www.xes-standard.org/lifecycle.xesext"/>
  <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.xesext"/>
  <classifier name="Event Name" keys="concept:name"/>
  <classifier name="(Event Name AND Lifecycle transition)" keys="concept:name lifecycle:transition"/>
  <string key="concept:name" value="coaching_before_retention.csv (filtered on event attributes) (filtered on event attributes)"/>
  <trace>
    <string key="concept:name" value="5bc8a6abc31c8b8b333820e40856b077"/>
    <event>
      <string key="roleType" value="manager"/>
      <string key="event" value="Visit Village"/>
      <date key="time:timestamp" value="2022-05-20T04:22:36.000+02:00"/>
      <string key="lifecycle:transition" value="complete"/>
    </event>
```

[1]https://www.promtools.org/doku.php

```
    <event>
      <string key="roleType" value="manager"/>
      <string key="event" value="Click Button"/>
      <date key="time:timestamp" value="2022-05-20T04:24:20.000+02:00"/>
      <string key="lifecycle:transition" value="complete"/>
    </event>
  </trace>
</log>
```

Listing 2.2: XES Example

The converted XES is the required format to generate process models. In this research, Plug-in "*Mine for a Fuzzy Model* [62]" and "*Interactive Data-aware Heuristic Miner (iDHM)* [37]" are used. The former plugin implements *fuzzy miner*, and the latter implements *heuristic miner*. The detailed implementation can be found in Günther and van der Aalst [24] and Mannhardt et al. [38].



(a) Fuzzy Miner

(b) Interactive Data-aware Heuristic Miner

Figure 2.4: ProM Screenshots

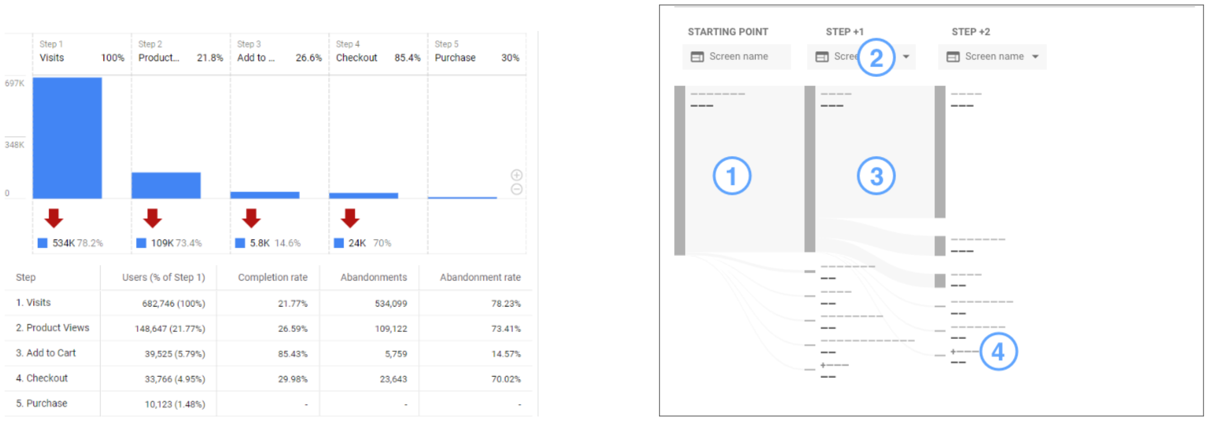
3 BACKGROUND AND RELATED WORKS

3.1 Commercial Analytics Tools

Thanks to the widely spread data-driven concept, most companies are used to storing and leveraging data in order to support business decision-making. Meanwhile, many analytic tools, such as Google Analytics, Mixpanel, Hubspot, and Tableau, are adopted for commercial use. However, many of the researchers mentioned the limitation of such tools.

Vinod et al. [63] compare the disciplines for e-commerce analytics. Although web analytics tools can improve the site navigation and conversion goals, the improvements can only be made by a clear understanding of site structure and user needs.

Terragni and Hassani [58] argued that these tools are too simplified by understanding website user behavior. The process model is able to capture the causality behavior pattern that leads to the conversion and is analyzed by different kinds of instances and perspectives.e conversion, and analyzed by different kind of instance and perspectives.



(a) Funnel Exploration Report [47]

(b) Path Exploration Report [22]

Figure 3.1: The User Flow Reporting in GA4

Take the latest version of Google Analytics, GA4, as an example, the analysis regarding flow or process can be seen in two reports: Funnel Exploration and Path Exploration (See Figure 3.1). According to Google's official instructions [22, 21], the steps need to be predefined in order to show user behavior. That is, if the site owner does not have a clear understanding of the primary process, some of the insights will be missed.

3.2 Web Mining

Web mining is a branch of data mining using weblogs with the adaption of web scenarios. As one of the SaaS characteristics is that the product or service is delivered by web browser [44], this research has a clear overlap between process mining and web mining.

Sharma et al. [53] categorizes web mining into web content mining, structure mining, and usage mining. Web usage mining aims to discover usage patterns from web-based applications using data mining techniques. Srivastava et al. [55], Sharma et al. [53] divides the process into three steps: preprocessing, pattern discovery, and pattern analysis. Pattern discovery contains some topics, such as association rules or sequential rules. The association rules aim to identify the hidden relationship between each instance or activity. The sequential pattern focuses on analyzing time-ordered set [55]. Apparently, such topics overlap with process mining. In fact, some of the research combines web usage mining and process mining while the data source originated from web applications [25]. However, such researches apply web usage mining as the primary methodology for preprocessing and process mining techniques. After all, process mining treats the process as the research subject, while web mining is a more broaden term.

3.3 Related Work of Process Mining

3.3.1 Process Mining in Navigation Behavior

Due to the noisy nature of web log data and the flexibility of user path, some of the researchers will add dummy instances during the preprocessing phase as a helper for the mining algorithm. Vinod et al. [63] add Start, End, and BuyerEnd to each web session. BuyerEnd are specified in order to analyze purchased sessions.

Vinod et al. [63] apply process mining to analyze an online travel agency's user navigation and purchase behavior. The research saturates the dataset to address purchased customers' low-frequency traces. The saturated dataset is chosen from the web sessions that are longer than average, which increases the percentage of purchased visits. The subsets are then mined by a knowledge-based miner and a heuristic miner.

Zaim et al. [71] aims to determine the information needed for the online store evaluation by analyzing navigation pathways. The research was divided into three aspects by analyzing usability, content adequacy, and reliability. The data apply fuzzy mining to determine the extent of each aspect. The result can be used in making strategies.

Husin and Ismail [25] analyzes the navigation behavior of a news website by the Alpha algorithm. Meanwhile, web usage mining is utilized for data cleaning and preprocessing. The research analyzes the navigation between pages and within sections, showing the users' interest in different types of information. The result shows the content interest of the users and the trends, which provide advice for website owners to attract more traffic.

3.3.2 Process Mining in Learning Behavior

The previous researchers also show interest in analyzing online learning behavior, as LMS is a trendy topic these days. It is called Educational Process Mining (EPM) while applying process mining in the educational domain [5]. According to the extended definition of customer that customers are users who experience the service [68, 4], considering students as customers in the educational process is acceptable and is seen in previous researches [13].

Taub et al. [57] tracks the changes in students' self-regulated learning behaviors. The research creates causal nets in terms of different conditions, for example, the number of attempts and whether the student passes the exam or not. The insight of extracting valuable factors from the process will be further applied in this project.

3.3.3 Process Mining v.s. Customer Journey

Although the research aims to support product development from a business perspective, the research subject is customers, which are the users who experience the service [68, 4]. Previous researches show some integration between process mining and customer journey mapping.

Customer Journeys Analysis is a method of understanding customer experience, the customers' interpretations of the service process, and their emotions [26]. It focuses on how customers interact with touch points, trying to understand the various paths for customers to complete the goal [33]. Meanwhile, customer journey mapping is the analysis process to identify customer journey [19]. Such a technique originated from marketing and service management. In the previous literature, some of the researchers tried to fill the gap between data science and customer journey maps. Process Mining is intuitively the closest discipline because they both address sequential user behaviors.

Bernard and Andritsos [4] push the customer journey maps (CJMs) further to assist decision-making by showing the potential of integrating process mining and CJMs. The paper proposes a CJM model mapping the CJM components to process mining analyzable XES format. The concept of divided journeys into expected and actual journeys meets the conventional process mining approaches [58].

Customer journey modeling language is an attractive idea for analyzing customer behavior and seems suitable for the research questions. However, CJML is a comparative tool requiring an expected journey in advance for comparison. Since Kaizo has no expectation of user flow, this method can keep for future research.

3.3.4 Process Mining v.s. Human-Computer Interaction

Human-Computer Interaction (HCI) is a popular field of understanding how users interact with the system. Most of the research is design-oriented and extends to other senses except senses of sight [31]. However, software process mining is a overlap between Human-computer Interaction aiming to analyze software execution data from a process perspective [50, 34, 59]. The result of mining users' behavior from software event logs can be used to improve the UI design and usability.

Theis and Darabi [59] apply the HCI concepts while collecting data. Apart from the general UI and frontend events, such as click or visit data, the authors tracked mouse movements and keystrokes, which are not commonly tracked in process mining and business contexts. The proposed approaches are used to detect whether the users interact with the system optimally.

Else et al. [15] apply process mining techniques to usability testing during product development. The process mining models extracted from logs can be used to generate evidence-based usability test scenarios, while traditionally, the scenarios are created based on user researchers' choices and experiences.

Liu et al. [35] discover behavior model based on pre-identified interface. The research answers the question of which the given interface provides function.

Cerone [8] discuss the vast field of which process mining can apply except business process management. The research shed light on the possibility of applying process mining techniques to learning, HCI, cognitive modeling, traffic, and emergency management.

Human-computer Interaction is a discipline in which human factors and cognitive science play essential roles. As the researchers stand on understanding human behavior, process mining can be a handy tool for HCI scientists to enrich the research.

3.4 Summary

This chapter introduces the possible disciplines to solve the research questions and why process mining is more ideal than others. The process mining techniques used in this research are also introduced. Furthermore, the previous process mining papers are discussed.

During the literature review, although many researchers target customers, the users' businesses serve as their research interests, seldom do they apply process mining approaches. Previously, business internal process management gained more interest for process mining researchers. The reason might be that the internal business process is more invisible, and the result, especially bottleneck research, can significantly reduce the operational cost. On the other hand, customer analysis has more existing methods, such as marketing research and interviews. However, the existing techniques have the disadvantage of expensive or unsuitable for user flow exploration. Meanwhile, no previous research was found in analyzing customer activation in the SaaS industry.

4 BUSINESS AND DATA UNDERSTANDING

4.1 Business Overview

4.1.1 Software-as-a-Service

Software-as-a-Service (SaaS) is a software deployment model where software is provisioned over the internet as a service [44]. It is an increasing trend and popular service provided by software vendors nowadays. Popular SaaS services include Dropbox, Zoom, Microsoft, Adobe Creative Cloud, and Google Cloud. A typical SaaS product often starts with free usage and charges if the account meets specific criteria. For instance, under the basic free plan, Zoom users can only host meetings for less than 40 minutes.

Mäkilä et al. [44] collected multiple SaaS definitions and extracted the main characteristics that:

- the product is used through a web browser;
- the product is not tailor-made for each customer;
- the product does not include software that needs to be installed at the customer's location;
- the product does not require special integration and installation work;
- the pricing of the product is based on actual usage of the software.

According to the definitions, Zendesk and Kaizo apply the SaaS model.

4.1.2 Zendesk

Zendesk is a B2B platform to support customer relationship management (CRM) by connecting customer service agents and customers. The platform integrates all the messages from multiple channels, such as calls, website chats, and social media, into one place, allowing the customer service team to manage all customer requests easily.

Zendesk is a well-known CRM platform service for CRM and has become an ecosystem with other extensions [1]. As SaaS products are not customized for each customer [44], the extensions act as a supplement to the service, allowing clients to customize the platform in order to fulfill their business needs.

[1] https://www.zendesk.com/marketplace/

4.1.3 Kaizo App as a Zendesk Extension

The Kaizo app is an extension that can be found in the Zendesk marketplace. The product supports CRM teams in tracking performance and optimizing the operational workflow. While Kaizo is an independent service, it shares a different pricing policy than Zendesk.

The company starts to charge when the users hit a specific headcount. Unfortunately, as a startup, the pricing plan changes dynamically without records. Since the pricing policies can potentially influence user behavior, the data will be extracted by new installs as the policies are more stable.

4.2 Stakeholder Identification

Stakeholders indicate the various groups of actors in and around the firm who can affect or are affected by the firm. [16, 18]. Due to the comprehensive definition, the stakeholders can be far away from governments and environmentalists, which is not the focus of this research. Generally, this section only discusses and identifies the relevant stakeholders that can influence the research.

4.2.1 External Stakeholders

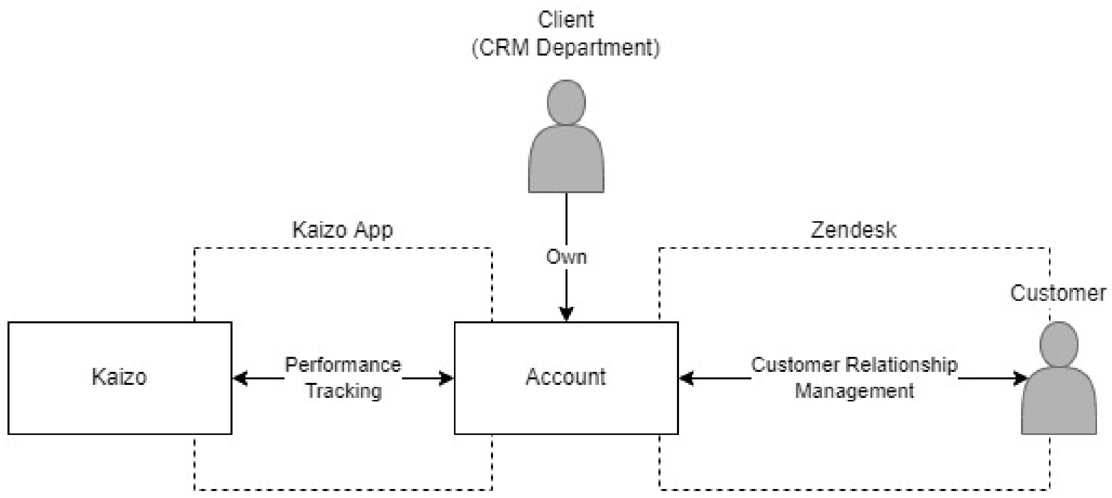


Figure 4.1: Interaction between each instance

The external stakeholders indicate the parties outside Kaizo company. Figure 4.1 shows the relationship between external stakeholders and Kaizo company. The CRM teams in other businesses provide customer service to their customers on the Zendesk platform. As Zendesk accounts, clients can install Kaizo's product on the Zendesk marketplace. Under the permission to use the logs generated from Zendesk, the Kaizo extension presents metrics for clients to track performance and optimize workflow.

Users on the Kaizo platform can have multiple roles. Each role corresponds to different permissions. The default set contains five different roles. A user can have multiple roles at the same time:

$$R_i \subseteq \{r_1, r_2, r_3, r_4, r_5\},\ R_i \neq \emptyset$$

where R_i denotes the single user's role set and r_j denotes the roles (See Figure 4.2).

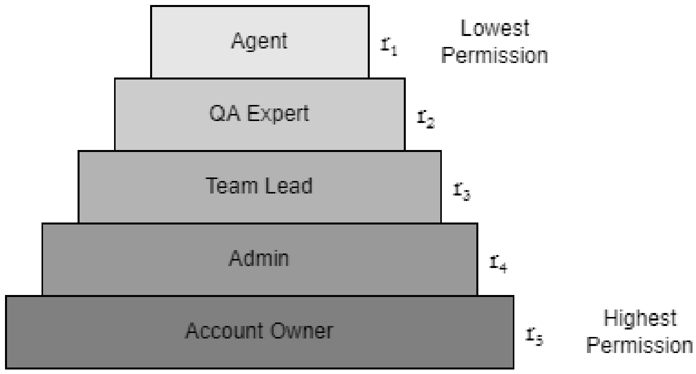


Figure 4.2: Default Roles

As the terms could be complicated because the business model contains multiple parties, Table 4.1 clarify the terms used in this paper. This research analyzes the behavior of clients and users using the Kaizo app.

Terms	Description
Kaizo	The Kaizo company
Kaizo App	The software product (Zendesk extension) developed and operated by *Kaizo*.
product	Indicate *Kaizo App*.
company	Indicate *Kaizo* company.
account	The account installed *Kaizo App*. It refers to a CRM department in a specific business.
domain	The same as *account*
client	The CRM department in a specific business that Kaizo serves.
user	The user using the *product*.
customer	The users which the *client* serves

Table 4.1: Terms Used in This Paper

4.2.2 Internal Stakeholder

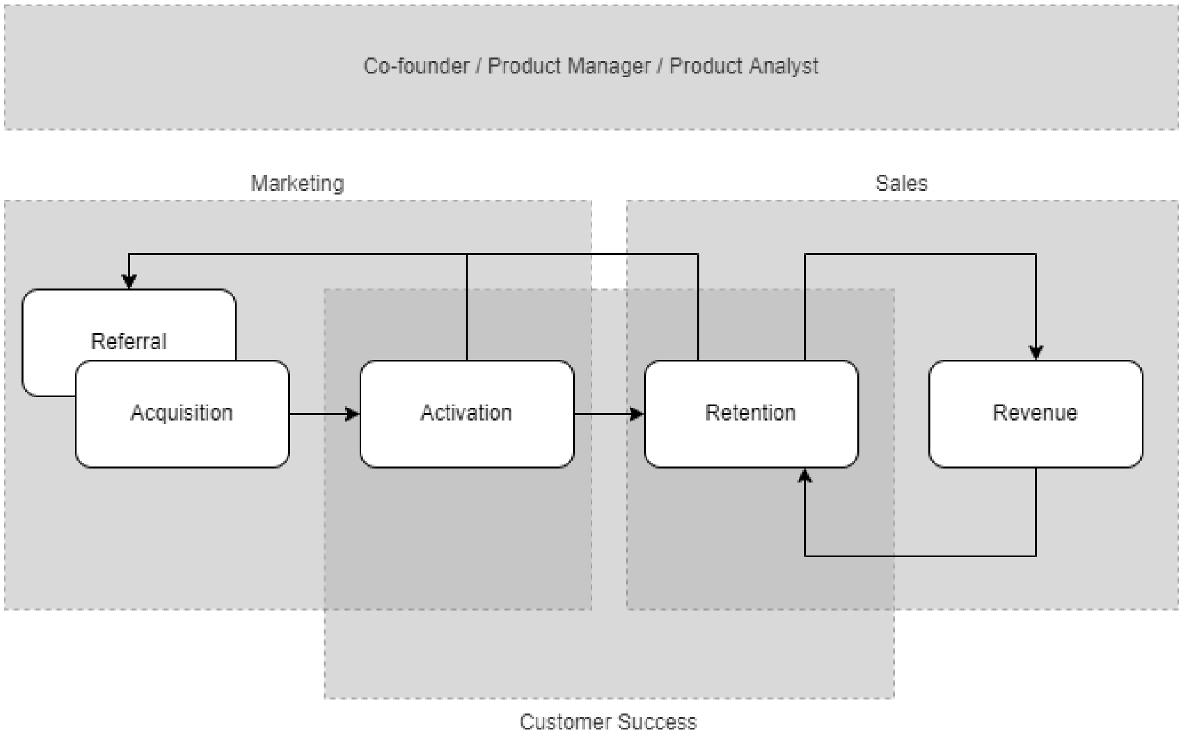


Figure 4.3: Stakeholders and Client Lifecycle

The research involves domain knowledge and stakeholders' opinions to utilize the process model from data preparation to evaluation. Figure 4.3 shows the stakeholders relevant to the research. Adapted from AARRR model [42], stakeholders focus on different phases of the client lifecycle.

The marketing team aims to acquire newly installed accounts and qualified clients, so marketers are interested in acquisition and activation. The sales team cares about converting clients to subscribe and pay for the service while renewing the contract is also considered. The paid clients with low retention rates are more likely not to renew the contracts, which decreases the expected revenue. Therefore, the salesperson considers not only revenue but also retention. The customer success team helps the clients to adopt the Kaizo app. The interest of the team is relatively broad because this role covers the most prolonged phase in clients' life cycle. Other than the teams with a specific interest, co-founders, product managers, and product analysts cover the whole life cycle.

Identifying the internal stakeholders and their interests provides guidance on whom the researchers should consult with. As the research mainly focuses on activation and retention, the research activity involves the customer success team's participation.

4.3 Product Solutions and Research Pillars

4.3.1 Transform Solutions to Pillars

Figure 4.4: Product Solutions and the Identified Pillars

From Kaizo's official website [28], the product contains five solutions, which can be considered as the features in the Kaizo app. As a Zendesk extension, the Kaizo app reorganizes the logs generated from the operational activities of the clients in Zendesk to support CRM operations. The organized metrics are the foundation of the product. Other solutions can be viewed as derived features of **Scorecards**. As **Reports** is an attached function of **Quality Assurance** and **Scorecards**, three pillars are identified: **Quality Assurance**, **Missions**, and **Performance Coaching**, where $\mathbf{P} = \{P_Q, P_M, P_C\}$, P_Q denotes **Quality Assurance**, P_M denotes **Missions**, and P_C denotes **Performance Coaching** (See Figure 4.4).

Quality Assurance focuses on service quality. Users rate the conversions between agents and customers as a quality check to supplement the quantitative metrics. **Missions** leverages the performance metrics to set up weekly goals for agents. Agents with assigned missions are required to commit and put effort into achieving the goals. For instance, the mission could be to increase the speed of response to the message from customers for a particular value. **Performance Coaching** turns the metrics into actionable advice according to the performance. The advice, transformed online, contains multiple types, such as meetings and messages.

4.3.2 Domain Object and Milestone Events

The introduction to the pillars in the previous subsection shows that each pillar has its domain object (See Table 4.2). The domain object is also called the object of interest, originating from the human-computer interaction discipline. According to Beaudouin-Lafon [3], domain Objects from the set of potential objects of interest for the user of a given application, which form the basis of the interaction and its purpose. Each domain object can be viewed as a virtual object

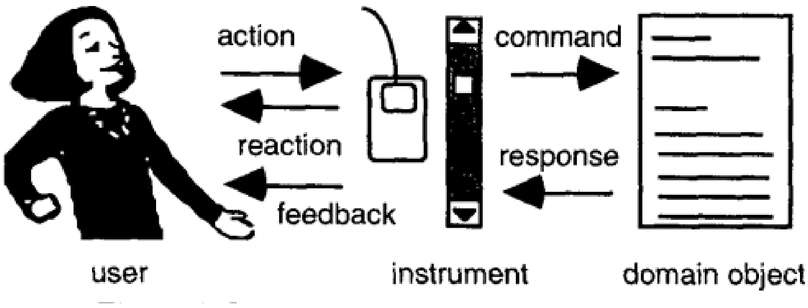


Figure 4.5: Interaction Instrument [3]

on the Kaizo app for users to create, edit, and change attributes. The domain object is denoted as $\mathbf{O}$, where $\mathbf{O} = \{O_Q, O_M, O_C\}$.

Pillar (P)	Domain Object (O)	Description
Quality Assurance (P_Q)	Ticket (O_Q)	The means through the end users (customers) communicate with agents in Zendesk Support [72].
Missions (P_M)	Mission (O_M)	The tasks with goals for agents to improve the performance.
Performance Coaching (P_C)	Card (O_C)	The advice for improving performance outcome.

Table 4.2: Domain Objects

Inspired by Peng and Cheng [49] of identifying transactions for session clustering, each pillar contains milestone activities associated with the domain objects. Fortunately, some milestone activities are pre-identified and tracked, see Table 4.3.

Pillar	Domain Object	Expected Action (Milestone)	
		Manager	Agent
Quality Assurance	Ticket	Rate	?
Missions	Mission	Create	Activate (Commit)
Performance Coaching	Card	Create	?

Table 4.3: Milestone Action on Domain Object

Specifically, the company considers performance tracking and management a two-party game in which users are expected to behave differently by roles. Although there are multiple sets of roles, they could be roughly categorized into two types: manager and agent. R_M denotes the manger's set of roles contains user roles more than r_1 (agent role); R_A denotes the agents' role which contains only r_1. The agents have the lowest permission, which managers supervise.

$$\begin{aligned} R_A &= \{r_1\} &&, |R_A| = 1 \\ R_M &\subseteq \{r_1, r_2, r_3, r_4, r_5\} \neq R_A &&, |R_M| \geq 1 \end{aligned}$$

The expected activities, denoted as E' for event logs, require specific permissions. The milestone events of managers in **Quality Assurance** ($E_Q^{'}$) is restricted to $r_2 \in R_i$; in **Missions** it is restrict to $r_3 \in R_i$; in **Performance Coaching** it is not restricted but users with R_M are able to undertake the expected actions $E_C^{'}$ for other agents.

Considering the users with R_A with the lowest permission, only **Missions** has milestone events tracked, enabling researchers to identify the retention. The other two pillars do not have specific event logs representing the agents actively adopting such feature.

4.3.3 Challenge Identification

The functions of each pillar released in different time and improved dynamically. **Quality Assurance** is the earliest pillar, which is the initial and primary solution of the product. The timestamp of earliest records of E_{P_M} is `2021-11-08 10:17:02.386000 UTC`, while the earliest timestamp of **Performance Coaching** is `2021-12-22 11:10:50.028000 UTC`.

Furthermore, **Missions** and **Performance Coaching** had changed the solution name once, affecting the event naming. Furthermore, some of the old clients are using a specific old version of the product, which contains old functions or previous versions of the pillars. The above cases bring challenges to data preparation.

4.4 Weekly Based Activities

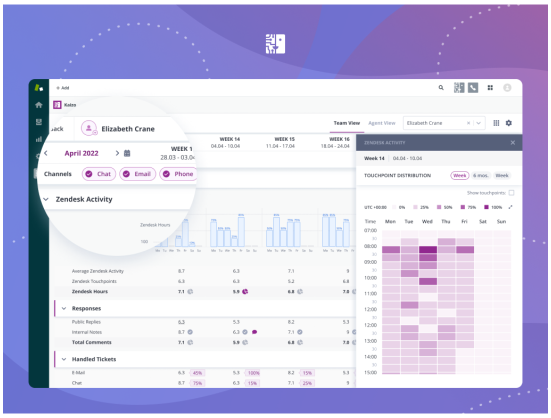
The business activities of clients are weekly based. From stakeholders' observations, the clients consider ISO week [2] as a business cycle. The milestone events of managers in **Missions** is also designed on a weekly base. This understanding will be applied to further research.

[2] https://www.iso.org/iso-8601-date-and-time-format.html

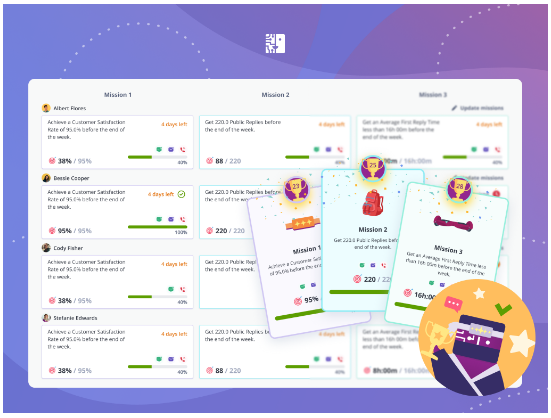
4.5 Interface Identification

Figure 4.6: Kaizo App's Landing Page

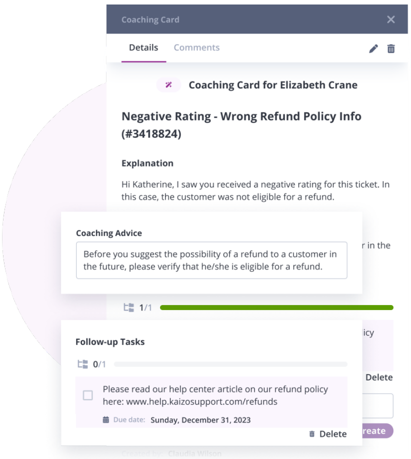
The event naming contains a UI component, so identifying the interface can help understand the data. Understanding how the company designs the product user flow can also help to identify the default landing page to benefit noisy data identification. Meanwhile, inspired by Liu et al. [35]'s research of component interface identification by process mining, the behavior model can be further segmented by the UI module, providing a new aspect of analyzing data. Such identification provides an initial concept of user journeys, but the exact process could not be clearly recognized at this stage due to the fact that UI designers follow the principle of providing flexibility and efficiency of use to users [45, 46], making the process difficult to predict.



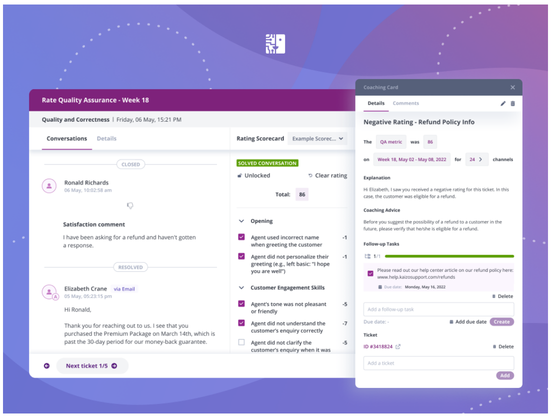
(a) **Scorecards** [29]



(b) **Missions** [29]



(c) **Performance Coaching** [30]



(d) **Quality Assurance** [29]

Figure 4.7: The Screenshot of Each Pillar

Kaizo app, embedded in the Zendesk platform, starts from the landing page and contains multiple UI modules (Figure 4.6). Each module has different functionality but is not limited to one pillar (Table 4.4). For instance, following the model of product solutions and pillars (Figure 4.4), **Scorecards** as UI components can appear in multiple modules such as **Scorecard** and **QA House**. **Performance Coaching**, as a coaching function based on performance and metrics, can appear in many modules as well.

Type	Module	Related Pillars		
		Quality Assurance	**Missions**	**Performance Coaching**
Static	**Dojo Room**		✓	
	Scorecard	✓	✓	✓
	QA House	✓		✓
	Team House			✓
	Missions Center		✓	
Generic	Settings	✓		
	Inbox		✓	✓
	Navigation			
Notification	**Missions Modal**		✓	
	Onboarding	✓	✓	✓
	Notification	?	?	?

Table 4.4: UI Modules

Meanwhile, the UI modules can be categorized into static, generic, and modal. Static module stays on the interface all the time in terms of permissions; generic modules are also static but are hidden and not involved in daily operation and performance tracking. Notification-type modules, in contrast to static ones, are UI components triggered by specific conditions and often overlap the main windows. The events associated with notification type of modules, however, can be viewed as noisy events because the display of notification breaks the event sequences of the main workflows. On the other hand, notifications can be shortcuts to users' desired destinations or notify important information to users, which drives actions.

Each module, especially the static modules, is expected to carry out one of the solutions in the product. Module **QA House** is designed for **Quality Assurance**; **Team House** is designed for **Performance Coaching**. Pillar **Missions** has a more specific module that **Missions Center** provides the function to create missions, and agents can view their mission progress from **Dojo Room** and **Missions Modal**. Again, the trigger of the events associated with each pillar does not limit to the designed modules.

Onboarding modules are instructions helping users to understand how to adopt the Kaizo app. As the company puts effort into improving activation, which is mentioned in Section 1.1, the onboarding flow changes dynamically in order to improve the activation and decrease the drop rate from activation to retention. In other words, onboarding can be a factor in influencing the research questions.

Some clients have a different interface than others due to the specific product version mentioned in Section 4.3.

4.6 Data Description

The data from Kaizo is stored in and extracted from Google BigQuery. Most software contains two types of product data: front-end and back-end, same as the Kaizo app. The front-end events are triggered by users' actions, such as clicks and page loads; the back-end data are generated by product API, which contains the user-generated contents in the pillars and the metrics data. To analyze process, this research mainly uses front-end data that records all users' actions.

However, as the service is delivered by the web browser and the event logs are triggered from the client side, the data is noisier than other types of data. Inspired by Vinod et al. [63]'s research, some of the issues are identified:

Challenge 1 - The front-end data and back-end data are not aligned. This might occur due to the fact that users are using ad blockers on their browsers or loss of connection.

Challenge 2 - The data are unbalanced. For instance, within the same period, the number of activated accounts is much smaller than inactivated accounts.

Challenge 3 - The presence of loop, duplicate activities, and parallel tasks.

Challenge 4 - Many event names have similar names and should be clustered or extracted more attributes.

Challenge 5 - The interface and the function show differently for some of the specific old clients

Challenge 6 - The change of pricing policy does not record in the data warehouse. The pricing plan might influence the activation because it is charged by headcount.

Challenge 7 - The dynamically changed onboarding flow might influence the activation.

Challenge 4 discusses how events are named. The event names are mixed with interface components, modules, pillars, and user activities, which are not aligned. Therefore, it is necessary to clean and extract information to create new attributes for analysis.

4.7 Summary

In this chapter, the dimensions which benefit the following research are identified. First, three pillars, **Quality Assurance**, **Missions**, and **Performance Coaching**, are defined, allowing the analysis to discuss under these contexts. Meanwhile, the associated milestone events are identified by stakeholders. The milestone events will be further used as "transaction" events to extract valuable sessions. Second, the users can be categorized based on permission scheme, R_A, and R_M. Managers and agents act as the two players in the game. Third, most of the clients do business activities weekly. The research will further discuss the interaction between the two parties.

On the contrary, many issues are held which threaten the data quality and data collection. First, the events naming and user flow is dynamically changing. Second, the company has no clear understanding of the milestone events. Third, the onboarding flow might affect the result. These issues should be further addressed in the next phase.

5 DATA PREPARATION

This chapter discusses the initial data preparation (**Step 2**). The collected dataset served as the basis of the following research steps. Meanwhile, test modeling is involved in exploring data and consulting stakeholders. At this phase, fuzzy miner [24] is selected due to the intuitive process map and the capacity to address unstructured processes.

To improve the readability, the data preparation works undertaken after **Step 2** are also recorded.

5.1 Data Tables

This section introduces the Google BigQuery tables used for data preparation and how the tables combined.

5.1.1 Events

Field	Level	Description
`id`	Event Level	The primary key of the records.
`timestamp`	Event Level	The timestamp event triggered.
`week_to_install`	Event Level	The ISO week difference from `timestamp` to `install_time`. The value starts from `0`, indicating the activities occur within the same week of installation.
`event`	Event Level	The event includes module, interface components, pillar, and user actions such as click and visit.
`event_category`	Event Level	The category extracted from event.
`session_id`	Session Level	A set of event sequence less than 30 minutes break.
`user_id`	User Level	The users' anonymous identification.
`domain_id`	Account Level	The clients' anonymous identification.
`install_time`	Account Level	~~The timestamp of the accounts' installation.~~ The timestamp of first recorded activity within the account in `event`.
`unInstall_time`	Account Level	The timestamp of the accounts' uninstallation.

Table 5.1: The Fields Used in Data Table `events` Table

As mentioned in Section 4.6, the front-end events are the main events used in this research. Table 5.1 shows the field used in `events` from Google BigQuery.

While exploring the data, the cases that the initial `install_time` recorded in the table, which is later than the timestamp of the first activity of the account, is found. Therefore, to reduce the confusion, the semantic meaning of `install_time` has been changed to the first activity time of the account.

`session_id` distinguishes the event sequence with the threshold of 30 minutes break. According to Jones and Klinkner [27], the session can be defined as

1. as a set of queries to satisfy a single information need
2. A series of successive queries, and
3. a short period of contiguous time spent querying and examining results.

The timeout threshold is set to 30 minutes based on the conventional web session timeout identification, which is initiated from Cooley et al. [9]. If the interval between two events is greater than 30 minutes, the following event will be considered as the start of the new session.

`session_id` can be seen as an ideal trace. However, the session is eventually only used in test modeling and is abandoned and not involved in the final analysis.

`event_category` is a derived attribute from **event**. As mentioned in Section 4.6, `event` contains multiple elements. Take event `Performance Journey - Coachable Moment - Task Updated` as an example, `Performance Journey` indicates a specific component in module **Team House**; `Coachable Moment` indicates the pillar **Performance Coaching**; `Task Updated` indicates the user action and the targeted object. However, this is only one naming pattern. There are more patterns that have been discussed in **Challenge 4**. Therefore, only the first element is extracted to show the primary and the highest level event category. Nevertheless, the extracted attribute still contains various semantic meanings, including different levels of UI components, specific functions, and specific user flows.

5.1.2 Users

To capture the important analysis component of role and permission mentioned in Section 4.2.1, the table records users' roles, and the changes are applied.

user_id	timestamp	user_roles
a	2019-06-25 12:38:03 UTC	agent
a	2019-06-26 13:20:11 UTC	agent, team lead
...	...	...
b	2020-03-20 07:12:45 UTC	account owner

Table 5.2: Example of User Table

The table records the daily changes of each user. For example, user `a` can be an agent at first and becomes a team lead the next day. The records reflect the role changes in the clients' teams, which also influence the permission and the capacity the users could do on the Kaizo app.

The Data Table `users` table will further merged with `events`. Due to the limitation of SQL language, the work is done by Python and its library `pandas`. By comparing the timestamp in Data Table `events` and Data Table `users`, the role schemes are assigned to each event in terms of users. The following shows the steps of assigning role schemes.

1. Prepare the data frame from `events` and `users`.
2. Keep the earliest timestamp of each role scheme within users.
3. If the user has only one role scheme in `users`, the roles apply to all events triggered by the corresponding user.
4. Start to check the role timestamps by order.
5. If the user has more than one role scheme, the events with a timestamp smaller than the unassigned role scheme will be assigned as the preceding role scheme (See Figure 5.1).
6. The events with timestamps larger than the timestamp of the last role scheme are assigned as the last role scheme.

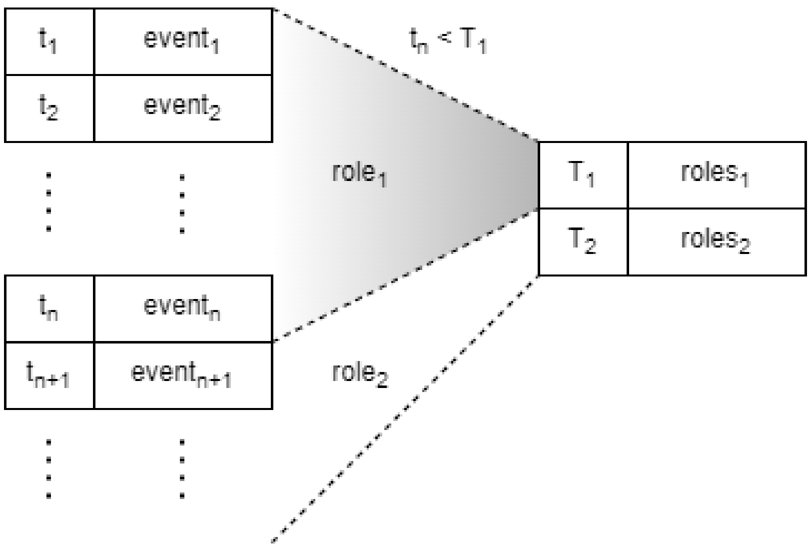


Figure 5.1: The Rule of Assigning Role Schemes in Single User

According to Section 4.3: Product Solutions and Research Pillars, the role schemes can be further categorized into manager and agent. Therefore, `user_roles` can derive to another event attribute `role_type` contains two values: `agent` and `manager`.

user_roles	role_type
agent	agent
agent, team lead	manager
account owner	manager

Table 5.3: Example of Categorize User Role Scheme

5.1.3 Tables Specific for Pillars

Through the progress of the project, some of the event attributes are further derived. The followings introduce the data tables that are used in other steps in the research. The table name with capital letter `E` denotes that there are multiple data tables with similar same table schema, which are the tables recording the detailed information of the front-end event logs.

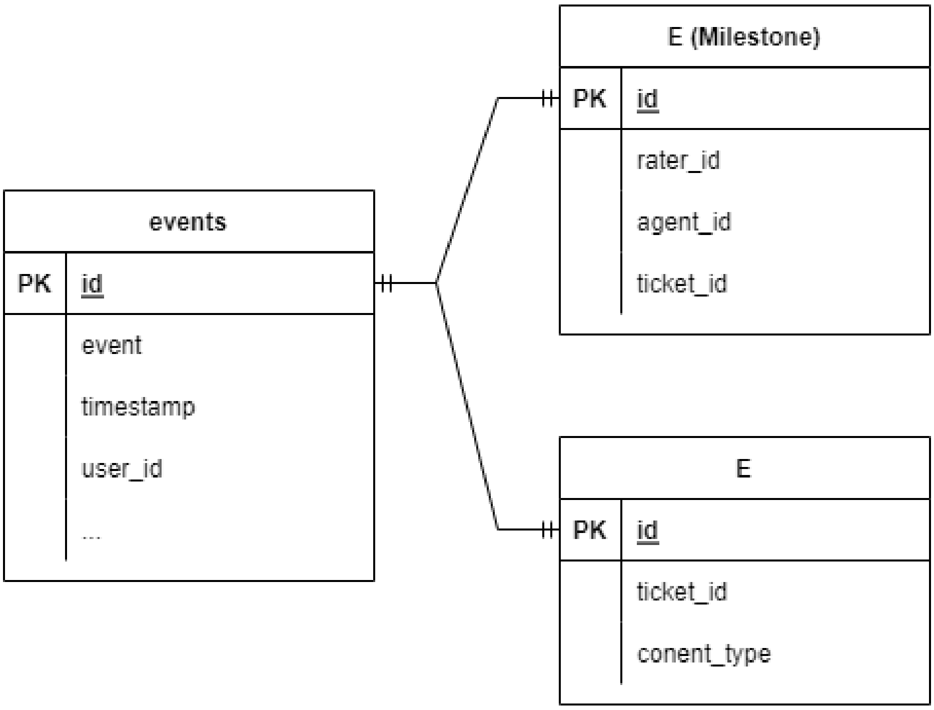


Figure 5.2: **Quality Assurance** Related Data Tables

Figure 5.2 shows the relations between the main Data Table `events` table and other tables carrying information related to **Quality Assurance**. Data Table `E` is not a single table but multiple tables that contain the column name `ticket_id`, which shows the event is relevant to Ticket O_Q and also **Quality Assurance**.

These table provides additional information `ticket_id` and `agent_id`, and `content_type`. `rater_id` is the same as `user_id` in Data Table `events`, which is already recorded. These additional fields are used to derive new event attributes

- `has_qa_info` in Section 6.2.2,
- `has_rated_ticket` in Section 6.2.3,
- `is_test`, `is_self_created` in Section 7.1.1,

and new event name in Section 6.3: Event Classification and Data Cleaning.

5.1.3.1. ***Missions***

The table is collected from API. The back-end data, however, is used to supplement the front-end events in this research. The initial table records every activity on the corresponding missions, but the table discussed here shows only the last record of each mission.

Field	Value Example	Description
`mission_id`	`156757`	The mission ID
`week`	`2022-06-06`	The starting date of the mission, which is always the start of the ISO week (Monday).
`agent_id`	`4b57e...05fec7e`	The agent who is assigned to the mission.
`user_id`	`77d2339...ed1fb60de`	The manager who create the mission.
`state`	`deleted`	The mission status.

Table 5.4: The Fields Used in Data Table `latest_missions` Table

5.1.3.2. ***Performance Coaching***

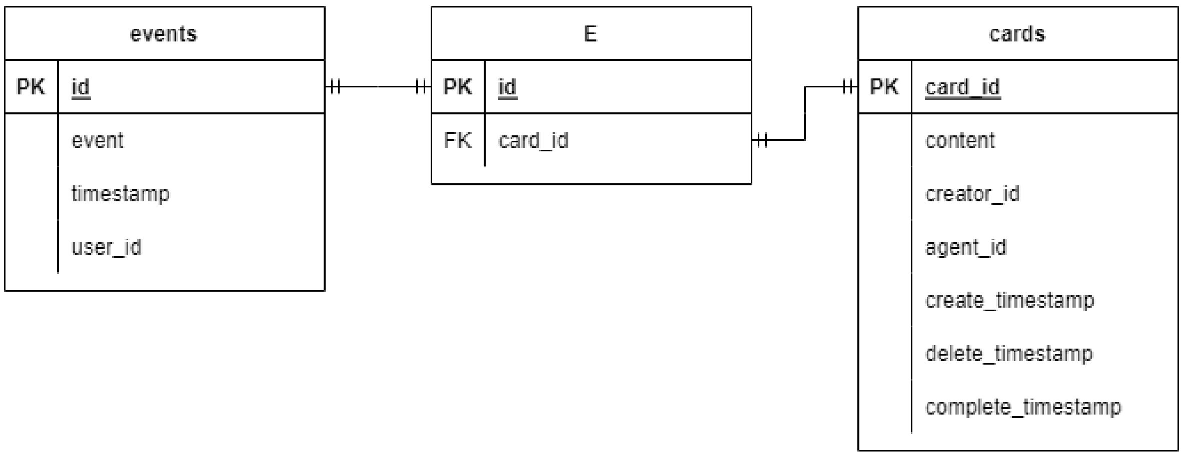


Figure 5.3: **Performance Coaching** Related Data Tables

Same as **Missions**, the information of **Performance Coaching** domain object O_C is recorded in another table and the source comes from API usage. With Data Table `E` as mediator, the attributes of O_C can be found in Data Table `cards` (See Figure 5.3). As mentioned above, Data Table `E` is not a single table. Under the current scenario of the tables with `card_id`, there are 78 tables considered as `E` are joined.

Further information of how to develop new event attributes is discussed in Section 7.1.1.

5.2 Data Collection, Cleaning, and Construction

5.2.1 Data Collection

To extract the main dataset, the following criteria are applied on `events`:

Criterion 1 - The test accounts are filtered out.
Criterion 2 - The client with multiple accounts are filtered out.
Criterion 3 - The records with `null` value in `event`, `user_id`, or `domain_id` are filtered out.
Criterion 4 - The date of `install_time` should be greater or equal to December 27, 2021.
Criterion 5 - Records from December 27, 2021 to July 3, 2022.
Criterion 6 - The account is not uninstalled or stayed on the Kaizo app for more than 7 days.

Criterion 4 aims to filter the new installed accounts to diminish the influence of **Challenge 7**, **Challenge 6**, and **Challenge 5**. The onboarding flow and price policies change over time. Extracting newly installed accounts can minimize such impact, and it is easier to consult stakeholders. Meanwhile, it is more valuable to analyze the current status of startup companies. Furthermore, as most of the interface and version differences are preserved for old clients, selecting newly installed accounts can best address the issue of interface difference.

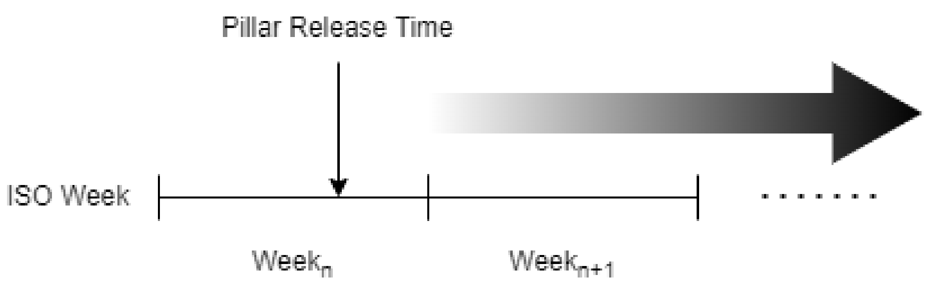


Figure 5.4: How to Select Installation Time

Each pillar has a different release time. As the research focuses on activation, it is necessary to consider when the features are introduced to clients. In general, the feature should be released before the account installs. As **Performance Coaching** released the latest, **Criterion 4** consider the release time of **Performance Coaching** as the standard. Considering the clients work in weekly base, the earliest `install_time` is set as the first day of the following ISO week of **Performance Coaching**'s release time (See Figure 5.4).

Initially, dividing the dataset by pillar and its `install_time` is considered. However, since **Quality Assurance** is the long-lasting feature, the records contain many abandoned events and functions. The idea of creating pillar datasets is failed to address **Challenge 7**, **Challenge 6**, and **Challenge 5**; meanwhile making unnecessary larger project scale.

For **Criterion 5**, The starting point of the timeframe is inherited by the **Criterion 4**, and the ending time is the end of ISO Week 26 in 2022.

Criterion 6 has been modified. The initial idea is to analyze accounts that are retained on the Kaizo app without uninstallation. However, the strict criterion deteriorates **Challenge 2** that there are not enough accounts with milestone events to analyze. Therefore, the condition that the accounts that had installed the Kaizo app more than 7 days but eventually uninstalled are considered.

After defining the main criteria, `users` are joined with `events` by the method discussed in Section 5.1.2.

A test dataset is extracted from the primary dataset by simple filtering criteria. The test dataset contains only one week of data with accounts that are not uninstalled. The smaller dataset is used for testing data cleaning and consulting stakeholders. The finalized methods can be further applied to the primary dataset.

5.2.2 Trace Identification

Two testing fuzzy models are created in order to define the trace. Figure B.1 apply `userId` as trace, while Figure B.2 apply `session_id` as trace. Apparently, the two models are too spaghetti-like and failed to extract any insights, but the session trace seems more promising. Therefore, the first trace considers a session to build process models.

5.2.3 Trial Cleaning

5.2.3.1. First Trial

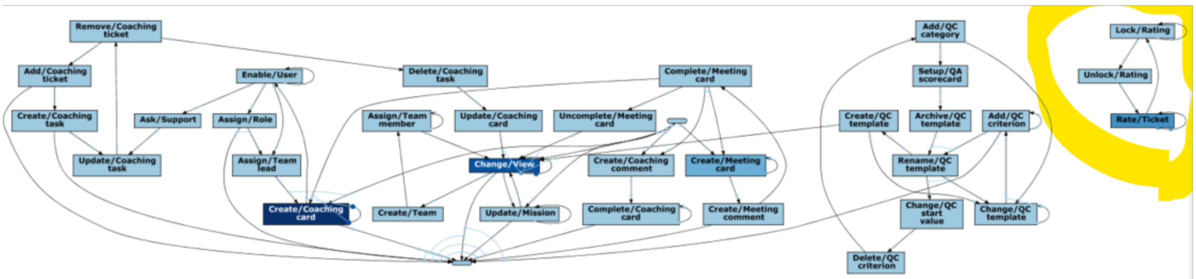


Figure 5.5: Example of the Broken Model

The first trial cleaning removes all the events that are not relevant to the pillars from the test dataset. However, the test model generated from heuristic miner shows the break processes 5.5), which is not allowed in casual net representation.

5.2.3.2. Second Trial

As the session is the first trace candidate, the second trial only removes the invalid sessions with only.

- events that are not triggered on the Kaizo app,
- events triggered outside the Kaizo app, such as the Zendesk navigation bar,
- notification events that users do not trigger,
- navigation events,
- error events, and
- default landing page visit event of Kaizo app;

and the session with only one event.

The result (See Figure B.3) is still difficult to read because there contains noisy events in valid sessions. Therefore, it is necessary to define the noisy events that could be filtered out globally to reduce the number of events and fuzzy nodes.

The first attempt is to extract sessions with milestone events to simplify the models. Appendix C shows the fuzzy models from sessions with milestone events in Table 4.3. Stakeholders point out several noisy events and provide more semantic meanings of the events to improve business understanding. Meanwhile, retention patterns are also identified during this phase. The detail will be discussed in Section 6.1.

The second attempt is to assign `event_category` as event to create new XES file and model (See Figure B.4). With the assistance of the XES summary and fuzzy model, the meaningless start events and events that interrupt the operation flow are identified:

- The error events
- The notification and modal events triggered by the system and are not related to pillars
- The event triggered outside the Kaizo app

5.2.4 Trace Re-identification

Ideally, the session should start from the default landing page. However, the start events in the second trial, after globally filtering out the noisy events, are still too diverse that the expected start events are only 65% of the total start events of sessions.

The result might occur because the embedded Kaizo app will not reload the page after 30 minutes so that users can continue the tasks after long breaks. Semantically, users might be busy working on Zendesk for operation and using the Kaizo app fragmentally.

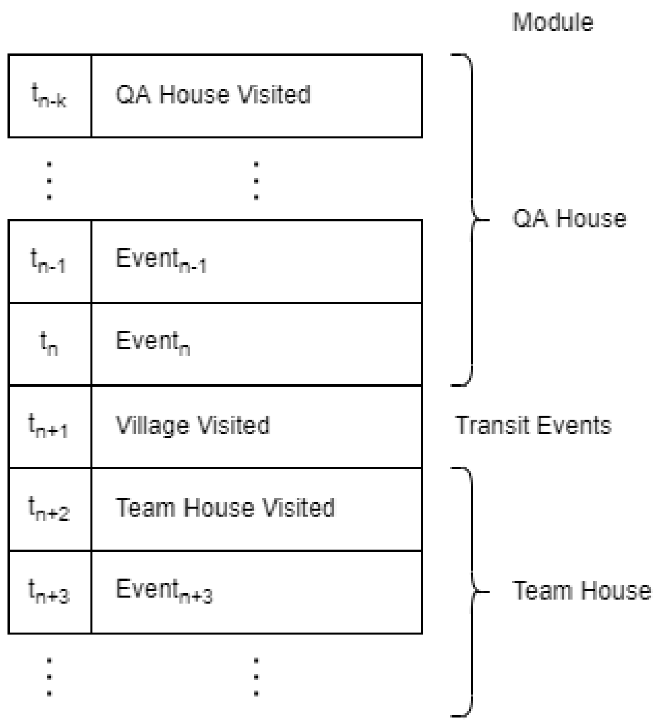


Figure 5.6: Assign `module` to Each Event

To validate the previous assumption, `module` and `module_id` are introduced. According to the understanding from Section 4.5, the Kaizo app contains fixed modules with corresponding visit events. Except for the transit visit events, such as the default landing page and navigation shortcut events, other events can be assigned a module by navigating the closed preceding visit events of each module (See Figure 5.6).

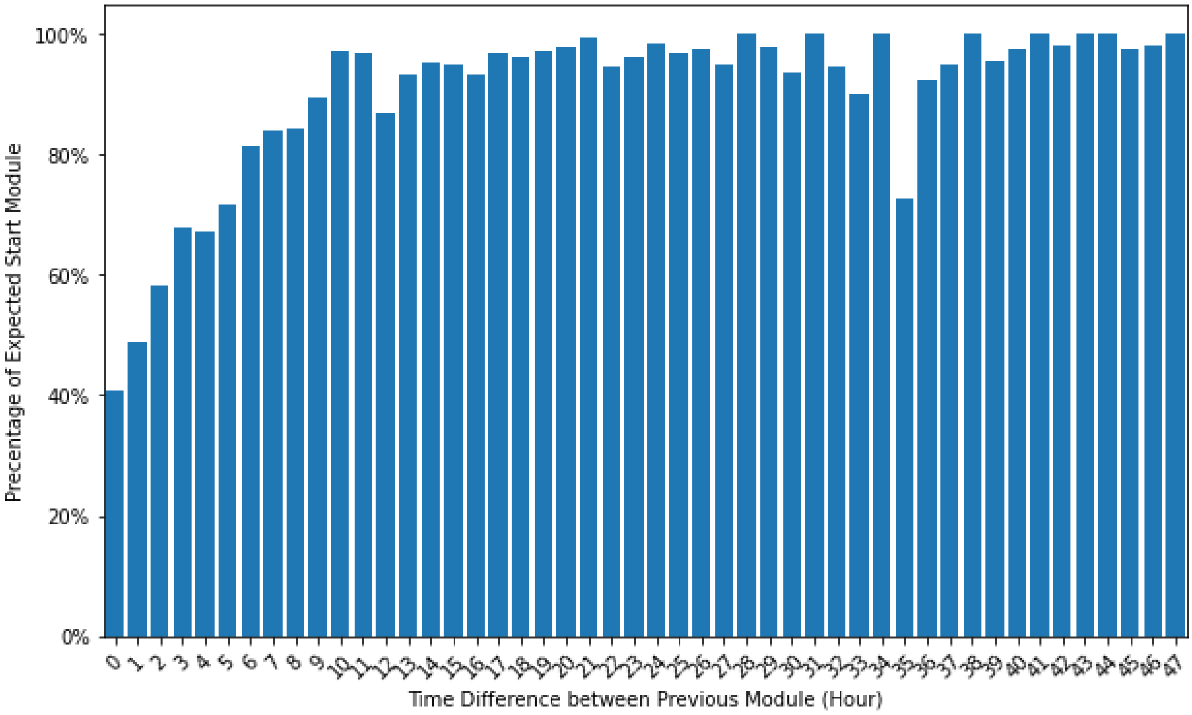


Figure 5.7: The Percentage of Expected Start Events and the Module Time Difference

After that, `module_id` can be defined. Switching the module within the user or the starting module of the user will be assigned a serial number. By calculating the hourly difference between the timestamp of a module's start event and the preceding module's end event, as well as the percentage of the expected start events, it is clear that users worked on a daily basis (See Figure 5.7). Therefore, the session as the trace input should be reconsidered.

As a result, `user_id` combined with the date of `timestamp` is used as the new trace for the following modeling. The methods and insights are applied to the primary dataset.

5.3 Summary

Event Attribute	Description	Cross-Reference
domain_id	The anonymous domain ID	Chapter 5
user_id	The anonymous user ID	Chapter 5
trace	The concatenation of the date of timestamp and user_id	Chapter 5
role_type	The type of user role.	Chapter 5, Section 4.2.1
module	The UI module where the event is triggered	Chapter 5, Section 4.5
module_id	The serial number of module	Chapter 5
timestamp	The time when the event is triggered	Chapter 5
event	The event name recorded in data warehouse	Chapter 5
week_to_install	The ISO week difference from the install_time to timestamp	Chapter 5
ticket_id	The domain object of **Quality Assurance** (O_Q)	Section 6.2.1
card_id	The domain object of **Performance Coaching** (O_C)	Section 6.2.1
is_self_created	Whether O_Q and O_C are rated/created by same user of the object owner	Section 7.1.1
is_test	Whether the object is test object or action	Section 7.1.1
has_qa_info	Whether the ticket and scorecard display **Quality Assurance** related information	Section 6.2.2
has_rated_ticket	Whether the agent has tickets that was being rated in the previous and current week	Section 6.2.3
mission_status	Check if the agent has mission or not and whether the missions are activated	Section 6.2.3
event_action	The action derived by the initial event	Section 6.3
event_object	The domain object derived by the initial event	Section 6.3
event_name	The concatenation of event_action and event_object	Section 6.3

Table 5.5: Event Attributes

As the initial data preparation, this chapter extracts the main dataset needed for this research. Also, data cleaning and data construction are applied. Data cleaning contains several iterations and the test modeling to identify noisy events. Meanwhile, some event attributes are derived, and the ideal trace is recognized. The result can be seen in Table 5.5. The table also includes the attributes derived from other sections.

6 RETENTION CLUSTERING

6.1 Retention Identification

As mentioned in Section 5.2.3, the models in the second trial are used to consult stakeholders. In this section, stakeholders provide more semantic meanings of the processes, which can benefit the retention clustering.

Appendix C shows four fuzzy models with yellow marks, indicating the pointed process, which could be seen as retention patterns. The input data is extracted from the test dataset with the sessions containing the pre-identified milestone events (Table 4.3).

Pillar	Managers' Expected Behaviors	Agents' Expected Behaviors
Quality Assurance	Figure C.1	(milestone events not clear)
Missions	Figure C.2	Figure C.3
Performance Coaching	Figure C.4	(milestone events not clear)

Table 6.1: Milestone Events and Test Models

After reviewing the models, more insights are created. From the manager side, more events co-occur with milestone events. Apart from the milestone events, other activities assisting the achievement of milestone events will occur, such as reviewing performance metrics. Also, some events are followed by milestone events, such as creating comments or notes. In general, the activities related to manipulating the object of interest occur.

Agents have limited permissions, so it could be challenging to see whether they are interested in the features or not. However, agents reviewing their performance and metrics can be considered proactively using the function, no matter which pillar. In order to capture agents' activities, additional information is required.

Performance Coaching is a unique pillar that the triggering of the milestone events does not limit to permission. That is, agents can create coaching cards on their own. Furthermore, the processes may turn to offline communications. The specific characteristics of **Performance Coaching** should be further discussed.

Meanwhile, according to AARRR model [42], users in the retention phase contain repeated behaviors. Stakeholders also agreed that the accounts which can be considered in the retention phase should show weekly patterns of the mentioned pillar. Specifically, the identification of retention should be defined based on the overall activities of the accounts because the users' activities can break due to vacation, and a colleague can undertake the required work.

6.2 Data Construction

Based on a further understanding of user behavior, more data should be merged in order to capture more details. The primary purpose of this section is to add more detailed information for analyzing agents' behavior in order to identify the missing agents' milestone events.

6.2.1 Domain Object

Pillar	Domain Object	ID (Attribute)
Quality Assurance	**Ticket**	`ticket_id`
Missions	**Mission**	(not capable of integration)
Performance Coaching	**Coaching/Meeting Card**	`card_id`

Table 6.2: The ID of domain objects

From the previous data understanding, the domain object id of **Quality Assurance** and **Performance Coaching** is capable of integrating to the dataset. The primary key of **Ticket** is `ticket_id`, and **Coaching/Meeting Card** is `card_id`.

timestamp	event_action	event_object	ticket_id	
t_1	Visit	Scorecard	--	
t_2	Open	Ticket	001	
t_3	Change View	Ticket	--	<- Add `001`
t_4	Create Rating	Ticket	001	
t_5	Open	Ticket	002	

Table 6.3: Example of An User's Activity on Ticket

Each ticket-related records contains a value of `ticket_id`. The relationship can be found in Figure 5.2 by joining on `id` of each record in Data Table `events`. However, some of the data are missing in the subtables in Figure 5.2. In such case, `ticket_id` is determined by the preceding and following events, as seen in Table 6.3.

The joining and `ticket_id` determination is done after later Section 6.3, so the explanation is displayed with the derived `event_action` and `event_object`. For better thesis structure, the explanation of event attribute `ticket_id` is added here.

The relationship between Data Table `coaching` and Data Table `cards` is simple and comprehensive. Therefore, `card_id` can directly merged by the relationship found in Figure 5.3.

6.2.2 Navigation Behavior in Quality Assurance

Except **Quality Assurance**, other pillars contains clear events showing the `Open` or `View` actions on domain objects. On the other hand, As mentioned in Section 4.5, **Scorecards** as the fundamental feature and the UI component, is designed to show in various modules. The information in **Scorecards** could be either relevant or irrelevant to **Quality Assurance**. The events that could indicate the navigation behavior of **Quality Assurance** information are mixed within **Scorecards** events.

Therefore, according to Figure 5.2: **Quality Assurance** Related Data Tables, `has_qa_info` is determined based on `content_type`. The scorecard with information relative to **Quality Assurance** would have the `has_qa_info` of `True`.

The attribute can be further derived to the **Quality Assurance** domain object to see whether or not the users are manipulating **Ticket** with the rating. Whether the ticket contains rating information can be further determined by `timestamp` that the ticket contains rating events before can be assigned `True` in `has_qa_info`, as seen in Table 6.4

timestamp	event_action	event_object	ticket_id	has_qa_info
t_2	Open	Ticket	001	False
t_3	Change View	Ticket	001	False
t_4	Create Rating	Ticket	001	False
t_5	Add Note	Ticket	001	True
t_6	Delete Note	Ticket	001	True

Table 6.4: Example of Deriving `has_qa_info`

As a result, the value of `has_qa_info` can be defined as:

`True` The corresponding ticket has rating scores, or the content type of scorecard is relevant to **Quality Assurance**.

`False` The corresponding ticket does not have rating scores, or the content type of scorecard is not relevant to **Quality Assurance**.

`null` Not relevant.

`has_qa_info` will be used in event classification later to rename the valuable events for research. The understanding of the corresponding ticket of event logs helps the modeling of domain object in Figure 6.3. Furthermore, the case of duplicated **Quality Assurance** milestone event is detected. The issue is solved in the next Section 6.3.

6.2.3 User Status

From the initial dataset, it is invisible to check users' status when the events are triggered. Users with the domain objects could behave differently than those who do not own the objects.

Involving those without objects would potentially reduce the frequency of the valuable events during modeling. Therefore, the event attributes `has_rated_ticket` and `mission_status`, are derived.

`has_rated_ticket` uncovers whether the agent has the tickets rated in the previous and current week. To be more specific, a manager rate the ticket in w_n. Meanwhile the agent who is rated, found in `agent_id` from detailed event tables (See Figure 5.2: **Quality Assurance** Related Data Tables), and then his or her events are triggered in w_n and w_{n+1} will be assigned considered as containing rated tickets. The value and definition can be seen as follows.

`True` The agent has rated tickets that are rated in the previous and current week.
`False` The agent has no rated tickets that are rated in the previous and current week.

Data Table `latest_missions` is used to check whether the agent has the mission or not because the front-end event failed to record the assigned agent, while the manager's information is well-recorded and easy to cluster. From the understanding of the business and **Missions**, the feature is operating by team base. That is, not all the agents have missions to carry out, even if their colleague has missions that week. Meanwhile, some of the events of **Missions Modal** can be considered as noisy events because the component contains no actionable information. Therefore, the event attribute `mission_status` is added to target the agents with valid missions that week.

The attribute is created by joining Data Table `latest_missions` mentioned in Section 5.1.3.1 on `agent_id` and `week`. First, deleted missions are filtered out. Second, `mission_status` is created on a weekly basis in terms of users. The value is defined as follows:

`none` The agent did not have any non-deleted mission that week.
`inactivated` The agent had non-deleted missions that week but did not activate any of them this week.
`activated` The agent had non-deleted missions and activated the missions this week.

6.3 Event Classification and Data Cleaning

This section is an essential pivot of the research job. The event classification and corresponding data cleaning diminish the interference of **Challenge 5**, the interface difference from different modules, devices, and versions.

After the event classification, the number of events successfully reduce from 338 to 128. Event classification addresses **Challenge 4** to rename the event based on motivations and actions. **Challenge 5**, **Challenge 7** are also addressed due to the removal of UI component naming. Meanwhile, the data size was reduced dramatically, making the modeling more efficient and easy to explore.

6.3.1 The Focus on Domain Objects

Inspired by Beaudouin-Lafon [3]'s framework, the interaction can be divided into objects of interest and actions. Therefore, the event attributes `event_action` and `event_objects` are defined. Where the events are triggered is recorded in another column, `event_location`. By roughly dividing the original event name into three columns, selecting data becomes more flexible and analyzable.

First, each pillar has its own object of interest, as discussed in Section 4.3.2, users can manipulate the attributes of the domain objects. For instance, `Note` is the attribute of **Ticket**. `Note` can be created, edited, or deleted. Users can also influence the existence of the object. For example, users can directly create or delete the domain object of **Missions**. Apart from pillars, there exist other objects of interest in the app. The same framework can also apply to other objects.

6.3.2 Specific Flows

On the other hand, many events are diminished due to the lack of semantic meanings, especially in specific onboarding flows and settings. For instance, the events representing the transition to the next step, such as `Click to Next` is removed because the following activity represents similar semantic meanings. The batch applies events of the sequences, for example, which records the changes with the button click `Save All` at the end for batch applying the change, will be removed. Meanwhile, if the button clicks do not occur, meaning that the changes do not apply, the whole sequence will be removed. The followings are examples of particular user flows.

- ~~`Open Form`~~, `Submit`
- ~~`Submit`~~
- `Config`, `Config`, ~~`Save All`~~
- ~~`Config`~~, ~~`Config`~~
- `First Step`, ~~`Click to Next`~~, `Second Step`

6.3.3 Derive New Event Node

This section aims to reduce the number of event nodes. However, some special cases need to derive new event concept names to identify the focused activities during modeling. The explanation of `has_qa_info` in Section 6.2.2 is a good example that the information mixed together in scorecards. Therefore, the scorecard events with `has_qa_info` is `True` are provided with a new name.

6.3.4 Duplicated Events

timestamp	event_action	event_object	ticket_id	
t_1	Open	Ticket	001	
t_2	Create Rating	Ticket	001	remove
t_3	Create Rating	Ticket	001	remove
t_4	Create Rating	Ticket	001	remove
t_5	Create Rating	Ticket	001	remove
t_6	Create Rating	Ticket	001	keep
t_7	Open	Ticket	002	
t_8	Create Rating	Ticket	002	remove
t_9	Create Rating	Ticket	002	remove
t_{10}	Create Rating	Ticket	002	keep
t_{11}	Open	Ticket	003	

Table 6.5: Example of Folding Duplicated Events

After the creation of `event_action`, `event_object`, and the derivation of `ticket_id` in Section 6.2.2, the duplicated events are identified. The mechanism of triggering `Create Rating` on **Ticket** is based on each click. Since the rating involves multiple criteria, repeated sequences occur in many cases. The duplicated issue highly influences the resulting modeling that `Create Rating` in all the models is remarkably significant. Since the duplicated events are difficult to resolve from *heuristic miner* [17], the event sequence then is folded by keeping only the last action within the same ticket and user window.

6.4 Data Preparation for Modelling

Recalling Section 5.2.4, a user's working day can be considered a case. Users' behavior can be modeled by selecting the `trace` with the milestone actions. The data selection for the pillar and role with pre-identified milestone actions follows the above framework for modeling. Agent's behavior modeling of **Performance Coaching** and **Quality Assurance** adopts a different method to select data, which will be explained in the corresponding section.

```
SELECT *
FROM table
WHERE week_to_install >= 3
QUALIFY COUNT(CASE WHEN event={milestone} THEN id END) OVER trace >= 1
WINDOW trace AS (PARTITION BY user_id, date ROWS BETWEEN UNBOUNDED PRECEDING
    AND UNBOUNDED FOLLOWING)
```

Listing 6.1: GBQ Example

Since this chapter is modelling the retention activities, to avoid the influence of dynamically changed onboarding flow, the records from the first two weeks are filtered out.

Furthermore, since `Visit` events have extremely high frequency, records with `event_action` equal to `Visit` are filtered out. Records with `event_action` equal to `Change View`, indicating the changing the information displayed on the screen, such as applying filters, are also integrated into one node to simplify the model.

6.5 Pillar: Quality Assurance

6.5.1 Manager

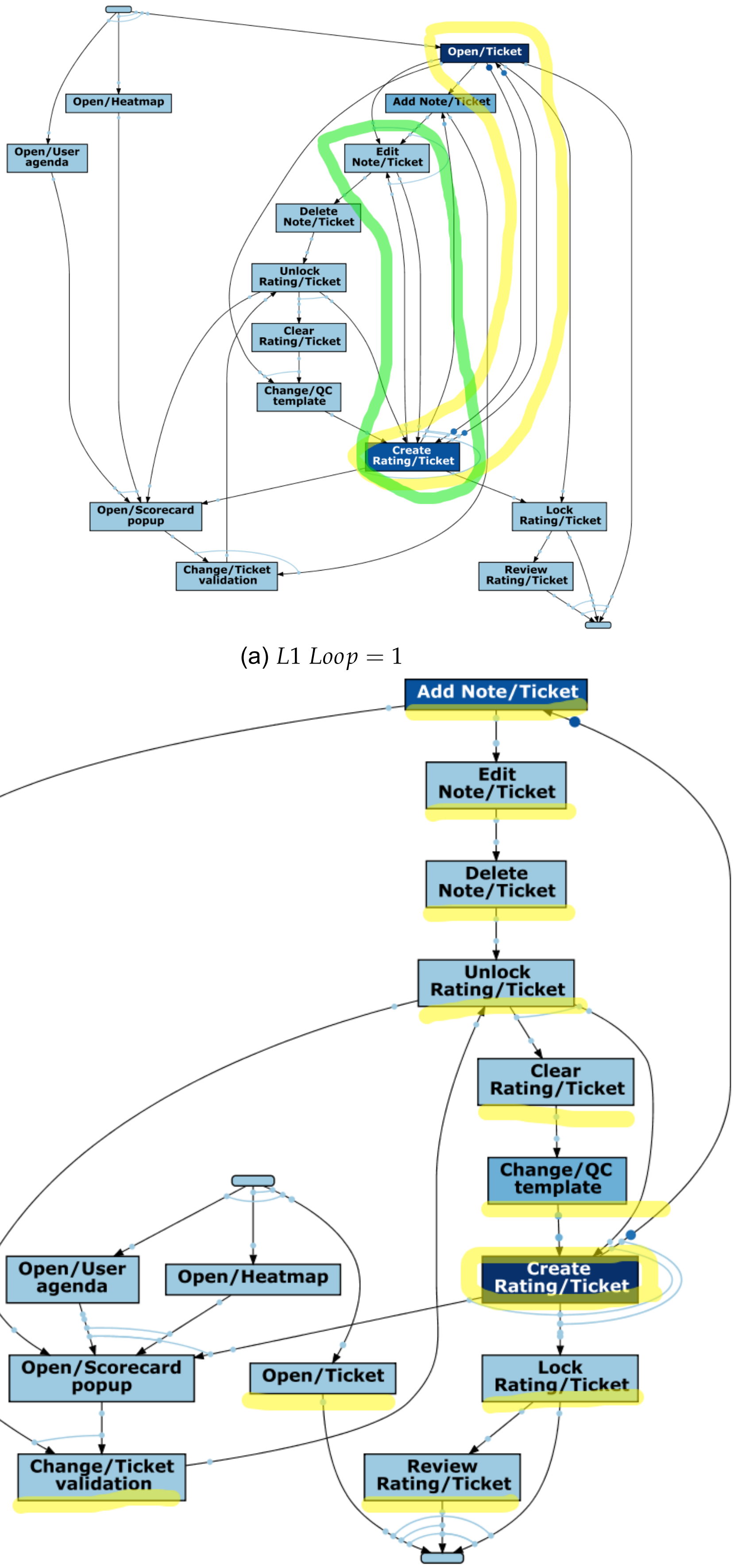
The dependency of the milestone node in the first model is highly influenced by self-loop. Although the previous data cleaning had removed the duplicated milestone action within the same ticket, the repeated activities occur between tickets that managers keep rating tickets within a day. To solve the repeated behavior issue, the second model removes *length-one loops*, as seen in Figure 6.1a. However, the loops between two nodes occur. Managers have back-and-forth behavior between `Create Rating/Ticket` and `Open/Ticket`, as well as `Create Rating/Ticket` and `Edit Note/Ticket`.

As a result, the next model configures the *length-two loops threshold* to 1 in order to create a more intuitive model. The result can be seen in Figure 6.1b. The follow-up activities of milestone action, `Create Rating`, manipulate other domain object attributes in **Quality Assurance**, such as adding notes on tickets.

Binding	Events	Frequency
Input	`Change QC Template`	91%
	`Unlock Rating`	9%
Output	`Add Note`	74%
	`Add Note`, `Lock Rating`	13%
	`Lock Rating`	12%
	`Add Note`, `Lock Rating`, `Visit Scorecard Popup`	0%

Bindings are extracted from Figure 6.1b: $L1\ Loop = 1, L2\ Loop = 1$.
`red event` denotes the event is acting on the domain object, **Ticket**.

Table 6.6: Bindings of Manager's Milestone Event (**Quality Assurance**)



(a) $L1\ Loop = 1$

(b) $L1\ Loop = 1, L2\ Loop = 1$

Figure 6.1: **Quality Assurance**: Manager's Behavior

6.5.2 Agent

Similar to Section 6.5.1: Manager, the input data consider whether a single `trace` (The combination of `date` and `user_id`) fulfills the following conditions:

1. The user is an agent (R_A).
2. The agent had rated tickets that had been rated in the current week or the previous week (`has_rated_ticket`).
3. The agent triggered the ticket-related events with the tickets with rating (`has_qa_info`, the additional symbol `*` is added for modeling event), or reviewed the overall **Quality Assurance** score (`Review Rating`).

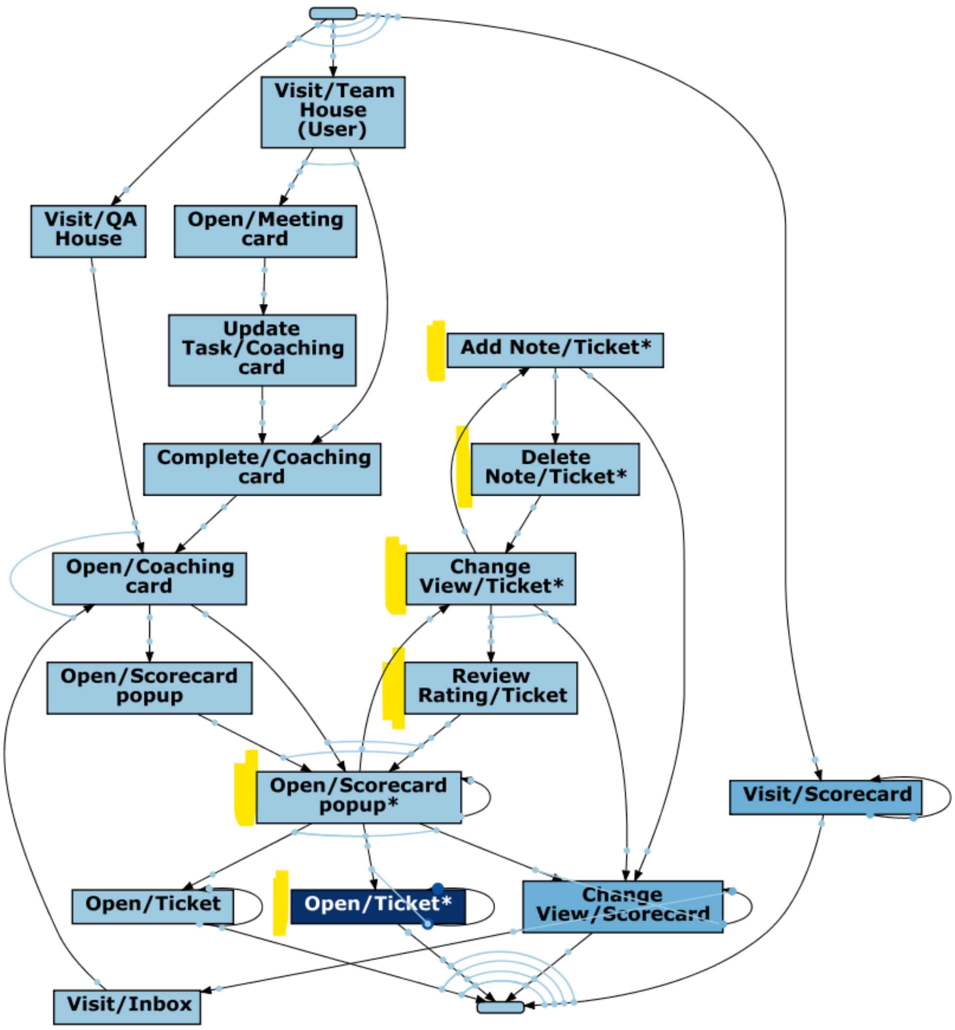


Figure 6.2: **Quality Assurance**: Agent's Behavior

In Figure 6.2, The highlighted nodes are associated to the third condition. The result shows that `Open` action on **Ticket** is the most frequent behavior for those who have rated tickets.

6.5.3 Domain Object Lifecycle

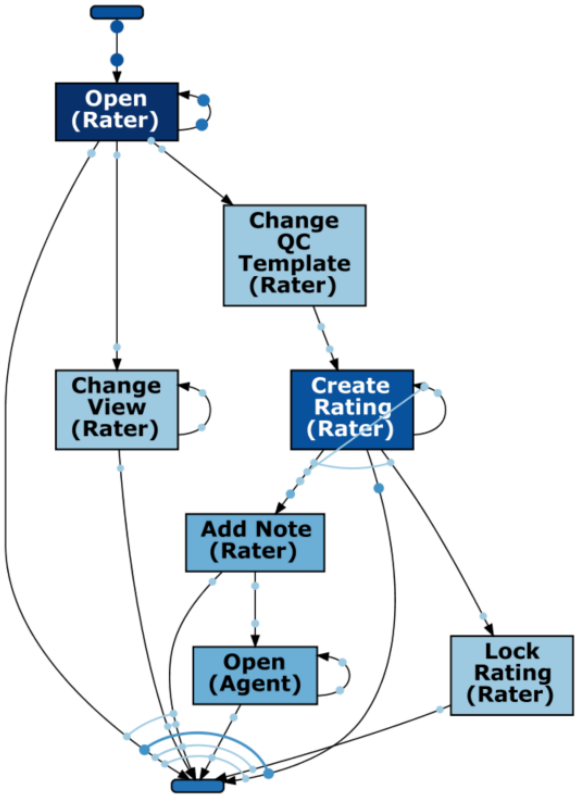


Figure 6.3: Tickets with Rating

With interest on the domain object, the third model considers using `ticket_id` as the trace, following the interest of the domain object. The nodes further added the notes to distinguish who triggers the events. Figure 6.3 shows the same pattern as found in **?? ??**, with meaningful prerequisite of opening the tickets. By heuristic filtering, only `Open` from agent is remain in Figure 6.3.

6.6 Pillar: Missions

6.6.1 Manager

To discover manager's behavior on **Missions**, the days of the manager containing milestone action of **Missions** are selected. During the modelling, *length-one loops* and *length-two loops* are removed in order to filter out the noisy repeated patterns (See Figure 6.4).

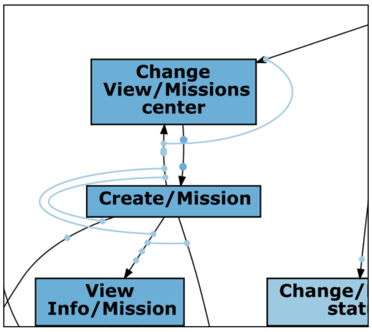


Figure 6.4: The Example of Repeated Pattern

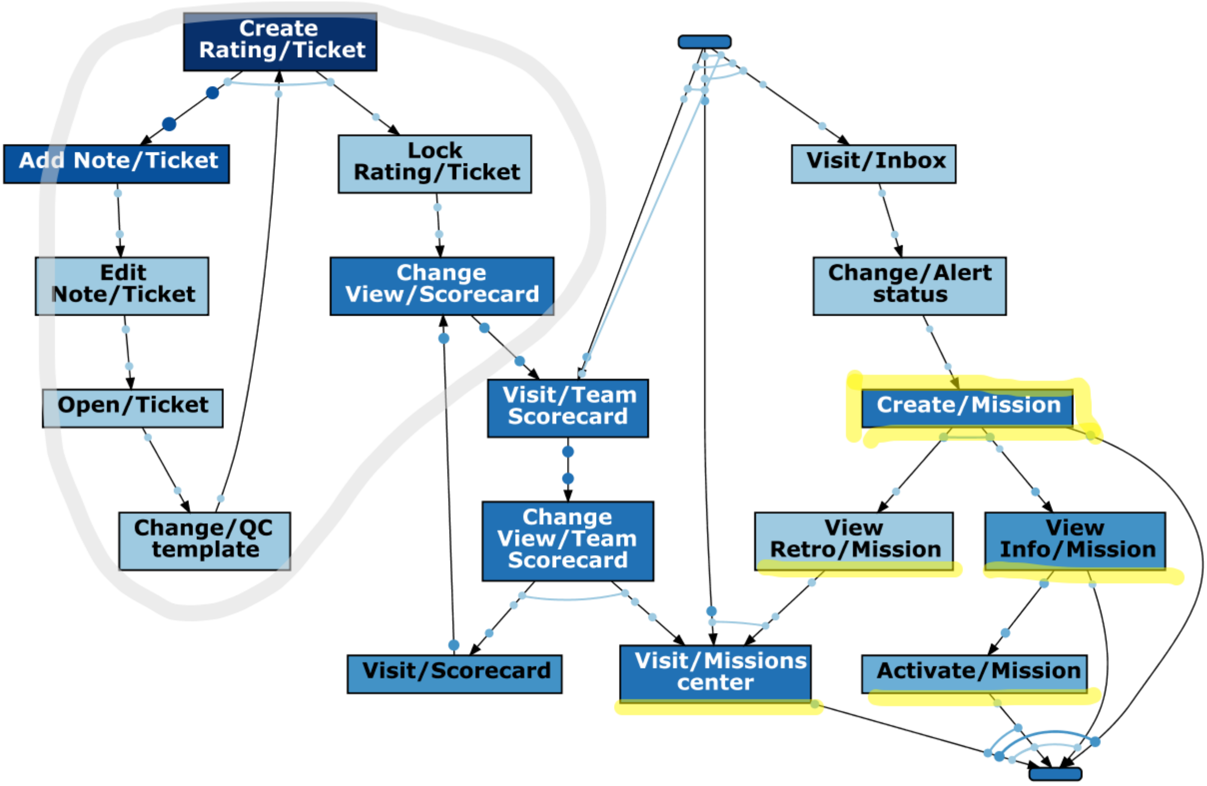


Figure 6.5: **Missions**: Manager's Behavior

Figure 6.5 shows the model without *length-one loops* and *length-two loops*. The nodes highlighted in yellow are the actions directly related to **Missions**; the left process is the discovered milestone process of **Quality Assurance** managers. Table 6.7 provides the bindings of the essential nodes.

Focused Node	Binding	Events	Freq.
(Start)	Output	`Visit`	29%
		`Visit`, `Visit (Team Scorecard)`	19%
		`Visit`, `Visit (Inbox)`	18%
		`Visit (Team Scorecard)`, `Visit (Inbox)`	14%
		`Visit`, `Visit (Team Scorecard)`, `Visit (Inbox)`	11%
		`Visit (Team Scorecard)`	6%
		`Visit (Inbox)`	3%
Visit Module	Input	`(Start)`	57%
		`Change View (Team Scorecard)`	23%
		`View Retro`	11%
		`(Start)`, `View Retro`	10%
	Output	`(End)`	100%
Milestone	Input	`Change Alert Status (Inbox)`	100%
	Output	`(End)`	51%
		`View Retro`, `View Info`	43%
		`View Info`	5%
(End)	Input	`Create`, `Visit`	59%
		`Activate`, `Visit`	28%
		`Activate`, `View Info`, `Visit`	13%

`red event` denotes the event is acting on the domain object, **Mission**.
`blue event` denotes the event acting on the associated module of **Missions**, which is **Missions Center**.

Table 6.7: Bindings of Focused Nodes in Manager's Behavior Model (**Missions**)

In general, the manager's task is precise. Managers start by reviewing performance metrics in other modules and attend **Missions Center** for the last stop for milestone action, creating missions. **Missions** and **Quality Assurance** also combined on the same day, as seen in the process with a grey circle.

6.6.2 Agent

Monday	Tuesday	Wednesday	Thursday	Friday
$E'_{M,R_A} \notin E_1$	$E'_{M,R_A} \in E_2$	$E'_{M,R_A} \notin E_3$	$E'_{M,R_A} \notin E_4$	$E'_{M,R_A} \notin E_5$
(Not selected)	The Day with Activation Figure 6.6a	After Activation Figure 6.6b		

Table 6.8: Example of Trace Selection

This subsection first selects the week with the activated mission. That is, the agents had missions that week and activated the missions. To simplify the model, `role_type` is limited to `agent`. The input data follow the same trace selection that an agent's workday is considered as a case. However, the event logs are divided into two subsets because the concept of **Missions** is designed to be a weekly activity.

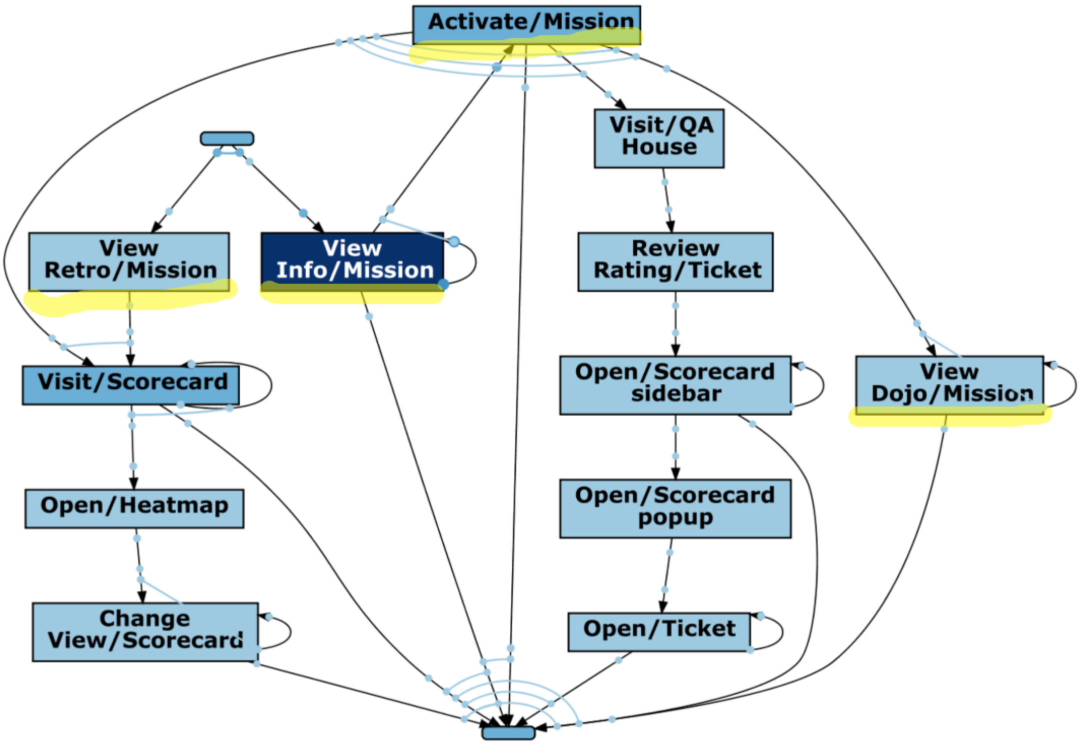
Unlike other trace selections with milestone events in this chapter, the agent's behavior in **Missions** should not consider only the day with milestone events E'_{M,R_A}. The days after activation within the week might contain potential behavior patterns such as reviewing the mission progress or performance, according to stakeholders. Table 6.8 shows an example of how to divide the two subsets. Consider an agent work five days a week and activating his/her missions on Tuesday. The events on Tuesday will be categorized to the first model Figure 6.6a: The Day with Activation. The following traces E_3, E_4, and E_5 will be categorized to the second model Figure 6.6a: The Day with Activation. The event sequence on Monday are not considered.

Figure 6.6a and Table 6.9 show how agents act the day when they activated the missions (E_2). According to the output binding column of the milestone events in Table 6.9, agents show interest in reviewing their performance afterward. Agents either visited **Dojo Room** to see the gamified mission performance or visited the scorecards in **Scorecard** or **QA House** to review the overall performance.

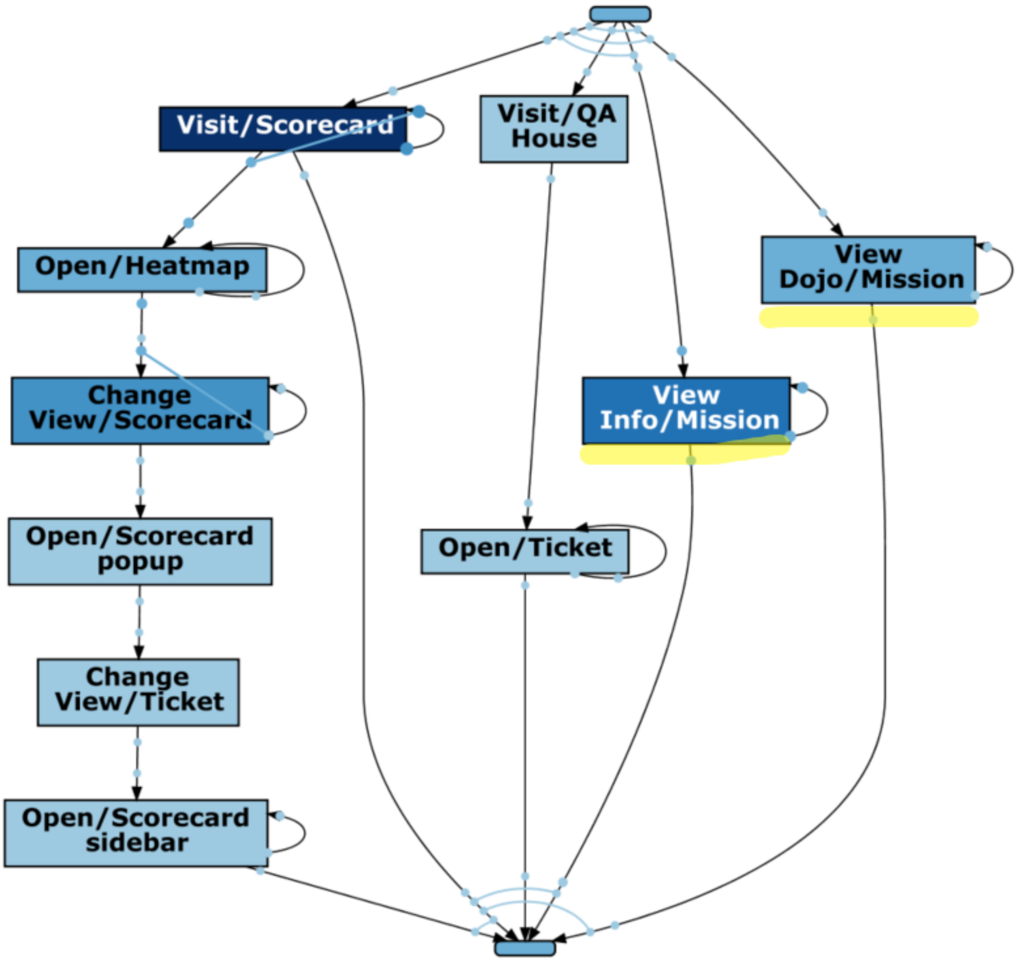
Focused Node	Binding	Events	Freq.
Milestone	Input	`View Info`	100%
	Output	`(End)`	44%
		`Visit (Scorecard)`	23%
		`View Dojo`	15%
		`View Dojo`, `Visit (QA House)`, `Visit (Scorecard)`	7%
		`View Dojo`, `Visit (Scorecard)`	5%
		`Visit (Scorecard)`	5%

Bindings are extracted from Figure 6.6a: The Day with Activation.
`red event` denotes the event is acting on the domain object, **Mission**.
Agents do not have permission to visit the associated **Missions** module.

Table 6.9: Bindings of Focused Nodes in Agent's Behavior Model (**Missions**)



(a) The Day with Activation



(b) After Activation

Figure 6.6: **Missions**: Agent's Behavior

Figure 6.6 shows agent's day after activation within the week of mission (E_3, E_4, E_5). Although the relevant nodes (marked as yellow) do not have any strong relationship to the interest of reviewing performance, Figure 6.7 still shows that agents reviewed their performance within the day.

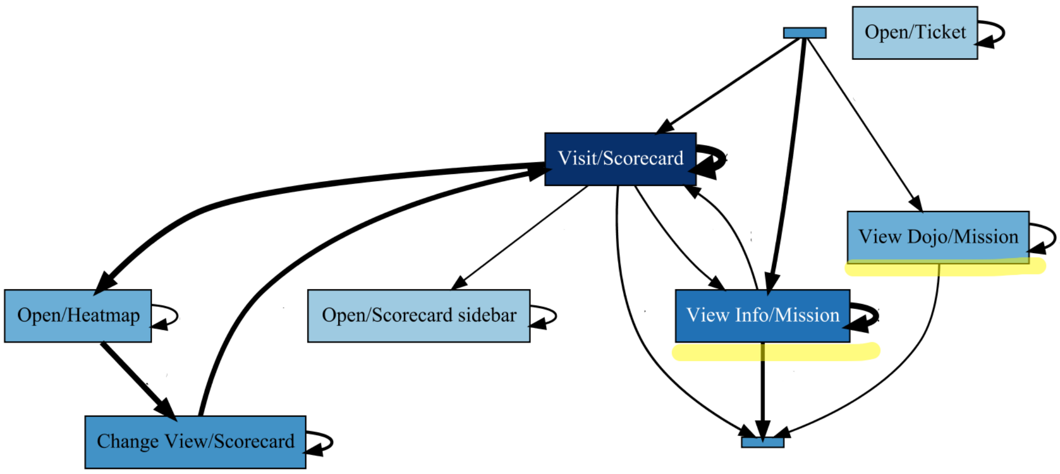


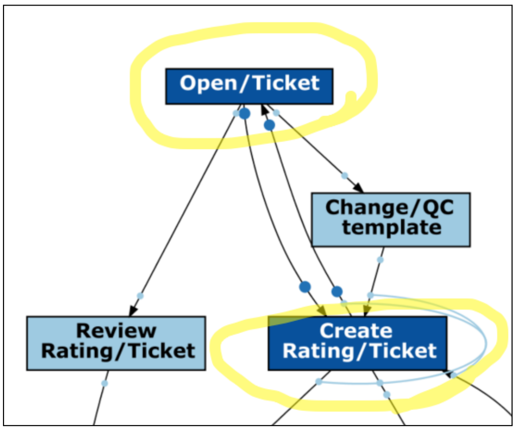
Figure 6.7: **Missions**: Agent's Behavior After Activation (Part of Directly-follows Graph)

6.7 Pillar: Performance Coaching

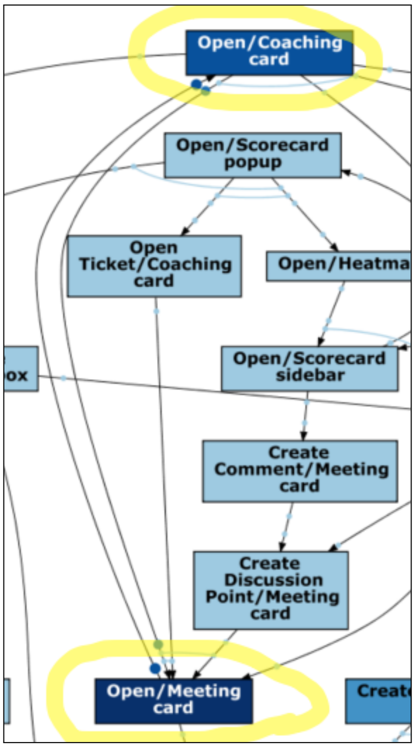
6.7.1 Manager

The traces, similar to other pillars, with milestone events in **Performance Coaching** are selected to model the manager's behavior. The initial model identifies some of the obligation steps before milestone action. The prerequisite steps, including visiting the corresponding module, are removed from the data to capture more casual relationships between functions. The trace containing self-creating milestone events is filtered out to simplify the model. Therefore, the cases of agents creating cards for themselves are excluded.

In the initial modeling, the self-loop occurs on the milestone node that more than 50% of the input bindings and the output bindings are the node itself. Therefore, the author rebuilt the model without *Length-one Loops*. Meanwhile, *all-tasks-connected heuristic* is applied because it is the more meaningful.



(a) Loops about Milestone Action in **Quality Assurance**



(b) Loops Reviewing Domain Objects in **Performance Coaching**

Figure 6.8: The Loops between Nodes

The model without *length-one loops* contains loops between two nodes (See Figure 6.8). Therefore, the next model is built without *length-two loops*, which can be seen in Figure 6.9. The nodes highlighted in yellow indicate the actions on the coaching card.

It is interesting to see that the creation and reviewing behavior on the domain objects are not adjacent in the process model. Therefore, Table 6.11 selects the milestone actions (`Create`) and the review actions(`Open`) as focused nodes. **Performance Coaching** contains two domain objects: coaching card and meeting card. The adoption of a coaching card is straightforward. Managers start to use the Kaizo app by coaching other agents since `Start` is the most frequent input binding of the coaching card, either `Create` or `Open` . After creating a coaching card, managers exit the app. However, after reviewing the coaching card, managers contain follow-up actions, as seen in the output bindings of `Open/Coaching card` .

The adoption of meeting cards is different. Managers treat scheduled meetings as the last activity in the app. However, when a meeting card exists, managers directly review the information and leave the platform. The associated bindings are more relevant to reviewing performance metrics. Since the meetings from the client side are not logged, this research can only speculate that the users shift to other meeting software for offline events.

The model discovers the creators' whole-day activities. The bottom-left and -middle process is relevant to milestone actions of **Quality Assurance** and **Missions**. Other activities are related to **Scorecards**. The result shows that **Performance Coaching** is an associated pillar that shows more value when adopting other pillars simultaneously.

Action	Domain Object	Bindings	Events	Freq.
Create (Milestone)	Coaching card	Input	`(Start)`	71%
			`Complete`	16%
			Assign Team Lead	13%
		Output	`(End)`	73%
			`Visit (Inbox)`	21%
			`View Dojo`, `Visit (Inbox)`	6%
	Meeting card	Input	`Change Ticket Validation`	50%
			`Complete`	50%
		Output	`(End)`	100%
Open (Review)	Coaching card	Input	`(Start)`	100%
		Output	`Create Comment`	43%
			`Complete`	43%
			`Create Task`	7%
			`Update`	7%
	Meeting card	Input	`(Start)`	67%
			`Open Ticket (Coaching card)`	24%
			`Create Discussion Point`	10%
		Output	`(End)`	47%
			`Open (Scorecard Popup)`	41%
			`(End)`, `Open (Scorecard Popup)`	6%
			`Create Task`	6%

Bindings are extracted from Figure 6.9.
`red event` denotes the event acting on the associated domain object.
`teal event` denotes the event acting on another domain object.

Table 6.10: Bindings of Focused Nodes in Manager's Behavior Model (**Performance Coaching**)

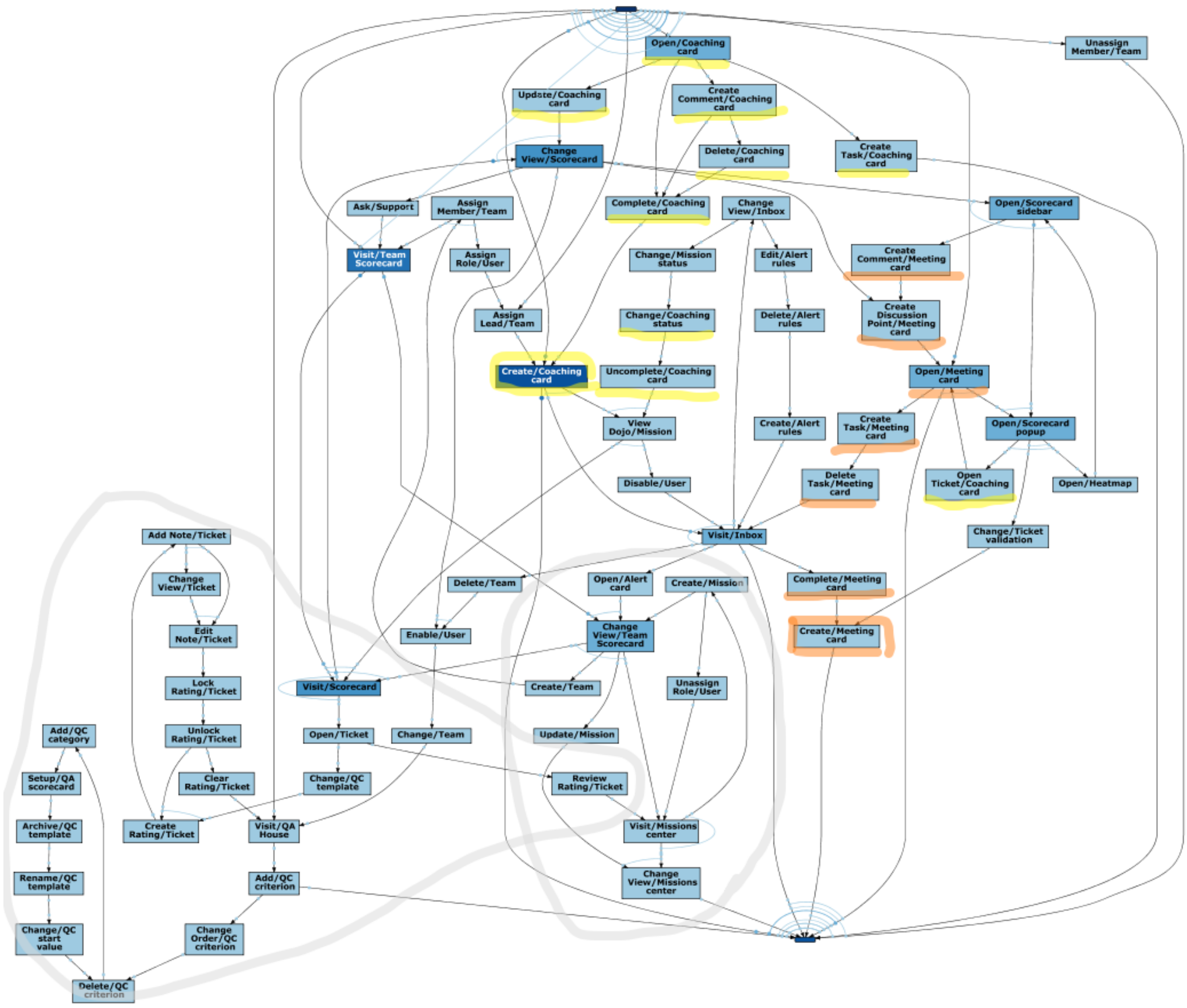


Figure 6.9: **Performance Coaching**: Manager's Behavior

6.7.2 Agent

The trace is selected whether the trace contains any action on the domain object from the corresponding agent. In other words, the day the agent with assigned cards reacts on his or her cards will be selected. To simplify the model, the users with permission higher than an agent (R_M) are filtered out. Agents reacting on test cards (`is_test`) and self-created cards (`is_self_created`) are not considered as well. The modeling is similar to the manager's model, and the removal of *length-one and length-two loops* and *all-tasks-connected* are applied based on the same reason.

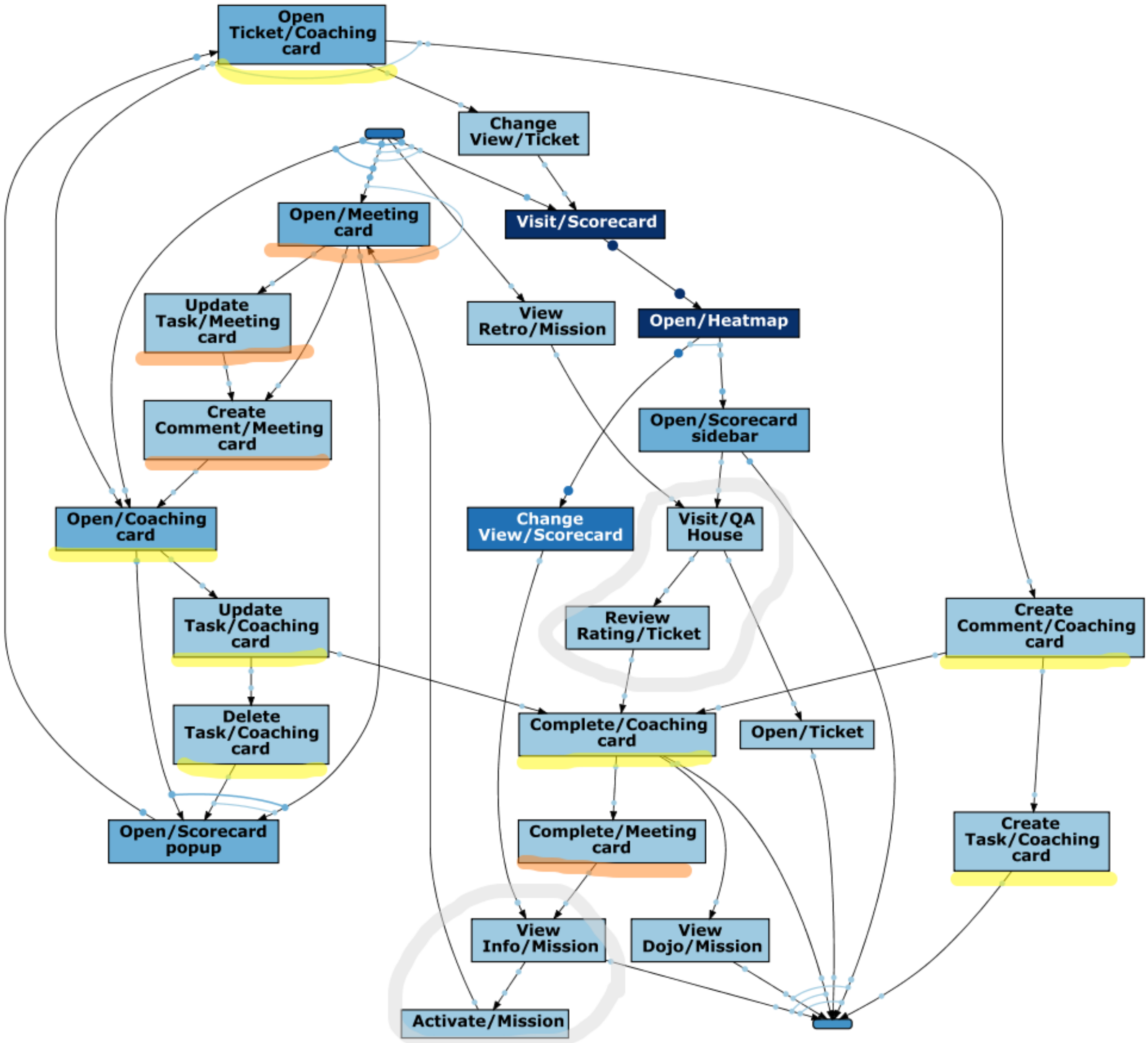


Figure 6.10: **Performance Coaching**: Agents' Behavior

According to Figure 6.10, agents show strong interest on reviewing the performance metrics in **Scorecards**. Other events that are less frequent but directly related to **Performance Coaching** are the `Open` action on the domain objects, `Coaching card` and `Meeting card`. The model also discovers the process of agents' milestone actions in the other two pillars, as discussed in the previous sections. The bottom process is associated with **Missions**, while the middle one is relevant to **Quality Assurance**.

Focused Nodes		Binding	Events	Frequency
Action	Domain Object			
	(Start)	Output	Open C.Card, Open M.Card, Visit (Scorecard)	48%
			Open C.Card, Open M.Card	40%
			Open M.Card, Visit (Scorecard)	6%
			Open M.Card, Visit (Scorecard), View Retro (Mission)	5%
Open	Coaching card	Input	(Start)	53%
			Open Ticket (C.Card)	43%
			Create Comment (M.Card)	3%
		Output	Open (Scorecard Popup)	65%
			Update Task (C.Card)	35%
	Meeting card	Input	(Start)	96%
			(Start), Activate (Mission)	4%
		Output	Open (Scorecard Popup)	82%
			Update Task (M.Card)	12%
			Create Comment (M.Card)	6%

Bindings are extracted from Figure 6.10
C.Card: Coaching card; M.Card: Meeting card
purple event: actions on domain objects; red event: actions on the corresponding domain object; teal event: actions on another domain object

Table 6.11: Bindings of Focused Events in Agent's Behavior Model (**Performance Coaching**)

Additional bindings can be seen in Table 6.11. Agents directly start with opening the cards. After that, agents show interest in reviewing the scorecards and the follow-up tasks on the domain objects.

6.7.3 Domain Object Lifecycle

Like **Quality Assurance**, the domain object in **Performance Coaching** contains comprehensive records for process discovery. Therefore, Figure 6.11 considers `coach_id` as trace, `event_action` as event, with additional information categorizing users into `Creator`, `Agent`, and `Other`. The input excludes the test and self-created cards. The model applies *all-task-connected* and remove *length-one loops*.

The creators' behavior is highlighted with a red circle, while the assigned agents' behavior is highlighted with a green circle. Other nodes are the actions triggered by other users, indicating the co-coaching behavior in some of the cases.

The process, of course, starts from `Create (Creator)`. The creator contains several follow-up processes, as seen in the upper part of the figure. After that, the assigned agent joins the process starting from opening the card (`Open (Agent)`). However, the frequency of `Open (Agent)` is not as high as `Create (Creator)`, indicating that not all the assigned agents review the cards. The following process highlighted with a green circle shows how agents interact on the card, such as `Create Comment` and `Update Task`.

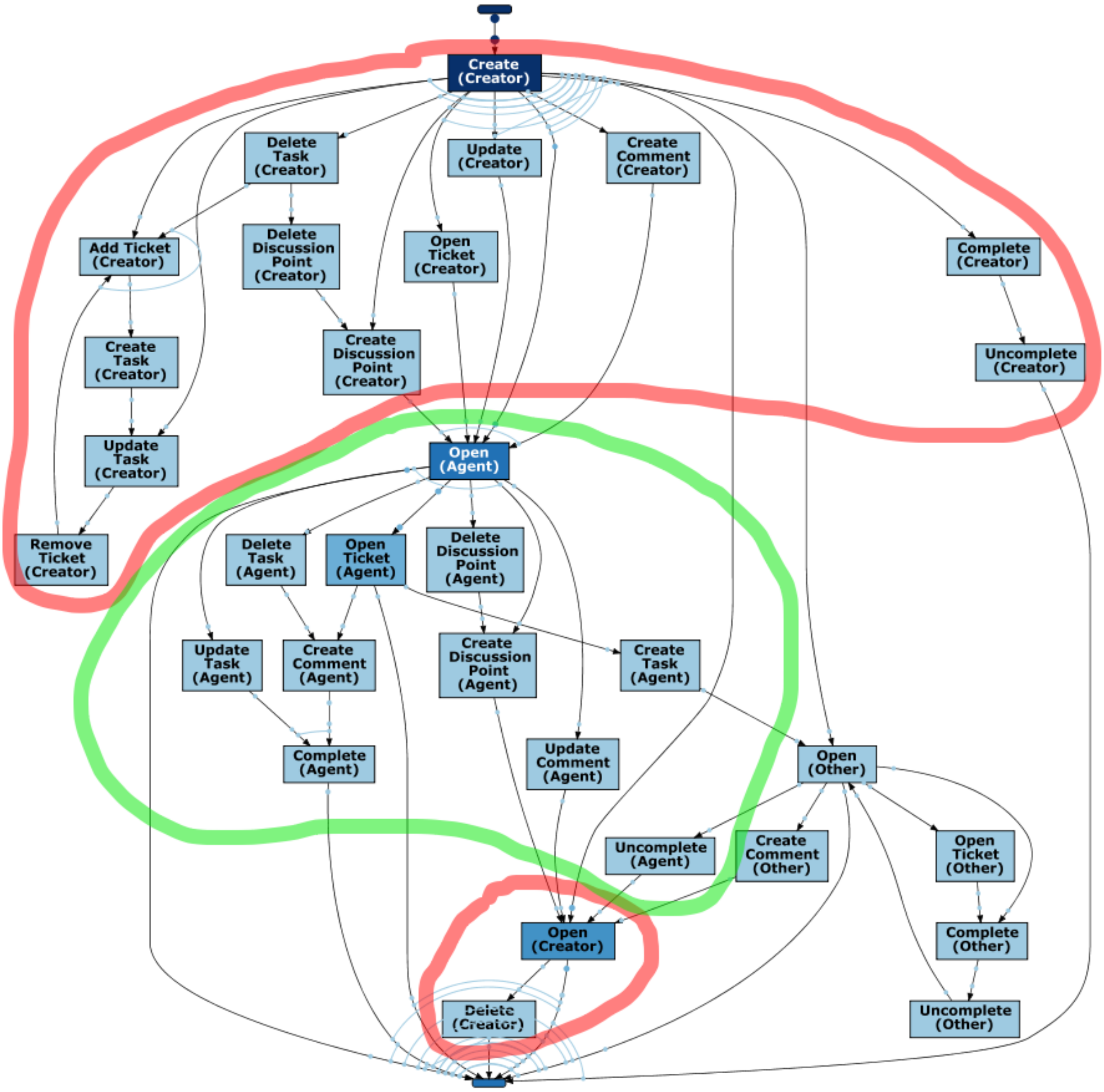


Figure 6.11: **Performance Coaching**: Domain Object Lifecycle

	Create to Agent Open	Agent Process
Mean	7 days 05:16:07	7 days 19:31:36
Median	1 days 15:47:16	0 days 00:13:38
75th Percentile	5 days 02:23:05	13 days 02:20:08

Table 6.12: Process Duration of Domain Object Life Cycle

From previous stakeholders' interviews, the retention is considered a continuous adoption from both parties, and the business activities are done weekly. Table 6.12 provides a brief summary of the process duration of domain object in **Performance Coaching**. Considering the median, the duration from creation to the agent's first opening is 1.5 days, and the duration of the agent interacting on a single card is less than a day. Considering the `Open` as agents' milestone action, most of the interaction can be done within a week.

6.8 Summary

This chapter discovers how users adopt the functions in terms of pillar, role, and domain object. Besides **Missions** with less comprehensive domain object data, other pillars contain three models from different perspectives.

Pillar	Model	Figure	Trace
Quality Assurance	Manager's behavior	6.1	`user_id` + `date`
	Agent's behavior	6.2	
	Ticket Lifecycle	6.3	`ticket_id`
Missions	Manager's behavior	6.5	`user_id` + `date`
	Agent's behavior	6.6	
Performance Coaching	Manager's behavior	6.9	`user_id` + `date`
	Agent's behavior	6.10	
	Domain Object Lifecycle	6.11	`card_id`

Table 6.13: Retention Models

Recalling Table 4.3, the agents' milestone events in **Quality Assurance** and **Performance Coaching** is not clear at the beginning. After modeling this chapter, all the missing events are clearly defined, as seen in Table 6.14.

Pillar	Domain Object	Role	
		Manager	Agent
Quality Assurance	**Ticket**	`Create Rating`	`Open`
Missions	**Mission**	`Create`	`Activate`
Performance Coaching	**Coaching/Meeting Card**	`Create`	`Open`

Table 6.14: Milestone Actions after Retention Analysis

The findings are used to cluster retention accounts. The next chapter will analyze the accounts' activation behaviors based on each cluster.

7 DISCOVER THE AHA! MOMENT

This chapter is going to answer **RQ2**, what is the aha moment look like in terms of pillar and user role? According to AARRR funnels [42], the clients moving to retention phase are obligatory to experience activation phase, which is also called aha! moment. Based on the findings from previous chapter, the accounts moving to retention phase can be identified. The selected accounts will further be treated as cases for modelling to discover the pre-retention behavior pattern in domain level.

7.1 Data Preparation

7.1.1 Data Construction

In order to capture the valid milestone actions to define retention, the identification of test activities is essential in this chapter. Therefore, `is_self_created` and `is_test` are derived.

According to Figure 5.2: **Quality Assurance** Related Data Tables and Figure 5.3: **Performance Coaching** Related Data Tables, the rating attribute of **Quality Assurance** domain object and the **Performance Coaching** domain object have their own data tables. Ratings or cards with the same `user_id` and `agent_id` is considered as self-created. That is:

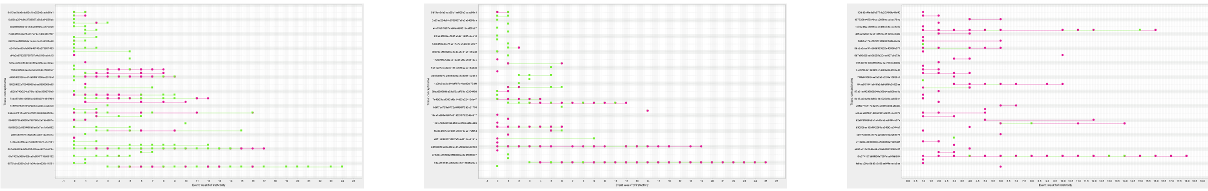
`True`	`user_id` is equal to `agent_id`
`False`	`user_id` is not equal to `agent_id`
`null`	The event is not relevant to the domain object.

The rating attribute of **Quality Assurance** domain object with `is_self_created` equals to `True` is intuitively considered as a test because it is meaningless to evaluate own service quality. On the other hand, the definition of `is_test` in **Performance Coaching** is relatively complicated which adopts nearly all the information in Data Table `cards`:

- All the content fields in cards containing string ”test” is considered a test card
- The card with a description of fewer than 10 characters is considered a test card.
- The card that is deleted within two days is considered a test card.
- The card that is completed within two days is considered a test card.

The milestone actions on self-created and test domain objects are considered invalid actions, which will not be included in the next selection phase.

7.1.2 Account Selection



(a) **Quality Assurance** (b) **Missions** (c) **Performance Coaching**

Figure 7.1: Weekly Domain Activity, by Role

After discovering the activities from both parties, the research returns to the domain level to see which account starts to adopt the features. Figure 7.1 displays the valid milestone actions from Table 6.14 in dotted charts. The logs are colored by the milestone events of both parties. The y-axis is the cases, and the x-axis is `week_to_install` to capture the weekly patterns.

The domains with two lines, indicating both parties with continuous weekly activities, can be viewed as retention. These accounts are selected for the following modeling.

7.1.3 Test Modelling

Due to the exploration of the new installed accounts, some of the characteristics of the new installed accounts are identified:

1. Only one or two users are active, and most of them are `manager` in the activation phase.
2. Users are experiencing various onboarding flows.
3. Users shift between modules and keep opening various UI components to explore the features.

The first finding provides an excellent opportunity to review the models from an account perspective. The primary activities for newly installed accounts are usually domain levels, such as setting or installation surveys. The second issue has been solved in Section 6.3: Event Classification and Data Cleaning. Only valuable actions are preserved.

It is difficult to identify whether users are seriously browsing the features or not. Since the original dataset does not record each activity's starting and ending timestamps for the calculation of completion time, it could be challenging to derive another timestamp. As a result, the following analysis would consider the browsing events and transit events as noisy events.

7.1.4 Trials

7.1.4.1. Early Trials

The first trial modeling inherits the previous account selection and applies the limitation of `week_to_install` $= 0$. `domain_id` is applied as a trace in order to see the account overview.

However, applying the heuristic miner shows the low significance of causal dependency due to the small number of records (See Figure E.1).

Furthermore, considering the lack of records, the limitation of `week_to_install` $= 0$ is abandoned because the coverage is not comprehensive. The records could start from Wednesday and cut off till Sunday, which is only three working days. Therefore, adjusted timeframe considers the entire 7 days from `install_time` to `timestamp` in order to capture the entire business week.

The second trial, applying the conclusions of the first previous trial, generated three fuzzy models (See Appendix E.2). The models seem ideal for reviewing the processes, but this research eventually focuses on which object the clients care about. Therefore, the subsequent trial considers using `event_object` an event to show the interest of the domain object.

The third trial finds out that users within the account are active at a similar time, and using the domain as the trace might mess up the timestamp and order. Therefore, the subsequent trial would consider data-awareness [41] and *long-distance dependency* [66] feature in *heuristic miner* to discover the behavior within users.

7.1.4.2. The Second Last Trial

The fourth trial reconsider the use of `event_action`. To avoid interrupt the event order within user, only the events with ”constructive” `event_action` are considered.For example, events with `event_action` such as ”Create” and ”Assign” are filtered in, while events with `event_action` such as ”Delete” and ”Edit” are filtered out. Specifically, `Change/QC template` are filtered in due to the high frequency and the highlight from stakeholders. `Ask/Help` and `Click/Payment` are filtered out due to the confusion from stakeholders.

From the previous understanding, accounts start retention less or equal to a week. Therefore, the first seven-day event logs are extracted. Considering the first week includes many settings complete by domain level, the models consider the client an entire unit. Therefore the `domain_id` is assigned as trace.

In the first week of the account, only one or two users are active within the account, and the user’s activities and sessions seldom overlap. Therefore, the model only configure `user_id` as data-awareness [**?**]. Furthermore, long-term dependency [66] is applied. Meanwhile, the dependency threshold is adjusted based on the number of traces.

Figure E.5 shows the associated interest of **Quality Assurance** related activities from the activation phase of the accounts with **Quality Assurance** retention. The accounts are purely **Quality Assurance** retention accounts. That is, the accounts do not show retention in other pillars afterward.

Figure E.6 represents the pre-retention phase of the accounts with **Missions** retention. The prerequisite of adopting **Missions** is a team, a group within the account that contribute to the same operational need. Therefore, the activities related to teams are also marked. In the model, `Create/Rating` is popped out because some of the accounts are also **Quality Assurance** retention. There is only one account with purely **Missions** retention (Figure E.9) and the **Quality Assurance** related activities are disappear. Furthermore, in Figure E.6, except the **Quality Assurance** and **Missions**-related activities, the configuration of ticket view and the care of metric/performance report are seen. `Add/Excluded tickets tag` is related to the calculation of **Scorecards**; `Add Field/TicketView` is relevant to the display of metrics; `Download/Scorecard report` represents the export of the performance metrics.

In Figure E.7 the follow-up **Performance Coaching** behavior is shown. `Create/Coaching moment` and `Add/Coaching ticket` can be considered as follow-up tasks when a coaching card is generated. Same as **Missions**, the model of the only account with pure **Performance Coaching** retention is shown in Figure E.10. In this graph, the activities indicating the interest of other pillars disappear.

Figure E.8 shows the first-7-day activities of the accounts that fail to move to the retention phase. Although the accounts contain several activities relevant to each pillar, the account did not move forward for more engagement.

7.1.4.3. The Last Trial

During the second last trial in discovering activation, the patterns of the first week regarding the account configuration are identified. However, the link between activation and retention is missing. Therefore, the final modeling considers the accounts' lifetime activities instead of the first seven days.

Figure E.11 shows the accounts with **Quality Assurance** retention without other pillar retention. `Change/Amount ticket rated value` is an important input binding of $E'_{Q,M}$, `Create Rating/Ticket` which defines the scale of manager's work on **Quality Assurance**.

Figure E.12 indicates the activities of the accounts with **Missions** retention but contains **Quality Assurance** retention as well because there is only one account that is purely **Missions** retention. The accounts show more interest in metrics that

- `Add Field/TicketView`,
- `Download/Scorecard report`, and
- `Change/Metrics Calculation`

are activities relevant to the display of ticket metrics, the reporting, and the calculation of performance metrics. These behaviors lead to the co-retention of **Quality Assurance** and **Missions** because the pure **Missions** retention account has no such interest in metrics (See Figure E.9).

Figure E.13 shows the model of **Performance Coaching**. **Performance Coaching**, similar to **Missions**, contains limited traces and fails to extract pure retention accounts. The accounts, mixed with **Quality Assurance** retention, show interest in **Quality Assurance** as well. However, compared with Figure E.11, the accounts show additional interest in the follow-up behavior in **Performance Coaching**, such as

- `Create Task/Coaching card`, and
- `Create Comment/Coaching card`.

Again, compared to Figure E.10, the nodes regarding **Quality Assurance** do not appear on the model anymore.

From stakeholders' feedback, the lifetime domain analysis, containing the activities from a different period, is confusing. The following modeling reconsiders the second last trial but elaborates on the sequence selection and parameter configuration.

7.1.5 Data Re-cleaning

In the previous trial, most of the account shows the significance of the **Quality Assurance**'s manager milestone events ($E'_{Q,M}$). However, the result might due to the sequence of the same events ($\{e'_{Q,M}, e'_{Q,M}, e'_{Q,M}, e'_{Q,M}, ...\}$) which apply on the same ticket. Such cases can be considered as duplicated tasks, but it is failed to address in heuristic miner [17].

The case can be found in the case that users rate and edit notes on the ticket. Therefore, the re-cleaning applies `ticket_id` to distinguish the duplicate tasks. According to Figure 5.2, `ticket_id` can be easily found by joined the primary key of the event log. The research only keeps the last activity within the long loop sequence. That is, consider a sequence with same events and same `ticket_id` within the user ordered by `timestamp`: $\{e_1, e_2, ..., e_n\}$. Events from e_1 to e_{n-1} are removed and e_n is kept.

7.2 Activation Behavior Modelling

7.2.1 Data Preparation and Model Configuration

7.2.1.1. Event Sequence Selection

The event sequence within an account is sliced into activation and retention. The retention criterion is the same as in Section 7.1.2 that only the week containing milestone events from two parties are considered, and the behavior pattern should repeat weekly.

Activation	Retention	
$e_1, e_2, ..., e_{i-1}$	e'_i	$e_{i+1}, e_{i+2}, ..., e_{i+n}$

Table 7.1: Example of Account Event Sequence Selection

Take Table 7.1 as an example, $\{e_1, e_2, ..., e_{i+n}\}$ is an event sequence of an account. e'_i is the first valid manager's milestone event in the retention week, with both parties' milestone events. Since the manager's milestone event in each pillar is leading actions before agents' response, the selection only considers the actions from the manager's side. Same as Section 7.1.2, the milestone actions should not be the test and self-created events. The earliest milestone action will be selected if the account contains two pillar retention. The event sequence before retention, $\{e_1, e_2, ..., e_{i-1}\}$, are selected for modelling. As a result, the start and end nodes in the model can be considered as `Installation` and `Start Retention`.

7.2.1.2. Data Filtering

As learned from the previous trials, the events that are not ”constructive” are filtered out to simplify the model. Such events include the visit events that occur too frequently, the events not triggered by users, and the follow-up actions such as `Edit`, `Disable`, and `Delete`.

7.2.1.3. Model Parameter Configuration

Inherited by the modeling experience from retention clustering, the models apply *all-tasks-connected heuristic* and remove *length-one and length-two loops*.

7.2.2 Pillar: Quality Assurance

According to Section 6.5 discussing retention, **Quality Assurance** is the most popular pillar and it’s more independent. Therefore, the **Quality Assurance** activation modeling includes only the accounts with pure **Quality Assurance** retention.

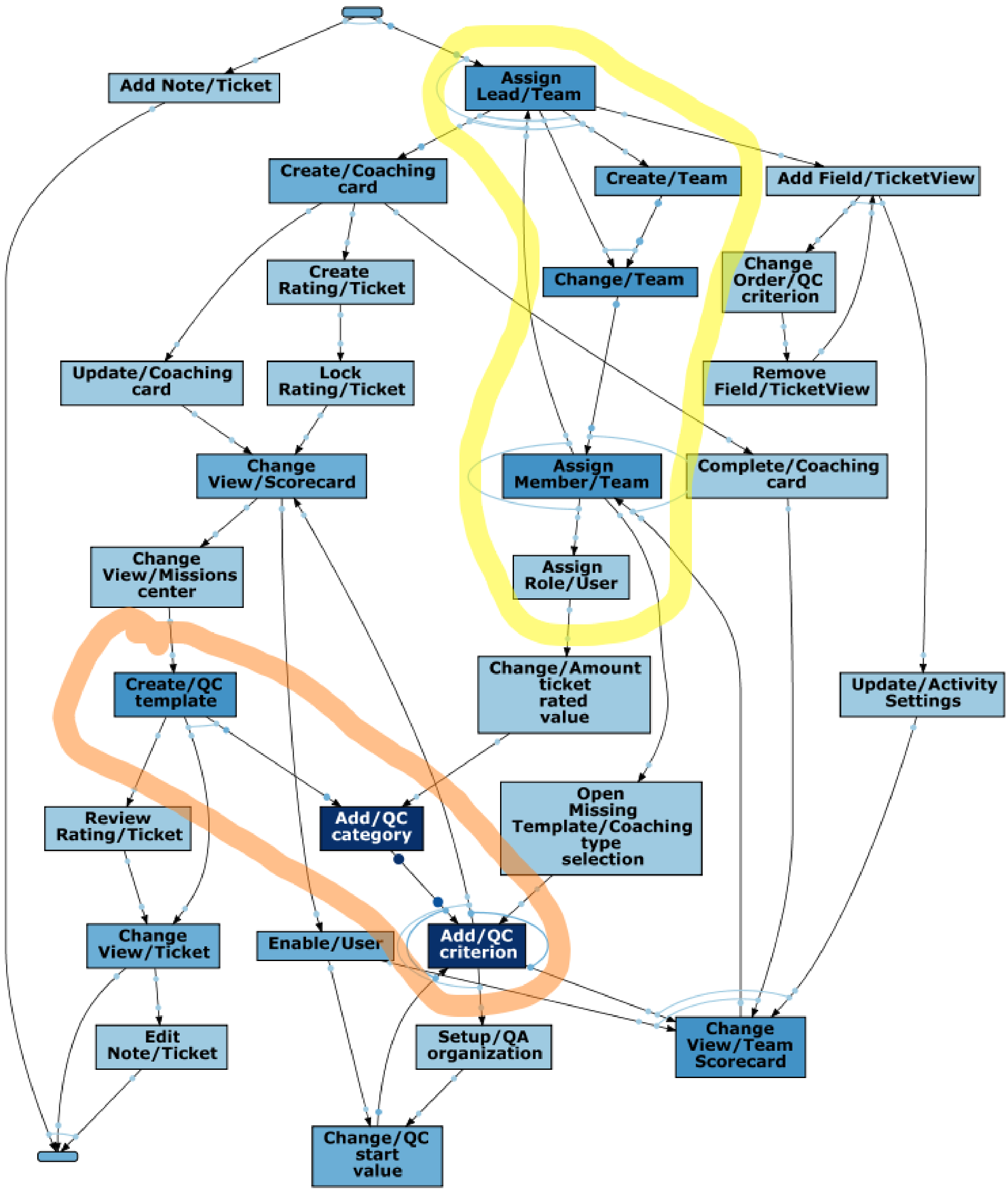


Figure 7.2: **Quality Assurance** Retention Accounts in Activation Phase

The process highlighted with orange circle in Figure 7.2 shows the accounts interest in **Quality Assurance** feature configuration. Users create their own criteria to shape ideal customer impressions based on their CRM guidelines, branding, and business goals. The process with a yellow circle indicates the nodes with the interest of team management. Since the CRM group could be big in the business, the functional team is the minimal operational unit in the CRM department.

7.2.3 Pillar: Missions

The account selection does not limit to purely **Missions** retention because it is natural that **Quality Assurance** and **Missions** retention coincide from the previous observation. Therefore, the model contains accounts with not only **Missions** retention but the retention of both **Quality Assurance** and **Missions**.

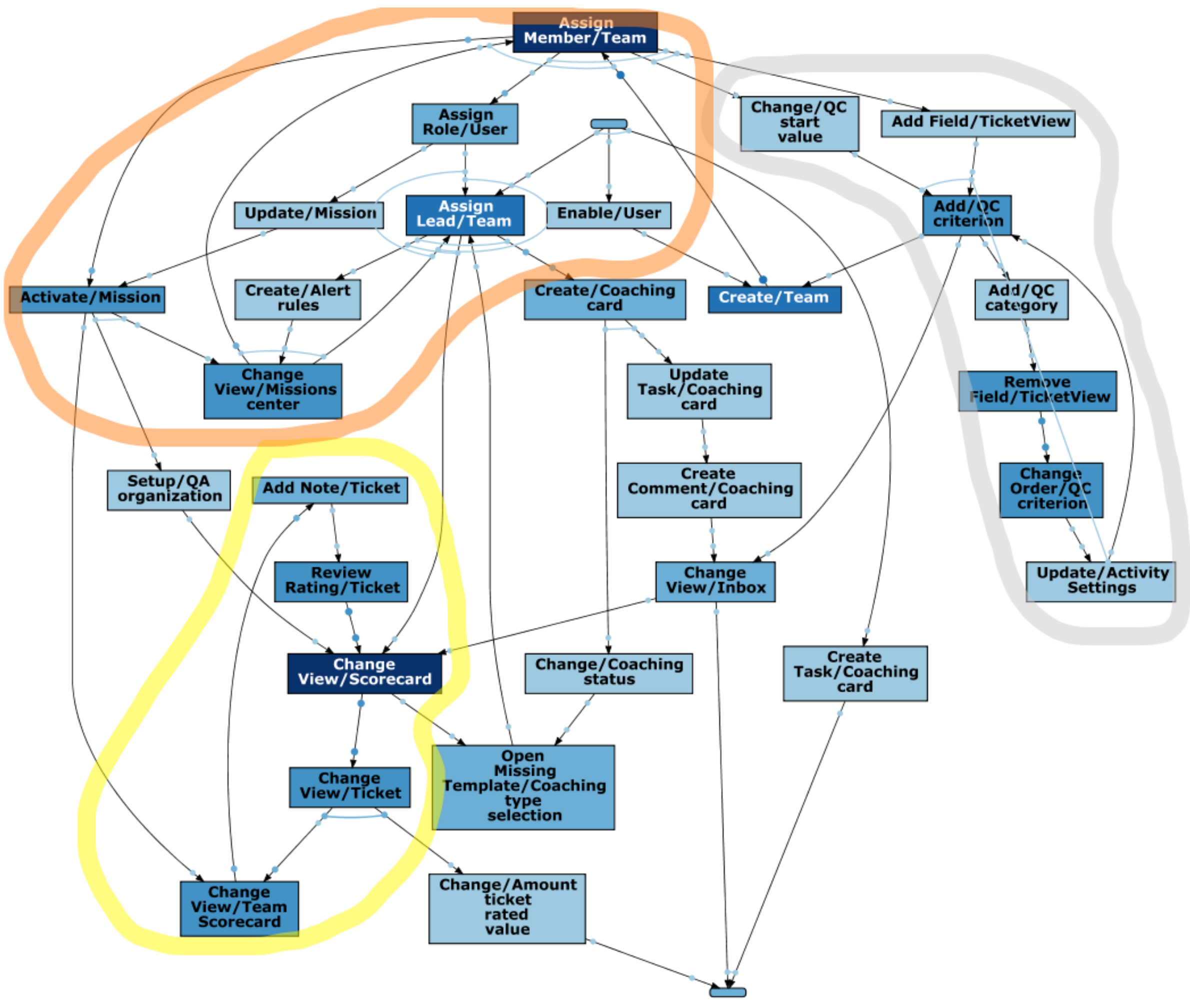


Figure 7.3: **Missions** Retention Accounts in Activation Phase

Apart from **Quality Assurance** activation as the highlighted process with grey circle in Figure 7.3, the accounts show interest in reviewing the performance metrics, for instance, the process highlighted with the yellow circle which contains the frequently occurred `Change View/Scorecard`, as well as the activities configuring the displayed metrics on the screen. The accounts also actively add new members to the team interested in manipulating the domain object of **Missions**.

7.2.4 Pillar: Performance Coaching

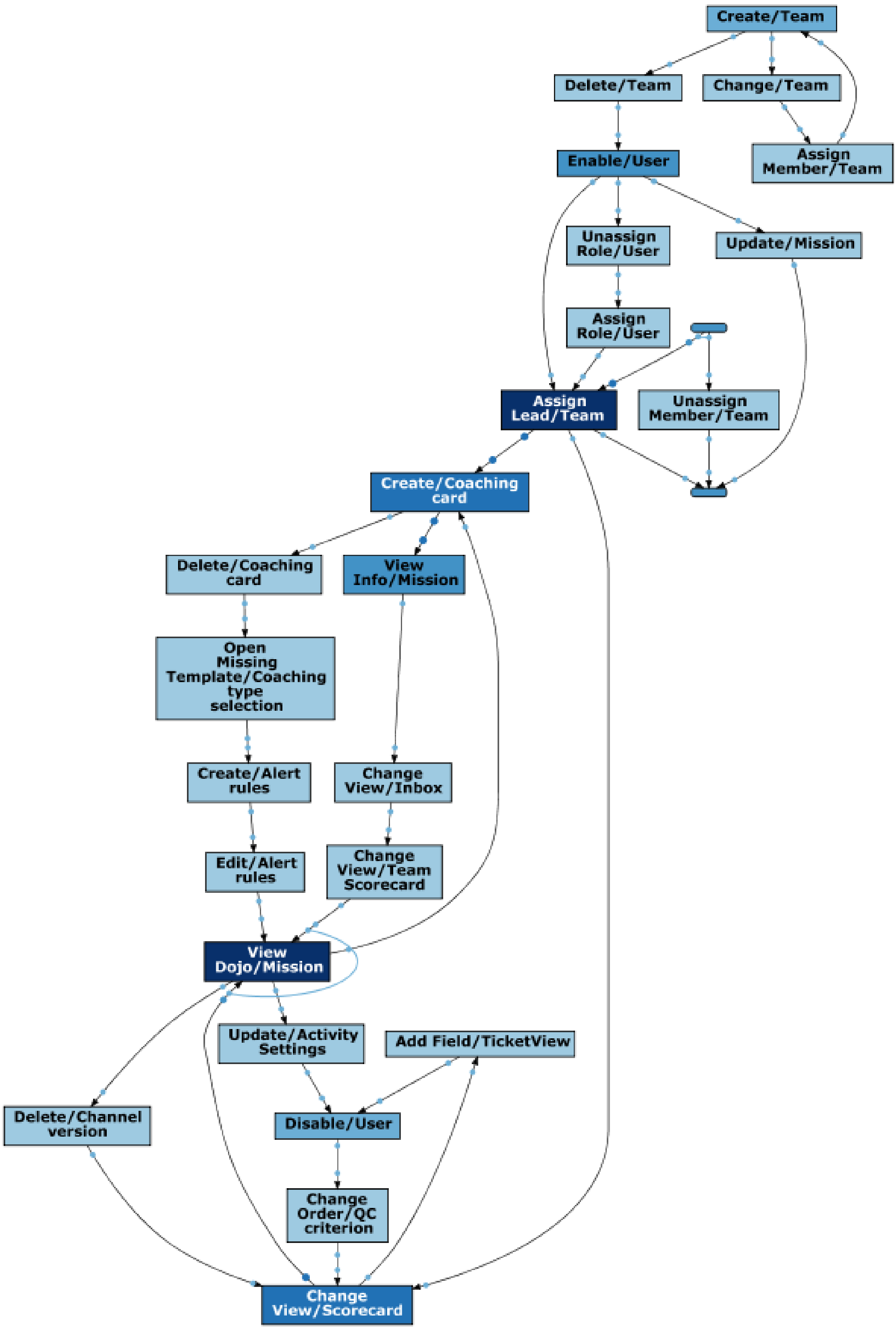


Figure 7.4: **Performance Coaching** Retention Accounts in Activation Phase

Performance Coaching modelling is special because after filtering all noisy events mentioned in Section 7.2.1.2, the model breaks into two processes like Figure 5.5. Therefore, the filtering narrows down to visit and passive events rather than follow-up events.

Apart from the events directly acting on **Performance Coaching**'s domain object, the activities in Figure 7.4 seem mixed, and it is difficult to highlight and assign the flow to specific motivation. The result might occur since the accounts also contain other pillar retention, which is reasonable as **Performance Coaching** is the less independent pillar observed in Section 6.7. On the other hand, the model shows more **Missions** activation characteristics as found in the previous subsection, indicating that **Performance Coaching** has a stronger correlation to **Missions**.

7.3 Summary and Feedback

This chapter answers **RQ2**, what is the aha moment of the clients? The analysis starts by identifying the accounts' retention week to see whether the clients have repeated behavior patterns week over week. In the meantime, the event attributes capturing invalid milestone events are defined. After several trials, the data is prepared by slicing the event sequences into activation and retention period, so the model becomes explainable that the start node indicates installation, and the end node denotes the start of retention.

The accounts with pure **Quality Assurance** retention show strong interest in configuring the settings related to creating **Quality Assurance** criterion in order to rate agents' performance. Therefore, determining the CRM quality guideline becomes the aha moment for clients using **Quality Assurance**.

Missions retention accounts focus on reviewing performance metrics. Users who are engaged in **Missions** treat **Scorecards** as an excellent partner to monitor agents' performance. The moment clients find the capacity of monitoring metrics becomes the key to **Missions** retention.

Performance Coaching is a complicated pillar that is more powerful when combined with **Quality Assurance** and **Missions**. The retention accounts show a broad interest in each pillar, and the number of traces is the least in this pillar. The stakeholders' impression of **Performance Coaching** retention clients is old accounts, which is not included in the initial dataset in this research. In other words, **Performance Coaching** retention is likely to occur in old clients. Considering the initial timeframe is defined based on the release of **Performance Coaching**, it can infer that old clients start retention from other pillars and find the value of **Performance Coaching** afterward. As a result, **Performance Coaching** can be considered as an assisted pillar, and retention is beneficial from **Quality Assurance** and **Missions**. Considering the models and the impression from stakeholders, the accounts with a broad interest in all three pillars, or the accounts with the adoption of **Missions** and **Quality Assurance** are likely to move to the retention phase.

The models are reasonable in domain knowledge. While reflecting the prerequisites and configurations, the models also validate some of the assumptions in Kaizo. On the other hand, how the model is displayed is still confusing. Although *heuristic miner* is capable of dealing with noisy scenarios, the method of presenting models, *causal net*, is less friendly for untrained stakeholders. Meanwhile, the stakeholders are also interested in discussing the interaction between pillars, but the research is failed to include such cases into modelling.

8 CONCLUSION

8.1 Result

This project aims to find the action or moment clients realize the value of the product they experience. To identify such Aha moment, the research adapts the funnel concept from AARRR model [42], assuming that all the retention users would experience an aha moment before they move to the retention phase.

During the research, domain knowledge plays an essential role. During the preparation phase before **Step 3** and **Step 3**, the important dimensions, pillar and roles are identified. After that, the partially identified milestone events assist in the first clustering in Chapter 6: Retention Clustering. Within the clustering, more insights are generated. This step specifically sheds light on agents' behavioral patterns which do not recognize before.

Meanwhile, the event classification in Section 6.3 is the key to simplifying the model by decreasing the number of nodes of 62%. Thanks to the previous trial modeling and the concepts of domain objects originating from human-computer interaction, the events are re-considered based on concepts, semantic meanings, and objects of interest. In the end, event classification successfully diminishes the influence of the user interface and the prerequisite UI binding events.

The insights further feed into the next step to re-cluster the retention accounts (Section 7.1.2). Chapter 7: Discover the Aha! Moment addresses the next research question **RQ2**. The research consider the account as the case to analyze its lifetime constructive activities such as `Create`, `Assign`, and `Setup`, instead of `Delete` and `Edit`. By further filtering out the "non-constructive" activities, the path representing the objects' interest is clearly shown.

The results can further guide the company on product development and onboarding strategies. The business can further optimize the user flow based on the process model, which could be considered the next step of process discovery.

8.2 Contribution

8.2.1 Academics

This thesis attempts to analyze customers' activation and retention by adopting process mining, involving marketing and human-computer interaction concepts.

According to Chapter 3: Background and Related Works, most process mining researchers focus on operational workflow management and employee behavior. Seldom do researchers analyze customer behaviors, which is traditionally the domain for marketing research and human-computer interaction. Besides the mining for the process within the organization, process mining papers targeting customer behaviors only focus on navigation and learning behaviors. None of the papers discuss customer lifecycle. It is new to involve a marketing funnel model to structure the process mining research.

The concept of domain objects also plays a vital role in data cleaning and event classification. It is intuitive for businesses to name the event logs by the user interface, such as `Button Click` or concatenate various components in the same column, such as `Homepage - Contact Company - Submit Contact Form`. However, for example, the behavior of contacting the company can appear not only on the homepage but on multiple pages.

The user interface interaction becomes more flexible these days due to the well-known principle derived from Nielsen [45]. One of the principles in heuristic evaluation guides that the design should provide users with the freedom to do any reasonable action. Although users benefit from the principle of experiencing more friendly interfaces, cleaning the UI logs becomes challenging. Therefore, future researchers can take advantage of the framework discussed in Section 6.3: Event Classification and Data Cleaning to reduce the data size and event nodes by diminishing the interference of dynamic interface.

8.2.2 Business

The customer value proposition is used to communicate how a company aims to provide value to customers [48]. The value can be pre-defined as a direction of developing products or services or be evolved while the company is pivoting. It is a top-down approach that the service provider defines the value first. It is difficult to tell whether the users experience the same value as proposed by the companies. The proposed value, however, could come from observations and research from users as well, but none of them originate from the event-based user process. This research fills the gap in finding the critical point of value engagement with data, a bottom-up approach to help companies reshape the value proposition.

As mentioned in Section 1.1: Aha! Moment, the onboarding or activation flow in B2B SaaS services plays a crucial role in showing clients' the product value. Only if the customers experience the value will they subscribe to the service. The identified and event-based value can precisely guide the product development team to design a better user journey and onboarding flow, leading more users to engage with the product.

The research framework can apply to future research for analyzing user behavior on web products, especially in the B2B industry. The dimensions in this research include the role and the interaction between two parties, which is suitable for the process focusing on supervision. B2B services often contain multiple solutions, and converting the solutions into pillars makes the research more organized.

The concepts in this paper can also benefit the projects analyzing customers, especially web users. Customer flows on software products are often more complicated and flexible than organizational workflow. Therefore, the idea of event classification and cleaning involving domain object concepts is suitable for cleaning noisy UI logs. Re-considering the events from a domain object perspective also helps the researchers to remove the follow-up actions in the logs, such as `Edit` and `Delete`, to simplify the model in particular cases. The capacity to reduce event nodes also shows the potential to scale up the project scope.

8.3 Discussion and Future Work

8.3.1 Pillar Interaction

The framework divides the pillar in the beginning to analyze the retention and activation but fails to discuss the interaction and relationship between pillars. On the other hand, according to the data and stakeholders' feedback, users can be active in multiple pillars, or a pillar can be a reason to use another.

Take **Performance Coaching** as an example. Companies that install after December 27, 2022, have a low percentage of being active on **Performance Coaching**, compared with the other two pillars because most of the accounts with high **Performance Coaching** retention are old clients, which are not in the research scale. That is, the clients adopt in **Missions** or **Performance Coaching**, then start using **Performance Coaching** after the function is released. The result might occur because **Performance Coaching** is such as function requiring the engagement of the other two pillars, or **Performance Coaching** is highly dependent on the usage of the other two pillars. However, this research does not further discuss this observation.

8.3.2 Process Cube

Process cube is a method to consider the process into multiple dimensions [54]. Each combination of the dimensions has its specific process model. In this research, week, user roles, and pillars can be viewed as dimensions. Therefore, the process cube concept can also be involved in future research.

Section 7.1.2 selects the retention account based on dotted chart. However, the method can be improved by conformance checking and process cube. For instance, according to the user role and the pillar, check if the account has a certain process per week to identify more robust retention.

8.3.3 Discovering Other Customer Phases

The research involves the marketing funnel concepts to organize the structure by discussing the reversed flow from retention to activation. However, vendors are also interested in other customer phases. For instance, the discussion of churn rate or the drop rate is also a popular topic in business management and customer analysis. Furthermore, due to the different role functions in the business, stakeholders only show interest in the flow they care about. For instance, the marketing team focuses on paths from acquisition to activation, and the sales team cares about which user flows can lead to revenue.

REFERENCES

[1] Wil Aalst, Arya Adriansyah, and Boudewijn Dongen. 2011. Causal Nets: A Modeling Language Tailored towards Process Discovery, Vol. 6901. 28–42. https://doi.org/10.1007/978-3-642-23217-6_3 A Causal-net (C-net) is a graph where nodes represent activities and arcs repre-sent causal dependencies.

[2] Sarabjot S. Anand, David A. Bell, and John G. Hughes. 1995. The role of domain knowledge in data mining. *Proceedings of the fourth international conference on Information and knowledge management - CIKM '95*, 37–43. https://doi.org/10.1145/221270.221321

[3] Michel Beaudouin-Lafon. 2000. Instrumental interaction: an interaction model for designing post-WIMP user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, 446–453. https://doi.org/10.1145/332040.332473

[4] Gaël Bernard and Periklis Andritsos. 2017. A process mining based model for customer journey mapping. *Forum and doctoral consortium papers presented at the 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017)* 1848, 49–56. Clarify CJM components and convert to XES format. This is a proposed modle withou empirical study. Further clustering algorithm and the prediction of the next touch point is required in this field.

[5] Alejandro Bogarín, Cristóbal Romero, Rebeca Cerezo, and Miguel Sánchez-Santillán. 2014. Clustering for improving educational process mining. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 11–15. https://doi.org/10.1145/2567574.2567604

[6] Longbing Cao. 2018. Data Science. *Comput. Surveys* 50 (5 2018), 1–42. Issue 3. https://doi.org/10.1145/3076253

[7] Wesley Carpenter. 2019. The Aha! Moment: The Science Behind Creative Insights. https://doi.org/10.5772/intechopen.84973

[8] Antonio Cerone. 2015. Process Mining as a Modelling Tool: Beyond the Domain of Business Process Management, Radu, Rumpe Bernhard Bianculli Domenico, and Calinescu (Eds.). *Software Engineering and Formal Methods*, 139–144.

[9] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. 1999. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems* 1 (2 1999), 5–32. Issue 1. https://doi.org/10.1007/BF03325089

[10] Angelo Corallo, Mariangela Lazoi, and Fabrizio Striani. 2020. Process mining and industrial applications: A systematic literature review. *Knowledge and Process Management* 27 (7 2020), 225–233. Issue 3. https://doi.org/10.1002/kpm.1630

[11] Alistair Croll and Benjamin Yoskovitz. 2013. Analytics Frameworks. , 45-54 pages.

[12] A. K. A. de Medeiros, A. J. M. M. Weijters, and W. M. P. van der Aalst. 2007. Genetic process mining: an experimental evaluation. *Data Mining and Knowledge Discovery* 14 (4 2007), 245–304. Issue 2. https://doi.org/10.1007/s10618-006-0061-7

[13] Branka Dropulić, Zoran Krupka, and Goran Vlašić. 2021. Student customer experience: A systematic literature review. *Management* 26 (12 2021), 211–228. Issue 2. https://doi.org/10.30924/mjcmi.26.2.12

[14] Sebastian Dunzer, Matthias Stierle, Martin Matzner, and Stephan Baier. 2019. Conformance checking: a state-of-the-art literature review. *Proceedings of the 11th International Conference on Subject-Oriented Business Process Management - S-BPM ONE '19*, 1–10. https://doi.org/10.1145/3329007.3329014

[15] Else, Sidorova Natalia, Brosens-Kessels Angelique, van der Aalst Wil M P van Eck Maikel L., and Markslag. 2019. Data-Driven Usability Test Scenario Creation, Kati, Lárusdóttir Marta Kristín, Palanque Philippe, Winckler Marco Bogdan Cristian, and Kuusinen (Eds.). *Human-Centered Software Engineering*, 88–108.

[16] Yves Fassin. 2009. The Stakeholder Model Refined. *Journal of Business Ethics* 84 (1 2009), 113–135. Issue 1. https://doi.org/10.1007/s10551-008-9677-4

[17] ProM forum. [n.d.]. Loops, Duplicate task and Heuristic Miner. - ProM forum. https://www.win.tue.nl/promforum/discussion/1261/loops-duplicate-task-and-heuristic-miner Online, accessed July 31, 2022.

[18] Andrew L Friedman and Samantha Miles. 2006. *Stakeholders: Theory and practice*. OUP Oxford.

[19] Asbjørn Følstad, Knut Kvale, and Ragnhild Halvorsrud. 2013. Customer journey measures-State of the art research and best practices. (2013). https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2390670

[20] Google. [n.d.]. Pricing Overview | Google Cloud. https://cloud.google.com/pricing Online, accessed August 23, 2022.

[21] Google Inc. 2022. [GA4] Funnel exploration - Analytics Help. https://support.google.com/analytics/answer/9327974 Online, accessed June 26, 2022.

[22] Google Inc. 2022. [GA4] Path exploration - Analytics Help. https://support.google.com/analytics/answer/9317498 Online, accessed June 26, 2022.

[23] Antonella Guzzo, Antonino Rullo, and Eugenio Vocaturo. 2022. Process mining applications in the healthcare domain: A comprehensive review. *WIREs Data Mining and Knowledge Discovery* 12 (3 2022). Issue 2. https://doi.org/10.1002/widm.1442

[24] Christian W. Günther and Wil M. P. van der Aalst. 2007. Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. , 328-343 pages. https://doi.org/10.1007/978-3-540-75183-0_24

[25] Husna Sarirah Husin and Shahrinaz Ismail. 2021. Process mining approach to analyze user navigation behavior of a news website. *2021 The 4th International Conference on Information Science and Systems*, 7–12. https://doi.org/10.1145/3459955.3460593

[26] Robert Johnston and Xiangyu Kong. 2011. The customer experience: a road□map for improvement. *Managing Service Quality: An International Journal* 21 (1 2011), 5–24. Issue 1. https://doi.org/10.1108/09604521111100225

[27] Rosie Jones and Kristina Lisa Klinkner. 2008. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. *Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08*, 699. https://doi.org/10.1145/1458082.1458176

[28] Kaizo Operations B.V. 2022. Kaizo. All your customer support operations in one place. https://kaizo.com/ Online, accessed June 24, 2022.

[29] Kaizo Operations B.V. 2022. Kaizo App Integration with Zendesk Support. https://www.zendesk.com/marketplace/apps/support/202996/kaizo/ Online, accessed June 29, 2022.

[30] Kaizo Operations B.V. 2022. Leverage coachable moments with Performance Coaching. https://kaizo.com/team-house/ Online, accessed June 29, 2022.

[31] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. 2008. Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems* 1 (1 2008), 137–159. Issue 1. https://doi.org/10.21307/ijssis-2017-283

[32] Fred Lemke, Moira Clark, and Hugh Wilson. 2011. Customer experience quality: an exploration in business and consumer contexts using repertory grid technique. *Journal of the Academy of Marketing Science* 39 (12 2011), 846–869. Issue 6. https://doi.org/10.1007/s11747-010-0219-0

[33] Katherine N. Lemon and Peter C. Verhoef. 2016. Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing* 80 (11 2016), 69–96. Issue 6. https://doi.org/10.1509/jm.15.0420

[34] Cong Liu, Boudewijn van Dongen, Nour Assy, and Wil M.P. van der Aalst. 2016. Component behavior discovery from software execution data. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8. https://doi.org/10.1109/SSCI.2016.7849947

[35] Cong Liu, Boudewijn van Dongen, Nour Assy, and Wil M. P van der Aalst. 2018. Component interface identification and behavioral model discovery from software execution data. *Proceedings of the 26th Conference on Program Comprehension*, 97–107. https://doi.org/10.1145/3196321.3196338

[36] Joos C A M., van Dongen Boudewijn F., van der Aalst Wil M P Verbeek H. M. W., and Buijs. 2011. XES, XESame, and ProM 6, Erik Soffer Pnina and Proper (Eds.). *Information Systems Evolution*, 60–75.

[37] Felix Mannhardt. [n.d.]. Interactive Data-aware Heuristic Miner. in ProM Lite 1.3.

[38] Felix Mannhardt, Massimiliano De Leoni, and Hajo A Reijers. 2017. Heuristic Mining Revamped: An Interactive, Data-aware, and Conformance-aware Miner. *Bpm (demos)*.

[39] Felix Mannhardt, Niek Tax, and Dennis M.M. Schunselaar. [n.d.]. Covert CSV to XES Plug-in. in ProM Lite 1.3.

[40] Fernando Martinez-Plumed, Lidia Contreras-Ochando, Cesar Ferri, Jose Hernandez-Orallo, Meelis Kull, Nicolas Lachiche, Maria Jose Ramirez-Quintana, and Peter Flach. 2021. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering* 33 (8 2021), 3048–3061. Issue 8. https://doi.org/10.1109/TKDE.2019.2962680

[41] Massimiliano, Reijers Hajo A., van der Aalst Wil M P Mannhardt Felix, and de Leoni. 2017. Data-Driven Process Discovery - Revealing Conditional Infrequent Behavior from Event Logs, Klaus Dubois Eric and Pohl (Eds.). *Advanced Information Systems Engineering*, 545–560. https://doi.org/10.1007/978-3-319-59536-8_34

[42] Dave McClure. 2022. Startup Metrics for Pirates. https://500hats.typepad.com/500blogs/2007/09/startup-metrics.html Retrieved from website ”Master of 500 Hats.”.

[43] Jorge Munoz-Gama. 2016. *Conformance Checking and Diagnosis in Process Mining*. Vol. 270. Springer International Publishing. https://doi.org/10.1007/978-3-319-49451-7

[44] Tuomas Mäkilä, Antero Järvi, Mikko Rönkkö, and Jussi Nissilä. 2010. How to Define Software-as-a-Service – An Empirical Study of Finnish SaaS Providers. , 115-124 pages. https://doi.org/10.1007/978-3-642-13633-7_10

[45] Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI ’94*, 152–158. https://doi.org/10.1145/191666.191729

[46] Jakob Nielsen. 2022. 10 Usability Heuristics for User Interface Design. https://www.nngroup.com/articles/ten-usability-heuristics/ Online, accessed July 7, 2022.

[47] Optimize Smart. 2022. How to use the Funnel Exploration Report in GA4 (Google Analytics 4) - Funnel Analysis - Optimize Smart. https://www.optimizesmart.com/how-to-use-the-funnel-analysis-report-in-google-analytics-4-ga4/ Online, accessed June 26, 2022.

[48] Adrian Payne, Pennie Frow, and Andreas Eggert. 2017. The customer value proposition: evolution, development, and application in marketing. *Journal of the Academy of Marketing Science* 45 (2017), 467–489. Issue 4. https://doi.org/10.1007/s11747-017-0523-z

[49] Sumian Peng and Qingqing Cheng. 2009. Research on Data Preprocessing Process in the Web Log Mining. *2009 First International Conference on Information Science and Engineering*, 942–945. https://doi.org/10.1109/ICISE.2009.942

[50] Vladimir A. Rubin, Alexey A. Mitsyuk, Irina A. Lomazova, and Wil M. P. van der Aalst. 2014. Process mining can be applied to software too! *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM ’14*, 1–8. https://doi.org/10.1145/2652524.2652583

[51] Christoph Schröer, Felix Kruse, and Jorge Marx Gómez. 2021. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science* 181 (2021), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

[52] Daniel Schuster, Sebastiaan J. van Zelst, and Wil M.P. van der Aalst. 2022. Utilizing domain knowledge in data-driven process discovery: A literature review. *Computers in Industry* 137 (5 2022), 103612. https://doi.org/10.1016/j.compind.2022.103612

[53] Kavita Sharma, Gulshan Shrivastava, and Vikas Kumar. 2011. Web mining: Today and tomorrow. *2011 3rd International Conference on Electronics Computer Technology*, 399–403. https://doi.org/10.1109/ICECTECH.2011.5941631

[54] Shengnan, Gorissen Pierre van der Aalst Wil M. P., and Guo. 2015. Comparative Process Mining in Education: An Approach Based on Process Cubes, Rafael, Cudre-Mauroux Philippe Ceravolo Paolo, and Accorsi (Eds.). *Data-Driven Process Discovery and Analysis*, 110–134. https://doi.org/978-3-662-46436-6_6

[55] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. 2000. Web usage mining: discovery and applications of usage patterns from Web data. *ACM SIGKDD Explorations Newsletter* 1 (1 2000), 12–23. Issue 2. https://doi.org/10.1145/846183.846188

[56] Hans Stuyck, Bart Aben, Axel Cleeremans, and Eva Van den Bussche. 2021. The Aha! moment: Is insight a different form of problem solving? *Consciousness and Cognition* 90 (4 2021), 103055. https://doi.org/10.1016/j.concog.2020.103055

[57] Michelle Taub, Allison M. Banzon, Tom Zhang, and Zhongzhou Chen. 2022. Tracking Changes in Students' Online Self-Regulated Learning Behaviors and Achievement Goals Using Trace Clustering and Process Mining. *Frontiers in Psychology* 13 (3 2022). https://doi.org/10.3389/fpsyg.2022.813514

[58] Alessandro Terragni and Marwan Hassani. 2018. Analyzing Customer Journey with Process Mining: From Discovery to Recommendations. *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, 224–229. https://doi.org/10.1109/FiCloud.2018.00040

[59] Julian Theis and Houshang Darabi. 2019. Behavioral Petri Net Mining and Automated Analysis for Human-Computer Interaction Recommendations in Multi-Application Environments. *Proceedings of the ACM on Human-Computer Interaction* 3 (6 2019), 1–16. Issue EICS. https://doi.org/10.1145/3331155

[60] Wil van der Aalst. 2012. Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems* 3 (7 2012), 1–17. Issue 2. https://doi.org/10.1145/2229156.2229157

[61] Wil van der Aalst. 2016. *Process mining : data science in action*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

[62] H.M.W. Verbeek. [n.d.]. Mine for a Fuzzy Model. in ProM Lite 1.3.

[63] Vinod, Carrera David, Khalaf Rania Poggi Nicolas, and Muthusamy. 2013. Business Process Mining from E-Commerce Web Logs, Jianmin, Weber Barbara Daniel Florian, and Wang (Eds.). *Business Process Management*, 65–80. https://doi.org/10.1007/978-3-642-40176-3_7 Structure user behaviour model intend to provide insights.

[64] Jan vom Brocke, Wil MP van der Aalst, Thomas Grisold, Waldemar Kremser, Jan Mendling, Brian Pentland, Jan Recker, Maximilian Roeglinger, Michael Rosemann, and Barbara Weber. 2021. Process Science: The Interdisciplinary Study of Continuous Change. *SSRN Electronic Journal* (2021). https://doi.org/10.2139/ssrn.3916817

[65] A.J.M.M. Weijters and J.T.S. Ribeiro. 2011. Flexible Heuristics Miner (FHM). *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 310–317. https://doi.org/10.1109/CIDM.2011.5949453

[66] A.J.M.M. Weijters, W M P Aalst van der, and A K Alves De Medeiros. 2006. *Process mining with the HeuristicsMiner algorithm*. Technische Universiteit Eindhoven.

[67] AJMM Weijters, Wil M P van Der Aalst, and A K Alves De Medeiros. 2006. Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP* 166 (2006), 1–34. Issue July 2017.

[68] Ana Wheelock, Marisa Miraldo, Anam Parand, Charles Vincent, and Nick Sevdalis. 2014. Journey to vaccination: a protocol for a multinational qualitative study. *BMJ Open* 4 (1 2014), e004279. Issue 1. https://doi.org/10.1136/bmjopen-2013-004279

[69] Rüdiger Wirth and Jochen Hipp. 2000. CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* 1, 29–39.

[70] Fitri Almira Yasmin, Faiza Allah Bukhsh, and Patricio De Alencar Silva. 2018. Process enhancement in process mining: A literature review. *CEUR workshop proceedings* 2270 (12 2018), 65–72. http://simpda2018.di.unimi.it/ 8th International Symposium on Data-driven Process Discovery and Analysis 2018, SIMPDA 2018 ; Conference date: 13-12-2018 Through 14-12-2018.

[71] Houda Zaim, Mohammed Ramdani, and Adil Haddi. 2018. Fuzzy-Based Mining Framework of Browsing Behavior to Enhance E-commerce Website Performance: Case Study from Kelkoo.com. *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, 1–6. https://doi.org/10.1145/3289402.3289528

[72] Zendesk. [n.d.]. Tickets - Zendesk Developer Docs. https://developer.zendesk.com/api-reference/ticketing/tickets/tickets/ Online, accessed July 29, 2022.

[73] Pierluigi Zerbino, Alessandro Stefanini, and Davide Aloini. 2021. Process Science in Action: A Literature Review on Process Mining in Business Management. *Technological Forecasting and Social Change* 172 (11 2021), 121021. https://doi.org/10.1016/j.techfore.2021.121021

[74] Zoom. [n.d.]. Zoom Video Communications Plans Pricing | Zoom. https://zoom.us/pricing Online, accessed August 23, 2022.

A EVENT CLASSIFICATION

Table A.1: Event Classification

Initial Event	`event_action`	`event_object`
alert_rules_modal_closed	–	–
alert_rules_modal_opened	–	–
channels_settings_channel_config_channel_icon_selected	–	–
channels_settings_channel_config_channel_name_is_changed	–	–
channels_settings_channel_config_channel_type_selected	–	–
channels_settings_channel_config_channels_order_changed	–	–
channels_settings_channel_config_form_opened	–	–
channels_settings_channel_config_tag_added	–	–
channels_settings_channel_config_tag_deleted	–	–
channels_settings_channels_versions_version_closed	–	–
channels_settings_channels_versions_version_opened	–	–
coaching_onboarding_agent_selected	–	–
coaching_onboarding_agent_selection_confirmed	–	–
coaching_onboarding_agent_selection_showed	–	–
coaching_onboarding_coaching_card_intro_fields_slide_showed	–	–
coaching_onboarding_coaching_card_intro_goals_slide_showed	–	–
coaching_onboarding_coaching_card_intro_metrics_slide_showed	–	–
coaching_onboarding_coaching_card_intro_monitor_slide_showed	–	–
coaching_onboarding_coaching_card_intro_notifications_slide_showed	–	–
coaching_onboarding_coaching_card_intro_tasks_slide_showed	–	–
coaching_onboarding_completed	–	–
coaching_onboarding_create_coaching_card_agent_selected	–	–
coaching_onboarding_create_coaching_card_agent_selection_confirmed	–	–
coaching_onboarding_create_coaching_card_completed	–	–
coaching_onboarding_create_coaching_card_started	–	–
coaching_onboarding_how_it_works_coaching_slide_showed	–	–
coaching_onboarding_how_it_works_completed	–	–
coaching_onboarding_quick_start_opened	–	–
coaching_onboarding_started	–	–
coaching_onboarding_team_leads_selection_showed	–	–
coaching_onboarding_team_leads_selection_user_selected	–	–
coaching_onboarding_team_leads_selection_user_unselected	–	–
coaching_onboarding_welcome_completed	–	–
coaching_onboarding_what_i_can_do_completed	–	–
coaching_onboarding_what_is_coaching_card_basic_fields_slide_showed	–	–
coaching_onboarding_what_is_coaching_card_challenge_slide_showed	–	–
coaching_onboarding_what_is_coaching_card_completed	–	–
coaching_onboarding_what_is_coaching_card_tasks_and_advices_slide_showed	–	–
error_authorization_error	–	–
error_contact_us_clicked	–	–
error_error_happened	–	–
features_onboarding_modal_dojo_link_clicked	–	–
features_onboarding_modal_missions_link_clicked	–	–
features_onboarding_modal_qa_link_clicked	–	–
features_onboarding_modal_scorecard_link_clicked	–	–
features_onboarding_modal_team_house_link_clicked	–	–
features_onboarding_modal_viewed	–	–
inbox_coachable_moment_creation_form_opened	–	–
inbox_missing_missions_create_team_mission_form_opened	–	–
installation_survey_form_submitted	–	–
installation_survey_rendered	–	–
kpi_dashboard_coming_soon_modal_joined_waiting_list	–	–
kpi_dashboard_coming_soon_modal_opened	–	–
kpi_dashboard_visited	–	–

kpi_dashboard_what_matters_opened	–	–
manage_ticket_settings_settings_saved	–	–
missions_center_create_agent_mission_form_open_mission_form	–	–
missions_center_create_bulk_missions_form_open_mission_form	–	–
missions_center_create_missions_form_open_mission_form	–	–
missions_center_update_agent_mission_form_open_mission_form	–	–
missions_center_update_bulk_missions_form_open_mission_form	–	–
missions_modal_closed	–	–
notifications_upgrade_plan_clicked	–	–
onboarding_assign_team_leads_opened	–	–
onboarding_assign_team_leads_user_selected	–	–
onboarding_assign_team_leads_user_unselected	–	–
onboarding_define_team_completed	–	–
onboarding_define_team_opened	–	–
onboarding_define_team_team_creation_selected	–	–
onboarding_define_team_team_joined	–	–
onboarding_define_team_team_joining_selected	–	–
onboarding_define_team_team_member_added	–	–
onboarding_define_team_team_member_removed	–	–
payment_required_modal_viewed	–	–
people_permissions_settings_people_tab_opened	–	–
people_permissions_settings_roles_tab_opened	–	–
people_permissions_settings_teams_tab_opened	–	–
performance_journey_coachable_moment_creation_form_opened	–	–
performance_journey_meeting_creation_form_opened	–	–
qa_exports_export_creation_closed	–	–
qa_exports_export_creation_custom_fields_list_updated	–	–
qa_exports_export_creation_opened	–	–
qa_exports_export_creation_range_changed	–	–
qa_exports_export_creation_scope_type_changed	–	–
qa_exports_export_creation_teams_list_updated	–	–
qa_exports_export_creation_users_list_updated	–	–
qa_exports_export_panel_closed	–	–
qa_exports_export_panel_opened	–	–
qa_exports_list_refreshed	–	–
qa_exports_visited	–	–
qa_house_assign_qa_roles_button_clicked	–	–
qa_house_coachable_moment_creation_form_opened	–	–
qa_house_edit_qa_template_button_clicked	–	–
qa_house_setup_qa_organisation_modal_showed	–	–
qa_house_welcome_modal_continue_button_clicked	–	–
qa_house_welcome_modal_showed	–	–
qa_house_welcome_screen_closed	–	–
qaonboarding_completed	–	–
qaonboarding_how_it_works_completed	–	–
qaonboarding_how_to_build_qa_scorecard_completed	–	–
qaonboarding_how_to_build_qa_scorecard_templates_article_visited	–	–
qaonboarding_quick_start_opened	–	–
qaonboarding_rate_ticket_completed	–	–
qaonboarding_setup_qa_organisation_opened	–	–
qaonboarding_setup_qa_scorecard_opened	–	–
qaonboarding_welcome_completed	–	–
qc_settings_category_group_closed	–	–
qc_settings_category_group_opened	–	–
qc_settings_create_template_button_clicked	–	–
qc_settings_mode_changed	–	–
qc_settings_submitted	–	–
qc_settings_template_archiving_canceled	–	–
qc_settings_template_reset_canceled	–	–
qc_settings_template_reset_clicked	–	–
qc_settings_templates_list_archive_clicked	–	–
qc_settings_templates_list_delete_clicked	–	–
qc_settings_templates_list_name_edit_clicked	–	–
qc_settings_templates_list_template_changed	–	–
restrictions_modal_closed	–	–
restrictions_modal_opened	–	–
restrictions_prevent_warning_checked	–	–
restrictions_prevent_warning_unchecked	–	–
restrictions_settings_opened	–	–
restrictions_warning_closed	–	–
scorecard_actionable_item_modal_closed	–	–
scorecard_actionable_item_modal_existing_items_list_closed	–	–
scorecard_actionable_item_modal_existing_items_list_opened	–	–

scorecard_actionable_item_modal_opened	–	–
scorecard_coachable_moment_creation_form_opened	–	–
scorecard_reopens_select_opened	–	–
scorecard_tutorial_started	–	–
scorecard_tutorial_step_changed	–	–
scorecard_tutorial_tutorial_finished	–	–
settings_game_engine_link_clicked	–	–
settings_manage_activity_visited	–	–
settings_manage_channels_visited	–	–
settings_manage_tickets_visited	–	–
settings_qc_settings_visited	–	–
setup_flow_agreement_continue_button_clicked	–	–
setup_flow_agreement_toggled	–	–
setup_flow_app_facts_rendered	–	–
setup_flow_privacy_link_opened	–	–
setup_flow_scorecard_tutorial_skipped	–	–
setup_flow_setup_flow_error	–	–
setup_flow_setup_flow_rendered	–	–
setup_flow_setup_privileges_error	–	–
setup_flow_terms_and_conditions_rendered	–	–
setup_flow_terms_link_opened	–	–
team_house_create_team_clicked	–	–
team_scorecard_tutorial_button_clicked	–	–
teams_setup_wizard_closed	–	–
teams_setup_wizard_opened	–	–
village_tutorial_started	–	–
welcome_modal_viewed	–	–
zendesk_top_bar_app_link_is_clicked	–	–
zendesk_top_bar_invitation_notification_is_shown	–	–
zendesk_top_bar_notification_marked_as_read	–	–
zendesk_top_bar_notifications_archived	–	–
zendesk_top_bar_opened	–	–
zendesk_top_bar_qa_notification_is_shown	–	–
activity_settings_activity_time_updated	Update	Activity Settings
activity_settings_break_time_updated	Update	Activity Settings
inbox_alert_card_opened	Open	Alert card
performance_journey_alert_card_opened	Open	Alert card
alert_rules_modal_rule_created	Create	Alert rules
alert_rules_modal_rule_deleted	Delete	Alert rules
alert_rules_modal_rule_edited	Edit	Alert rules
inbox_alert_status_changed	Change	Alert status
manage_ticket_settings_amount_ticket_rated_value_changed	Change	Amount ticket rated value
challenge_overturned	Overturn	Challenge
channels_settings_channel_config_channel_created	Create	Channel
channels_settings_channels_versions_version_deleted	Delete	Channel version
channels_settings_channel_config_channel_edited	Save	Channels Settings
inbox_coachable_moment_card_opened	Open	Coaching card
inbox_coachable_moment_comment_created	Create Comment	Coaching card
inbox_coachable_moment_deleted	Delete	Coaching card
inbox_coachable_moment_task_created	Create Task	Coaching card
inbox_coachable_moment_task_deleted	Delete Task	Coaching card
inbox_coachable_moment_task_updated	Update Task	Coaching card
inbox_coachable_moment_ticket_info_opened	Open Ticket	Coaching card
inbox_coachable_moment_ticket_removed	Remove Ticket	Coaching card
inbox_coachable_moment_updated	Update	Coaching card
performance_journey_coachable_moment_card_opened	Open	Coaching card
performance_journey_coachable_moment_coaching_card_marked_as_completed	Complete	Coaching card
performance_journey_coachable_moment_coaching_card_marked_as_uncompleted	Uncomplete	Coaching card
performance_journey_coachable_moment_comment_created	Create Comment	Coaching card
performance_journey_coachable_moment_created	Create	Coaching card
performance_journey_coachable_moment_deleted	Delete	Coaching card
performance_journey_coachable_moment_task_created	Create Task	Coaching card
performance_journey_coachable_moment_task_deleted	Delete Task	Coaching card
performance_journey_coachable_moment_task_updated	Update Task	Coaching card
performance_journey_coachable_moment_ticket_added	Add Ticket	Coaching card
performance_journey_coachable_moment_ticket_info_opened	Open Ticket	Coaching card
performance_journey_coachable_moment_ticket_removed	Remove Ticket	Coaching card
performance_journey_coachable_moment_updated	Update	Coaching card
qa_house_coachable_moment_created	Create	Coaching card
qa_house_coachable_moment_updated	Update	Coaching card
scorecard_actionable_item_modal_item_created	Create	Coaching card
scorecard_coachable_moment_card_opened	Open	Coaching card

scorecard_coachable_moment_comment_created	Create Comment	Coaching card
scorecard_coachable_moment_created	Create	Coaching card
scorecard_coachable_moment_deleted	Delete	Coaching card
scorecard_coachable_moment_ticket_info_opened	Open Ticket	Coaching card
scorecard_coachable_moment_updated	Update	Coaching card
team_house_user_overview_comment_created	Create Comment	Coaching card
team_house_user_overview_task_created	Create Task	Coaching card
team_house_user_overview_task_marked_as_completed	Complete Task	Coaching card
team_house_user_overview_task_marked_as_uncompleted	Uncomplete Task	Coaching card
inbox_coachable_moment_status_changed	Change	Coaching status
performance_journey_coachable_moment_type_selection_missing_template_opened	Open Missing Template	Coaching type selection
performance_journey_coachable_moment_type_selection_opened	Open	Coaching type selection
connection_settings_reauthorize_button_clicked	Reauthorize	Connection
scorecard_heatmap_opened	Open	Heatmap
inbox_filters_cleared	Change View	Inbox
inbox_filters_filtered_by_alert_status	Change View	Inbox
inbox_filters_filtered_by_inbox_status	Change View	Inbox
inbox_filters_filtered_by_owner	Change View	Inbox
inbox_filters_filtered_by_subject	Change View	Inbox
inbox_inbox_opened	Visit	Inbox
inbox_pagination_page_changed	Change View	Inbox
inbox_pagination_size_changed	Change View	Inbox
inbox_inbox_item_opened	Open	Inbox item
inbox_meeting_card_opened	Open	Meeting card
inbox_meeting_comment_created	Create Comment	Meeting card
inbox_meeting_deleted	Delete	Meeting card
inbox_meeting_discussion_point_created	Create Discussion Point	Meeting card
inbox_meeting_task_created	Create Task	Meeting card
inbox_meeting_task_deleted	Delete Task	Meeting card
performance_journey_meeting_card_opened	Open	Meeting card
performance_journey_meeting_comment_created	Create Comment	Meeting card
performance_journey_meeting_comment_updated	Update Comment	Meeting card
performance_journey_meeting_created	Create	Meeting card
performance_journey_meeting_discussion_point_created	Create Discussion Point	Meeting card
performance_journey_meeting_discussion_point_deleted	Delete Discussion Point	Meeting card
performance_journey_meeting_meeting_marked_as_completed	Complete	Meeting card
performance_journey_meeting_meeting_marked_as_uncompleted	Uncomplete	Meeting card
performance_journey_meeting_task_created	Create Task	Meeting card
performance_journey_meeting_task_deleted	Delete Task	Meeting card
performance_journey_meeting_task_updated	Update Task	Meeting card
performance_journey_meeting_updated	Update	Meeting card
inbox_meeting_status_changed	Change	Meeting status
manage_ticket_settings_excluded_tickets_tag_added	Change	Metrics Calculation
manage_ticket_settings_excluded_tickets_tag_deleted	Change	Metrics Calculation
dojo_room_visited	View Dojo	Mission
inbox_missing_missions_missions_created	Create	Mission
missions_center_create_missions_form_missions_created	Create	Mission
missions_center_update_mission_form_agent_mission_created	Create	Mission
missions_center_update_mission_form_agent_mission_deleted	Delete	Mission
missions_center_update_mission_form_agent_mission_updated	Update	Mission
missions_center_update_mission_form_mission_updated	Update	Mission
missions_center_week_retro_modal_showed	View Retro	Mission
missions_modal_congratulations_opened	View Retro	Mission
missions_modal_current_week_info_opened	View Info	Mission
missions_modal_missions_activated	Activate	Mission
inbox_mission_reminder_status_changed	Change	Mission status
mission_center_agent_changed	Change View	Missions center
mission_center_team_changed	Change View	Missions center
mission_center_visited	Visit	Missions center
mission_center_week_changed	Change View	Missions center
nrr_challenged	Challenge	NRR
navigation_dojo_classic_clicked	Click	Navigation
navigation_dojo_master_clicked	Click	Navigation
navigation_qa_clicked	Click	Navigation
navigation_settings_clicked	Click	Navigation
navigation_team_scorecard_clicked	Click	Navigation
payment_required_modal_book_a_call_clicked	Contact	Payment
payment_required_modal_pay_now_clicked	Click	Payment

people_settings_upgrade_billing_plan_clicked	Click	Payment
people_permissions_settings_roles_permission_added	Add	Permission
people_permissions_settings_roles_permission_removed	Remove	Permission
qa_house_visited	Visit	QA House
qa_exports_export_archived	Archive	QA export
qa_exports_export_creation_export_created	Create	QA export
qa_house_setup_qa_organisation_modal_role_toggled	Setup	QA organization
qaonboarding_setup_qa_organisation_completed	Setup	QA organization
qa_exports_report_downloaded	Download	QA report
qaonboarding_setup_qa_scorecard_completed	Setup	QA scorecard
qc_settings_category_added	Add	QC category
qc_settings_category_deleted	Delete	QC category
manage_ticket_settings_ticket_view_fields_order_changed	Change Order	QC criterion
qc_settings_categories_order_changed	Change Order	QC criterion
qc_settings_criterion_added	Add	QC criterion
qc_settings_criterion_deleted	Delete	QC criterion
qc_settings_criterions_order_changed	Change Order	QC criterion
qc_settings_starting_value_changed	Change	QC start value
qc_settings_template_archived	Archive	QC template
qc_settings_template_created	Create	QC template
qc_settings_template_deletion_confirmed	Delete	QC template
qc_settings_template_reset_confirmed	Reset	QC template
qc_settings_template_restored	Restore	QC template
qc_settings_templates_list_save_clicked	Rename	QC template
scorecard_qc_template_changed	Change	QC template
scorecard_agent_changed	Change View	Scorecard
scorecard_category_panel_closed	Change View	Scorecard
scorecard_category_panel_opened	Change View	Scorecard
scorecard_channels_filter_all_channels_selected	Change View	Scorecard
scorecard_channels_filter_all_channels_unselected	Change View	Scorecard
scorecard_channels_filter_channel_selected	Change View	Scorecard
scorecard_channels_filter_channel_unselected	Change View	Scorecard
scorecard_channels_filter_edit_channels_channel_selected	Change View	Scorecard
scorecard_channels_filter_edit_channels_channel_unselected	Change View	Scorecard
scorecard_channels_filter_edit_channels_dropdown_closed	Change View	Scorecard
scorecard_channels_filter_edit_channels_dropdown_opened	Change View	Scorecard
scorecard_channels_filter_edit_channels_selected_all	Change View	Scorecard
scorecard_channels_filter_edit_channels_unselected_all	Change View	Scorecard
scorecard_heatmap_timeframe_changed	Change View	Scorecard
scorecard_heatmap_touchpoints_view_toggled	Change View	Scorecard
scorecard_month_changed	Change View	Scorecard
scorecard_sidebar_scale_changed	Change View	Scorecard
scorecard_visited	Visit	Scorecard
scorecard_popup_opened	Open	Scorecard popup
scorecard_report_is_downloaded	Download	Scorecard report
scorecard_sidebar	Open	Scorecard sidebar
scorecard_support_opened	Ask	Support
coaching_onboarding_coaching_card_intro_opened	Assign Lead	Team
onboarding_assign_team_leads_completed	Assign Lead	Team
onboarding_define_team_config_confirmed	Assign Member	Team
onboarding_define_team_team_created	Create	Team
people_permissions_settings_teams_member_assigned	Assign Member	Team
people_permissions_settings_teams_member_unassigned	Unassign Member	Team
people_permissions_settings_teams_team_created	Create	Team
people_permissions_settings_teams_team_deleted	Delete	Team
people_permissions_settings_teams_team_info_changed	Change	Team
people_permissions_settings_teams_team_lead_assigned	Assign Lead	Team
people_permissions_settings_teams_team_lead_unassigned	Unassign	Team
teams_setup_wizard_config_confirmed	Assign Member	Team
teams_setup_wizard_team_created	Create	Team
teams_setup_wizard_team_member_added	Assign Member	Team
teams_setup_wizard_team_member_removed	Unassign Member	Team
team_house_teams_overview_visited	Visit	Team House (Team)
team_house_user_overview_visited	Visit	Team House (User)
team_scorecard_filters_updated	Change View	Team Scorecard
team_scorecard_metric_changed	Change View	Team Scorecard
team_scorecard_metric_data_sorted	Change View	Team Scorecard
team_scorecard_pagination_page_changed	Change View	Team Scorecard
team_scorecard_pagination_size_changed	Change View	Team Scorecard
team_scorecard_visited	Visit	Team Scorecard
scorecard_pop_up_displayed_ticket_changed	Open	Ticket
scorecard_popup_note_added	Add Note	Ticket

scorecard_popup_note_deleted	Delete Note	Ticket
scorecard_popup_note_edited	Edit Note	Ticket
scorecard_popup_opened	Open	Ticket
scorecard_popup_rating_cleared	Clear Rating	Ticket
scorecard_popup_rating_locked	Lock Rating	Ticket
scorecard_popup_rating_unlocked	Unlock Rating	Ticket
scorecard_popup_tab_changed	Change View	Ticket
scorecard_popup_ticket_swapped	Open	Ticket
scorecard_qc_rated_criterion	Create Rating	Ticket
scorecard_sidebar	Review Rating	Ticket
scorecard_popup_ticket_validation_changed	Change	Ticket validation
manage_ticket_settings_ticket_view_field_added	Add Field	TicketView
manage_ticket_settings_ticket_view_field_removed	Remove Field	TicketView
people_permissions_settings_dojo_classic_disabled	Disable	User
people_permissions_settings_dojo_classic_enabled	Enable	User
people_permissions_settings_roles_role_assigned_to_user	Assign Role	User
people_permissions_settings_roles_role_created	Create Role	User
people_permissions_settings_roles_role_deleted	Delete Role	User
people_permissions_settings_roles_role_info_changed	Change Role	User
people_permissions_settings_roles_role_unassigned_from_user	Unassign Role	User
team_house_teams_overview_user_agenda_opened	Open	User agenda
master_village_rendered	Visit	Village

B TEST FUZZY MODELS FOR DATA CLEANING

Figure B.1: Test Model before Data Cleaning (User as Trace)

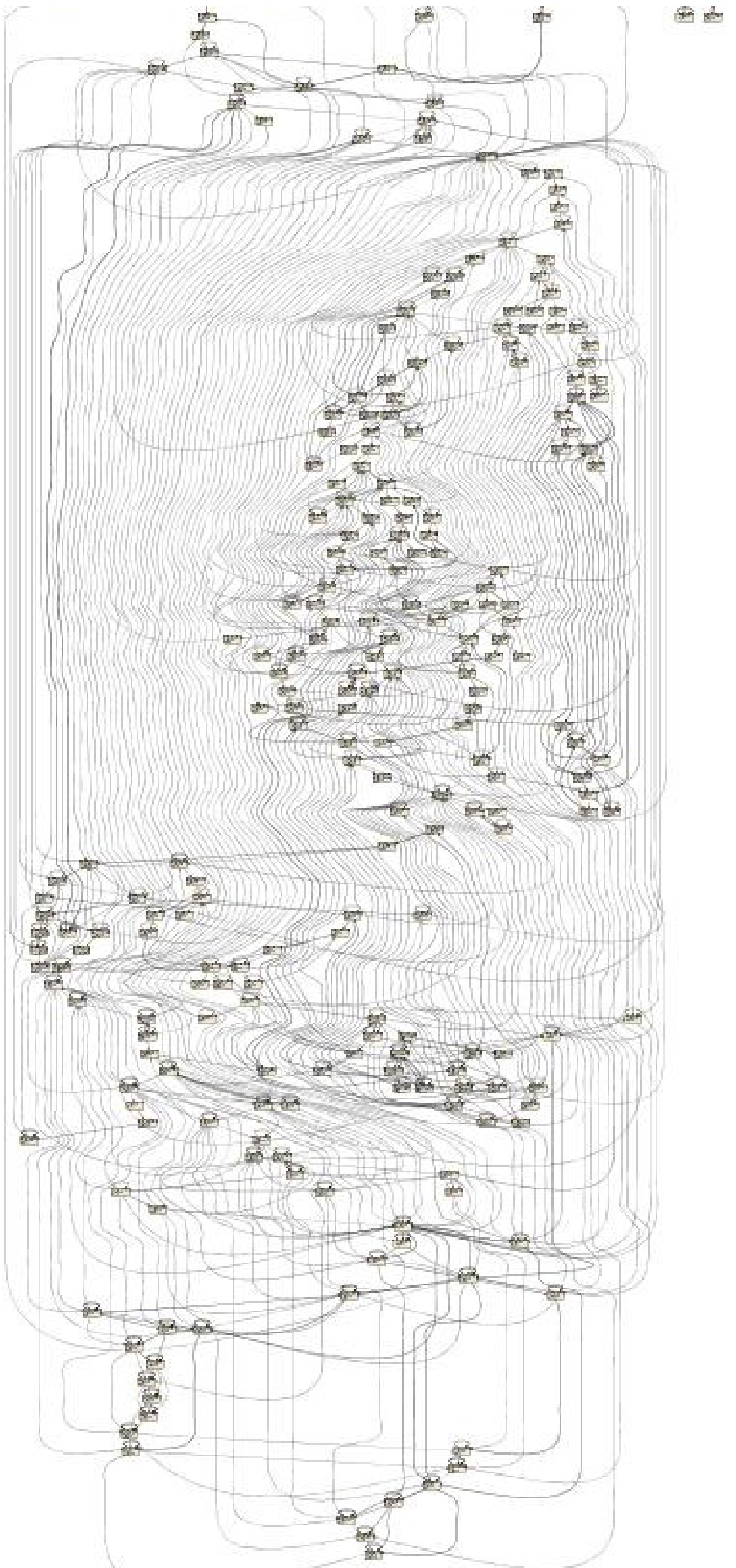


Figure B.2: Test Model before Data Cleaning (Session as Trace)

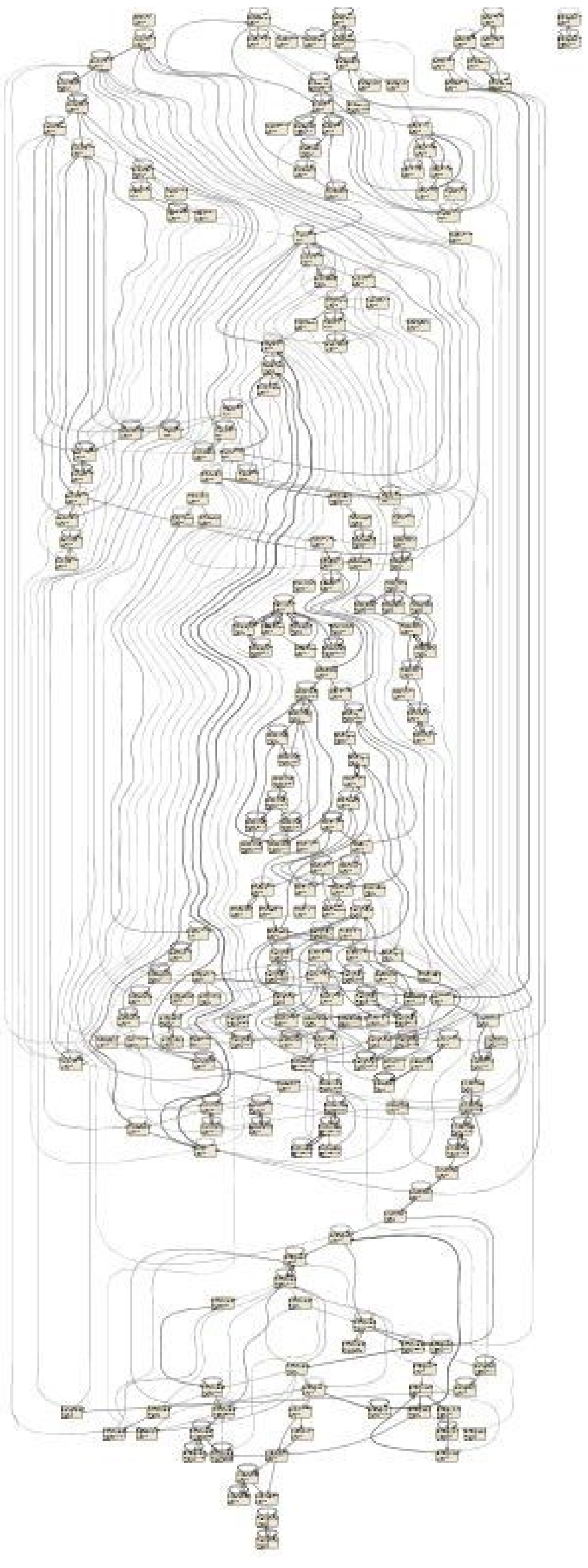


Figure B.3: Test Model after Removing Invalid Sessions

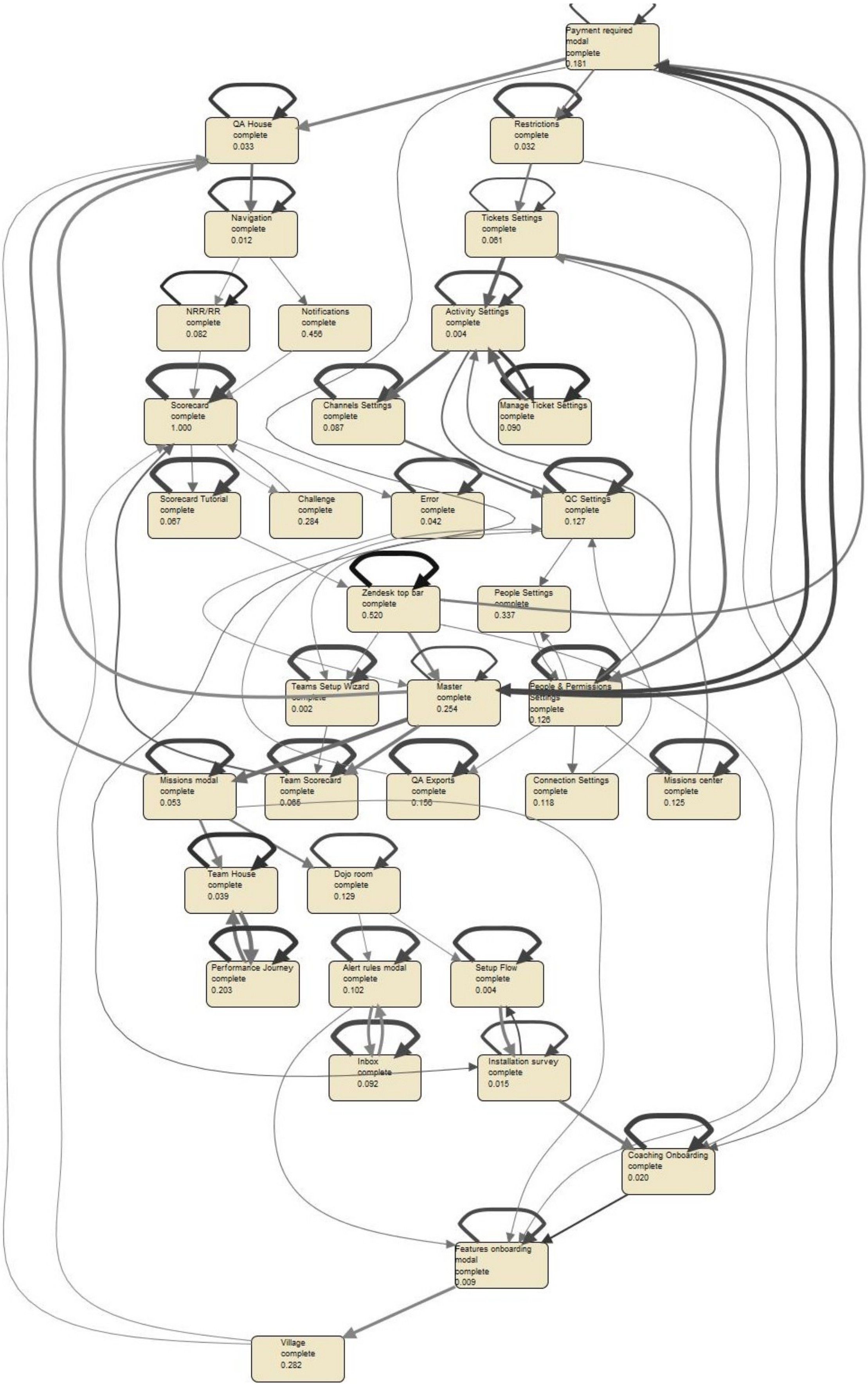


Figure B.4: Test Model Applying `eventCategory` as Event

C RETENTION PATTERN IDENTIFICATION

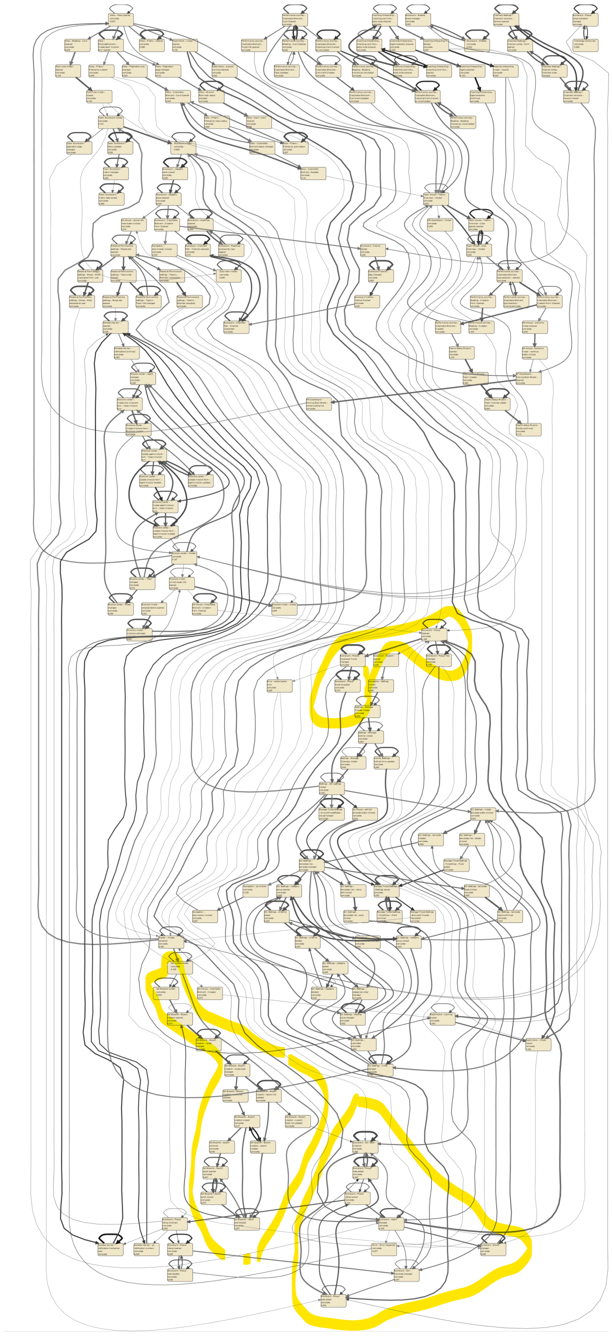


Figure C.1: Sessions contains Managers' Milestone Events in **Quality Assurance**

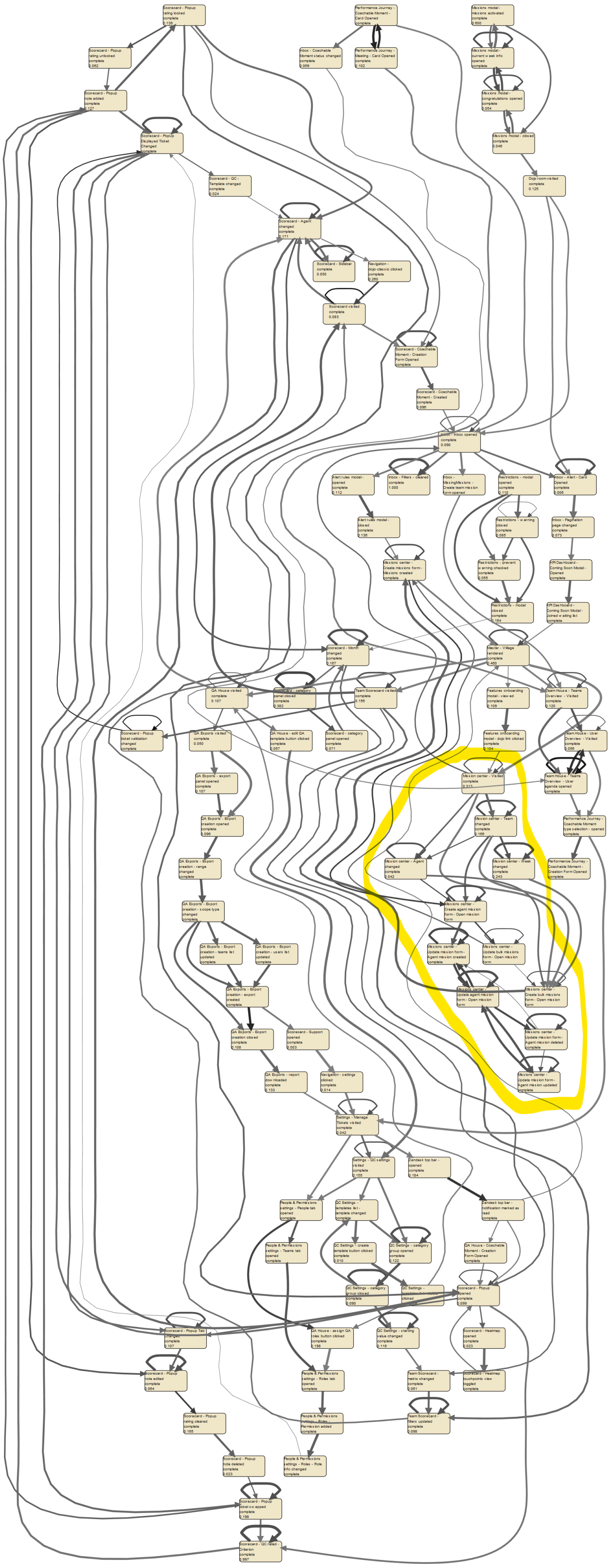


Figure C.2: Sessions contains Managers' Milestone Events in **Missions**

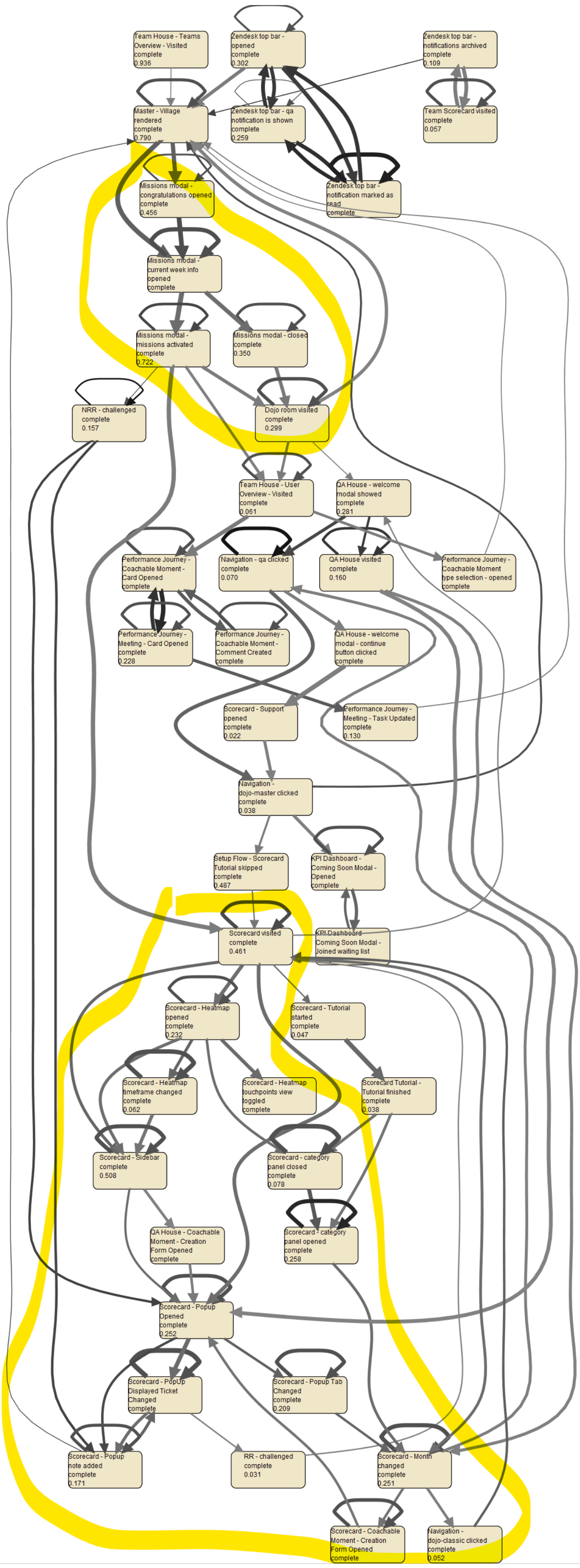


Figure C.3: Sessions contains Agents' Milestone Events in **Missions**

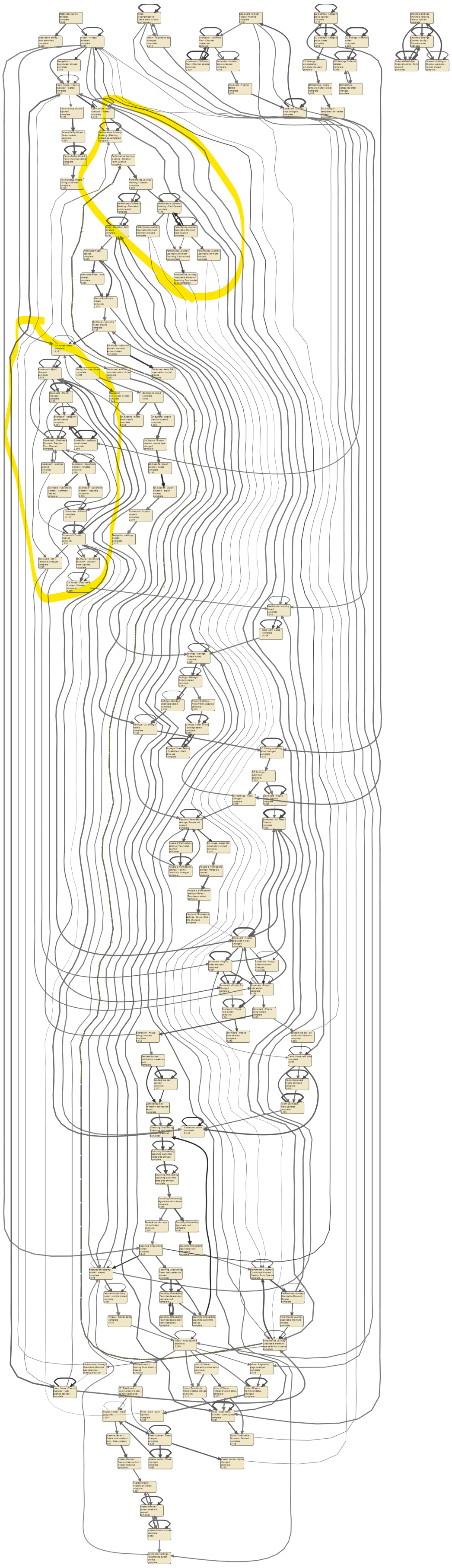
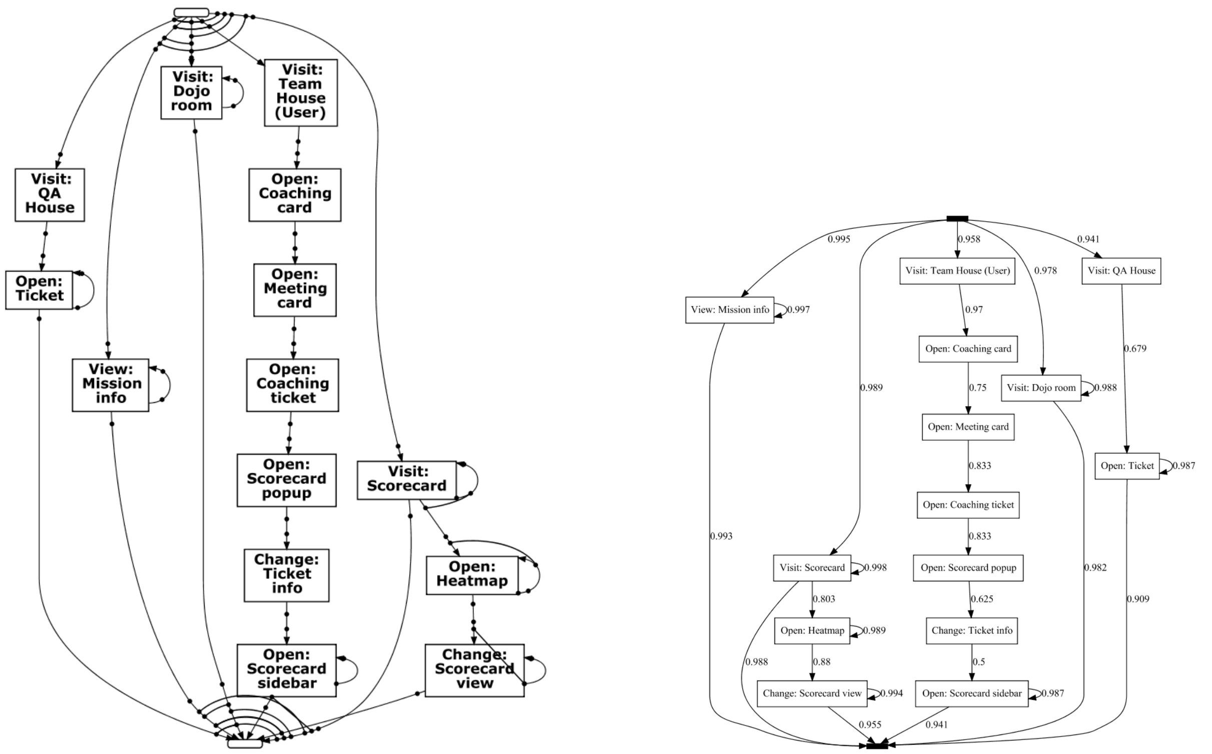


Figure C.4: Sessions contains Managers' Milestone Events in **Performance Coaching**

D AGENTS WITH MISSIONS

Corresponding to Subsection **??**, Figure D.1 describes the process of agents with assigned missions and the date after activation.



(a) C-net (b) Dependency Graph

Figure D.1: The Process Model of Agents who Activated the Missions

E MODELS IN DISCOVERING AHA MOMENT

E.1 First Trial

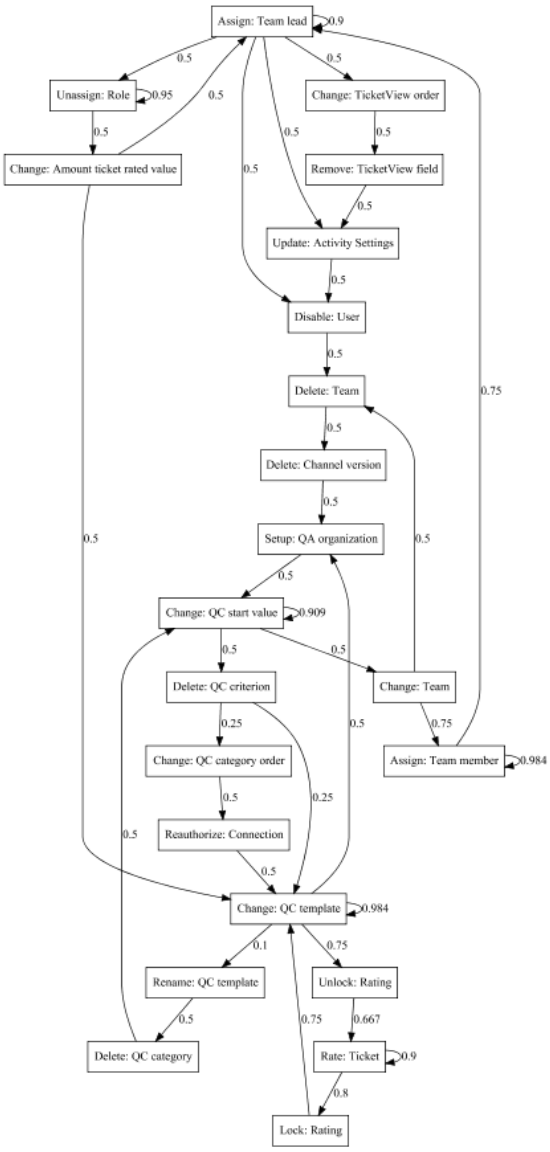


Figure E.1: The Example Dependency Graph of the First Trial Modelling

E.2 Second Trial

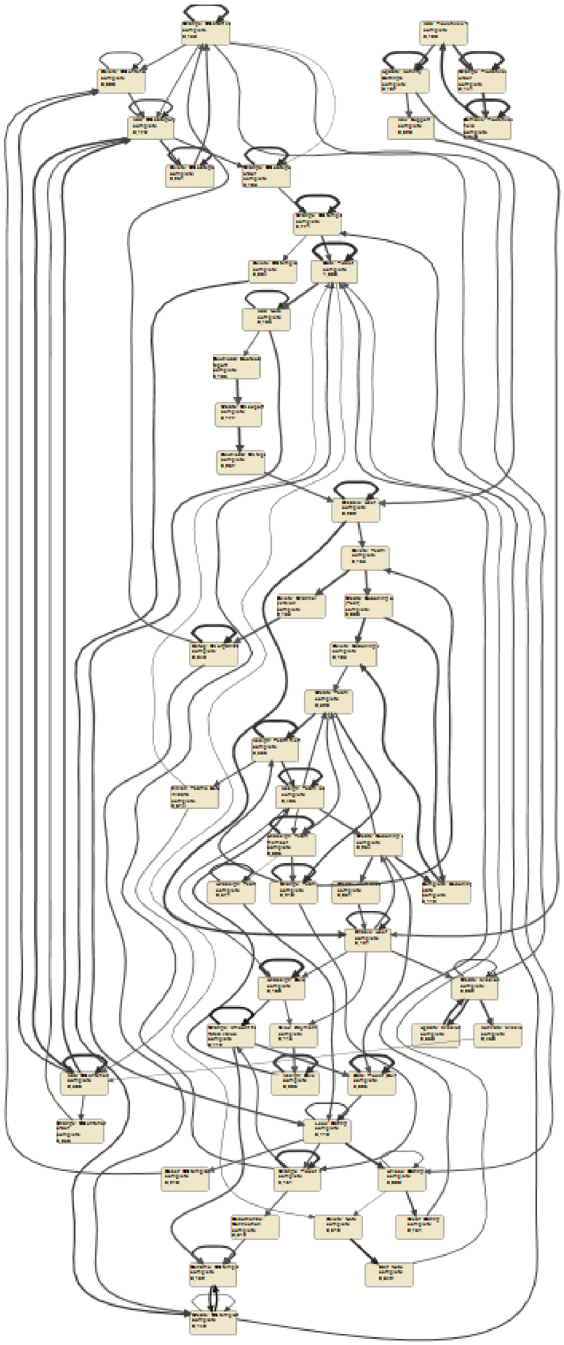


Figure E.2: The Activation Phase of the Account with High **Quality Assurance** Retention

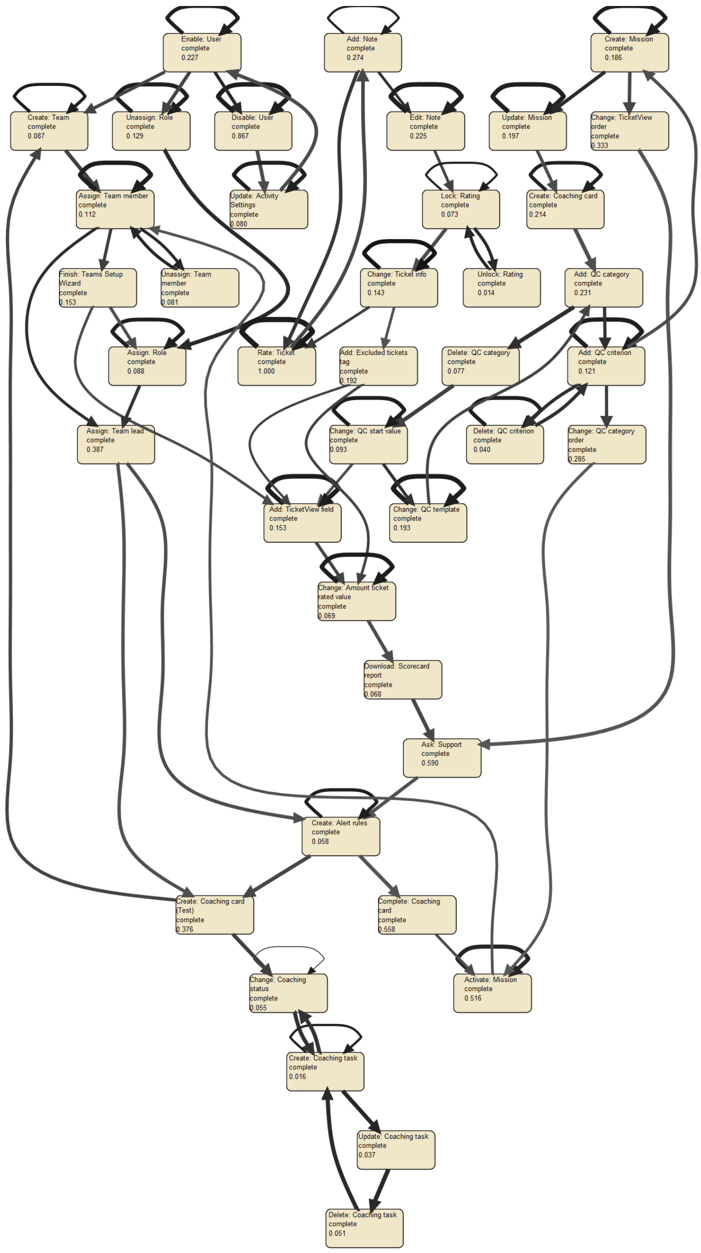


Figure E.3: The Activation Phase of the Account with High **Missions** Retention

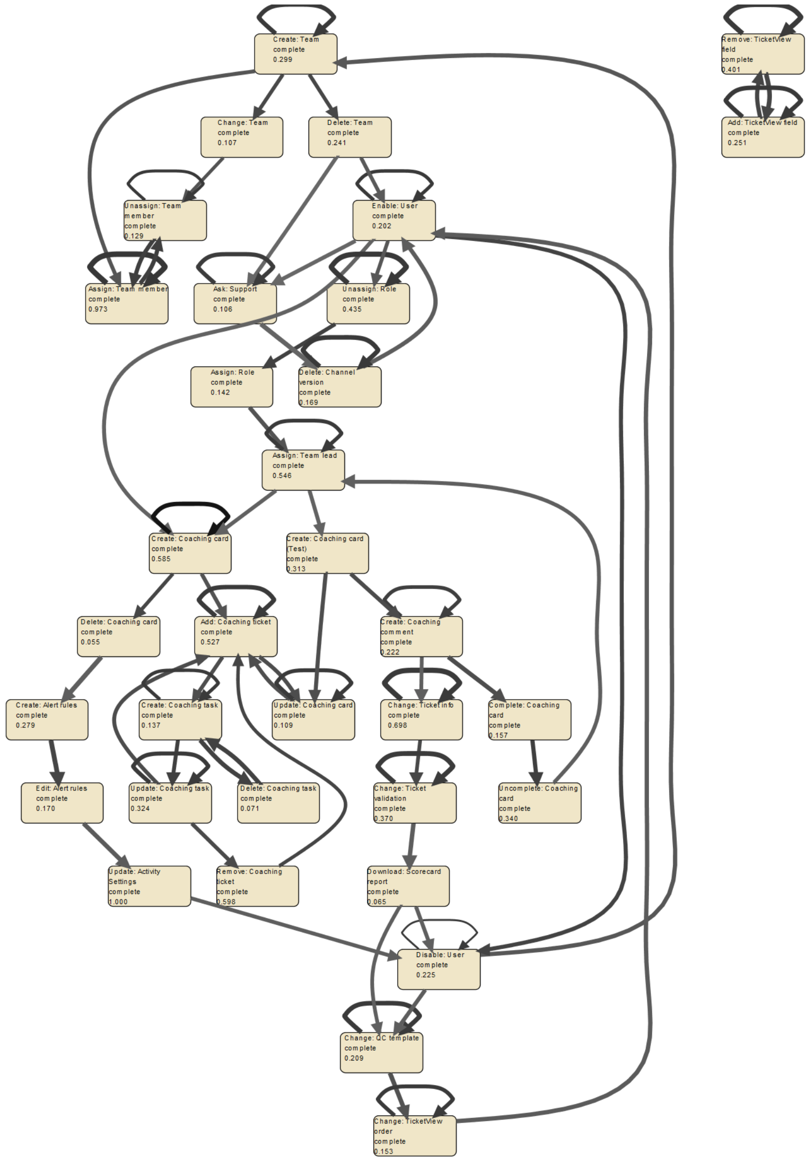


Figure E.4: The Activation Phase of the Account with High **Performance Coaching** Retention

E.3 Fourth Trial

This trial consider the the first-week activities of the accounts. The nodes with yellow marks in pillar models are the pre-identified activities that are directly-related to pillars.

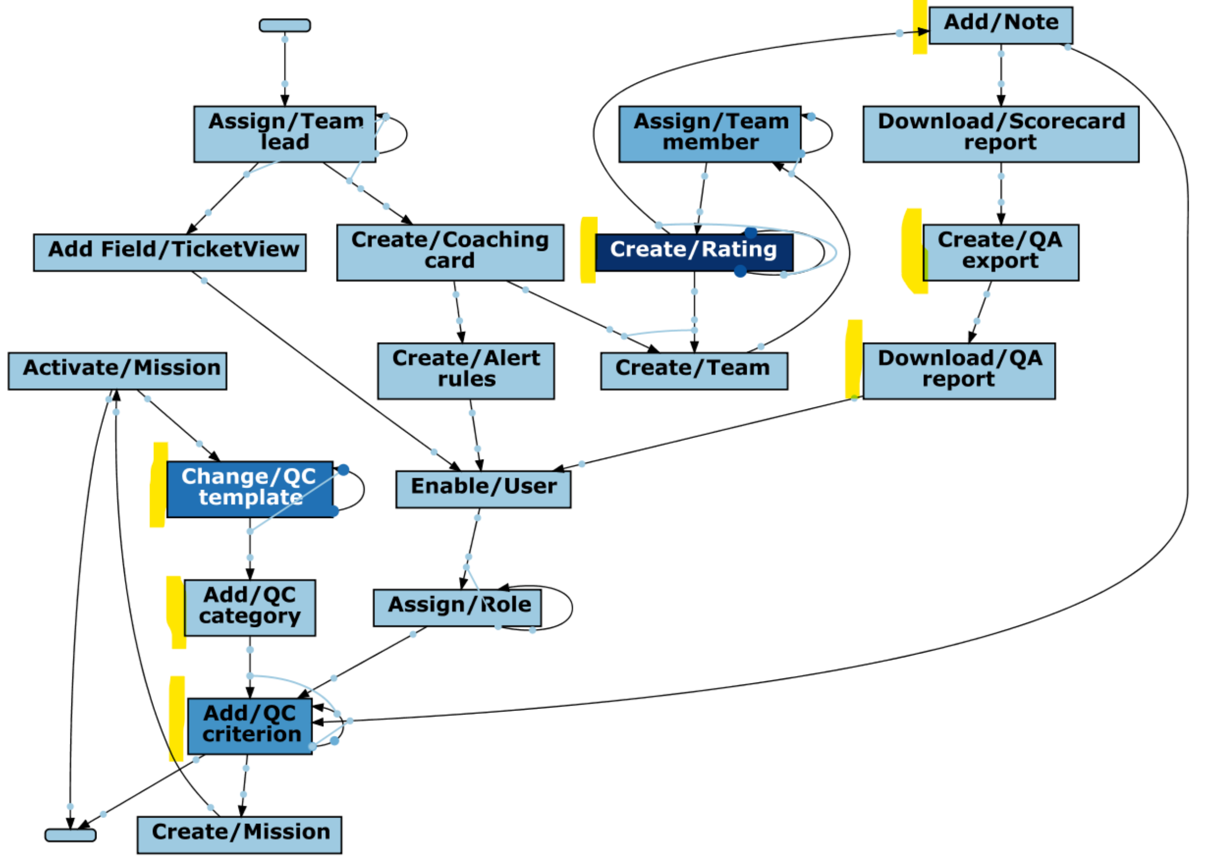


Figure E.5: Activation Phase of Accounts with **Quality Assurance** Retention

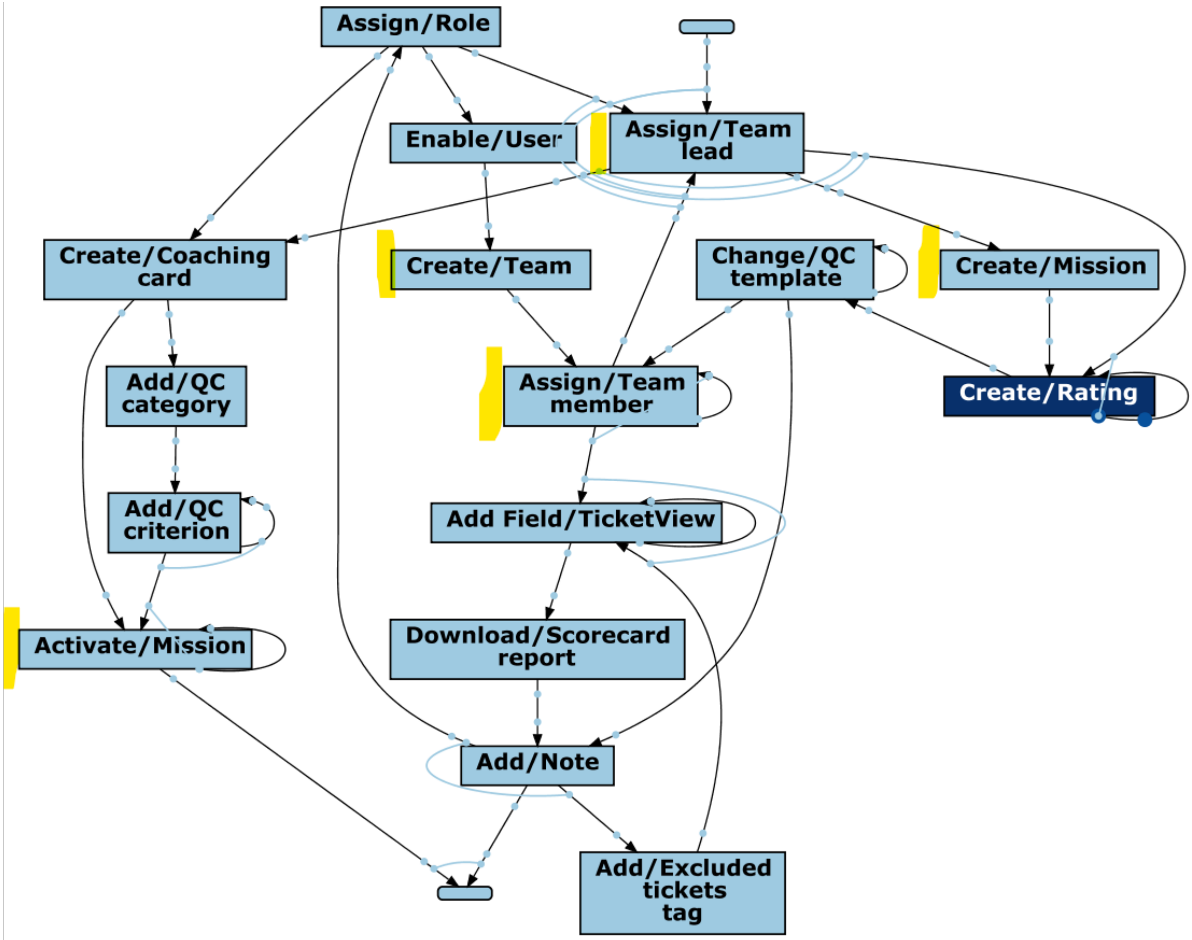


Figure E.6: Activation Phase of Accounts with **Missions** Retention

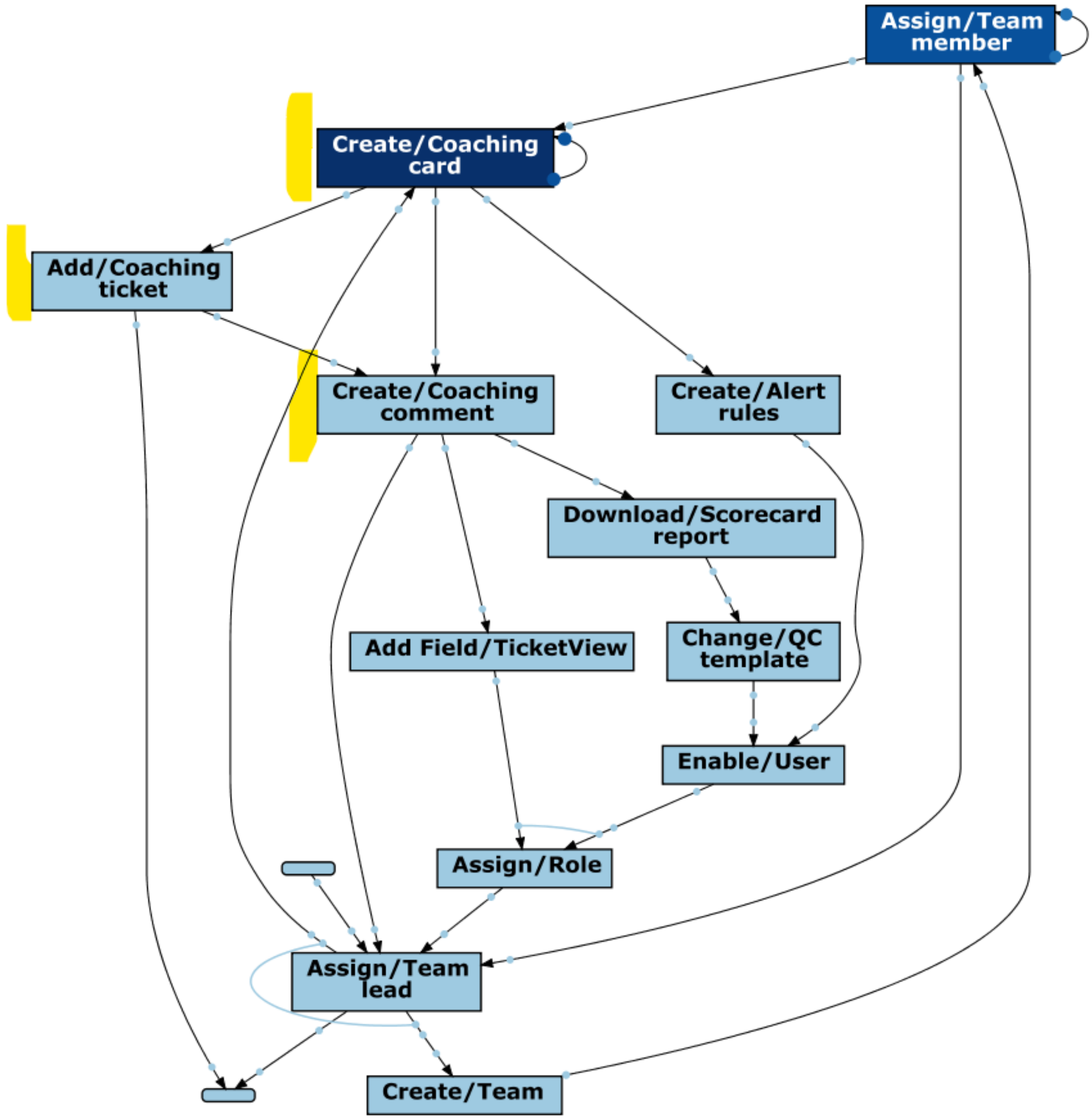


Figure E.7: Activation Phase of Accounts with **Performance Coaching** Retention

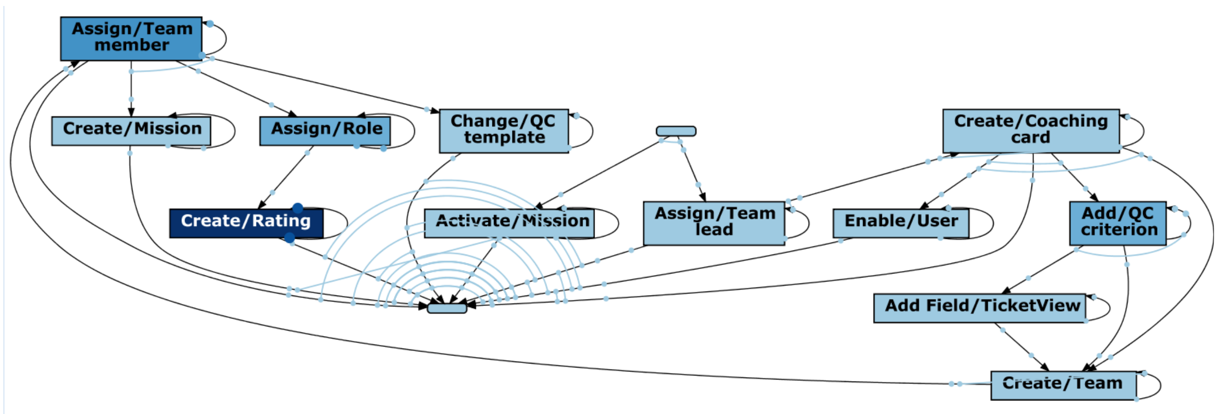


Figure E.8: First Week Behavior of Accounts without Retention

E.4 Pure Pillar Retention Accounts

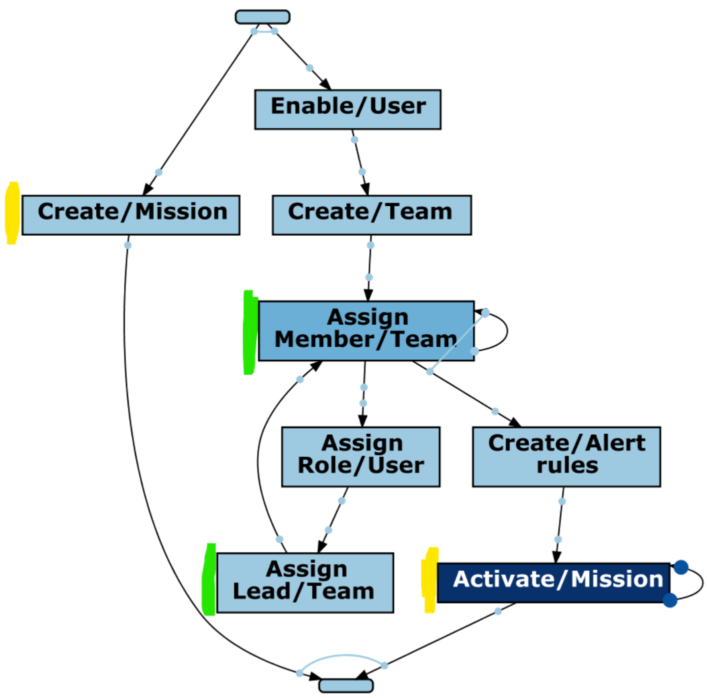


Figure E.9: Mission

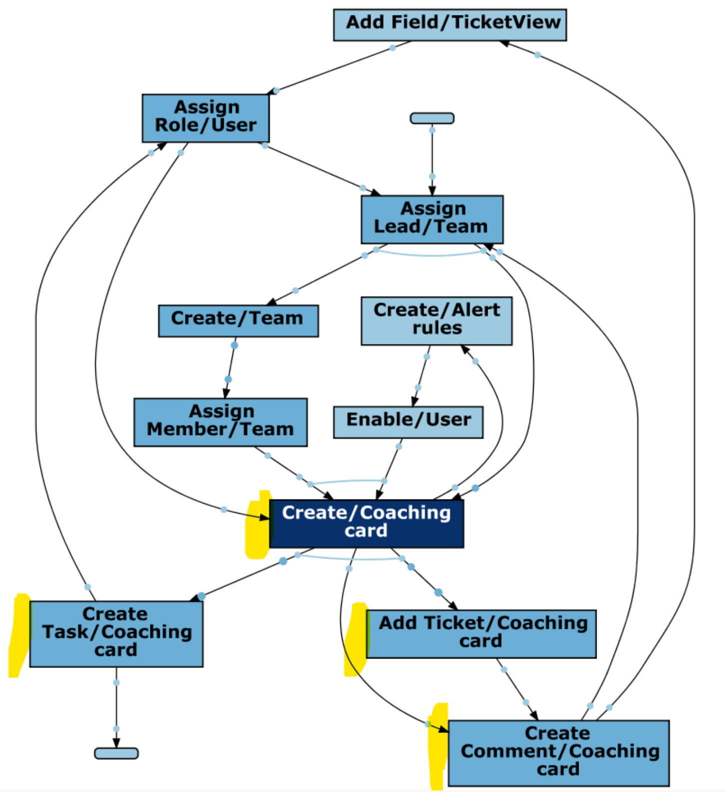


Figure E.10: Coaching

E.5 Last Trial

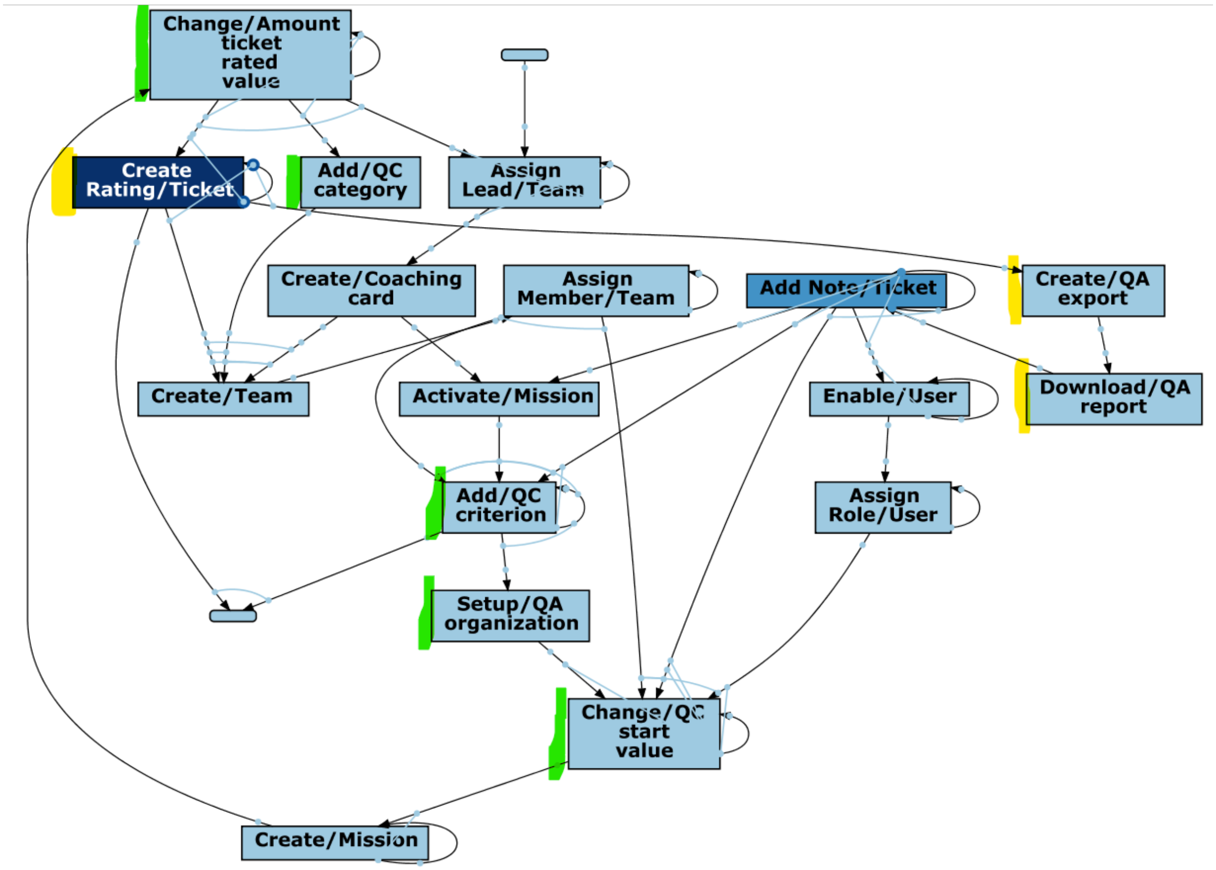


Figure E.11: **Quality Assurance**-Retention Account Activities

The nodes highlighted yellow are constructive activities directly related to **Quality Assurance**; the nodes highlighted green are settings or configuration related to **Quality Assurance**.

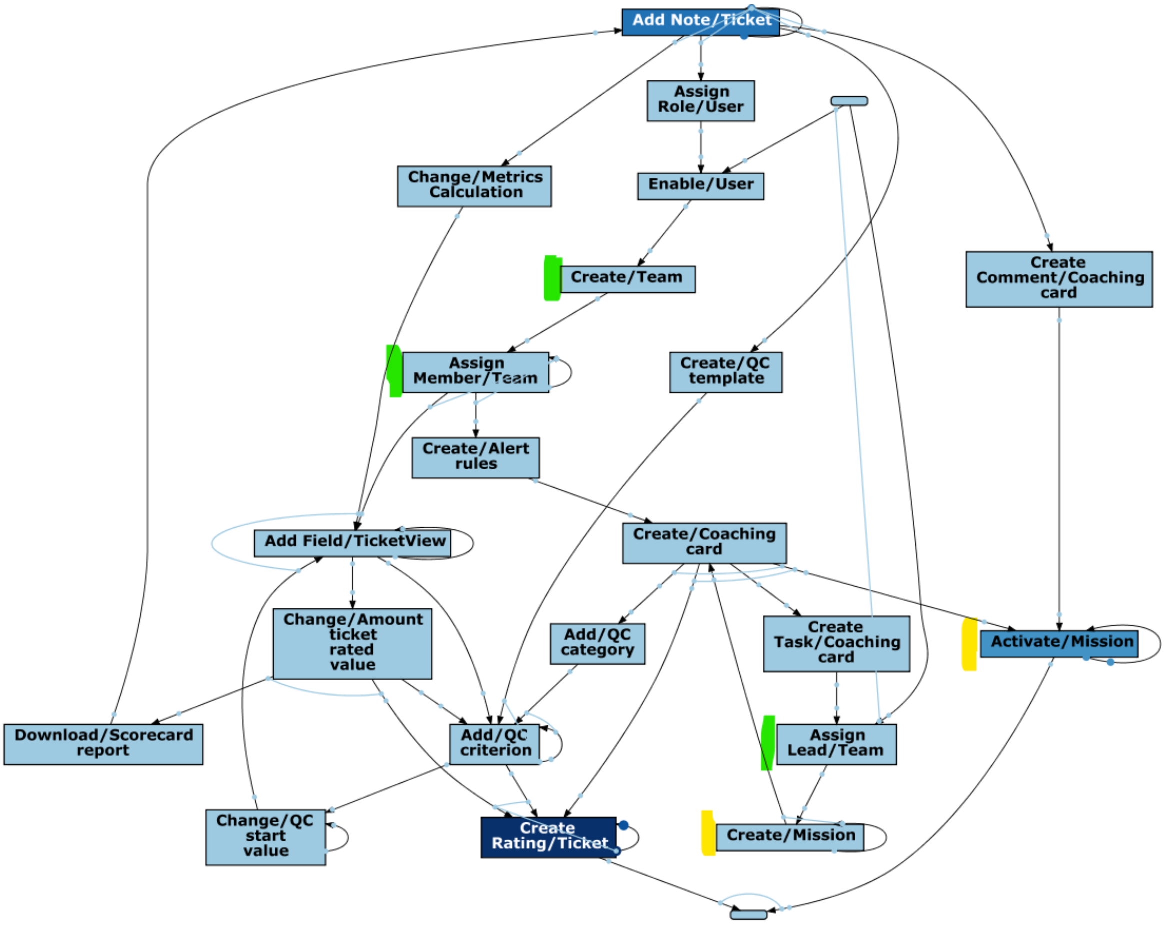


Figure E.12: **Missions**-Retention Account Activities

The nodes highlighted denotes the activities regarding **Missions**; the nodes highlighted green shows the prerequisite events of starting **Missions**.

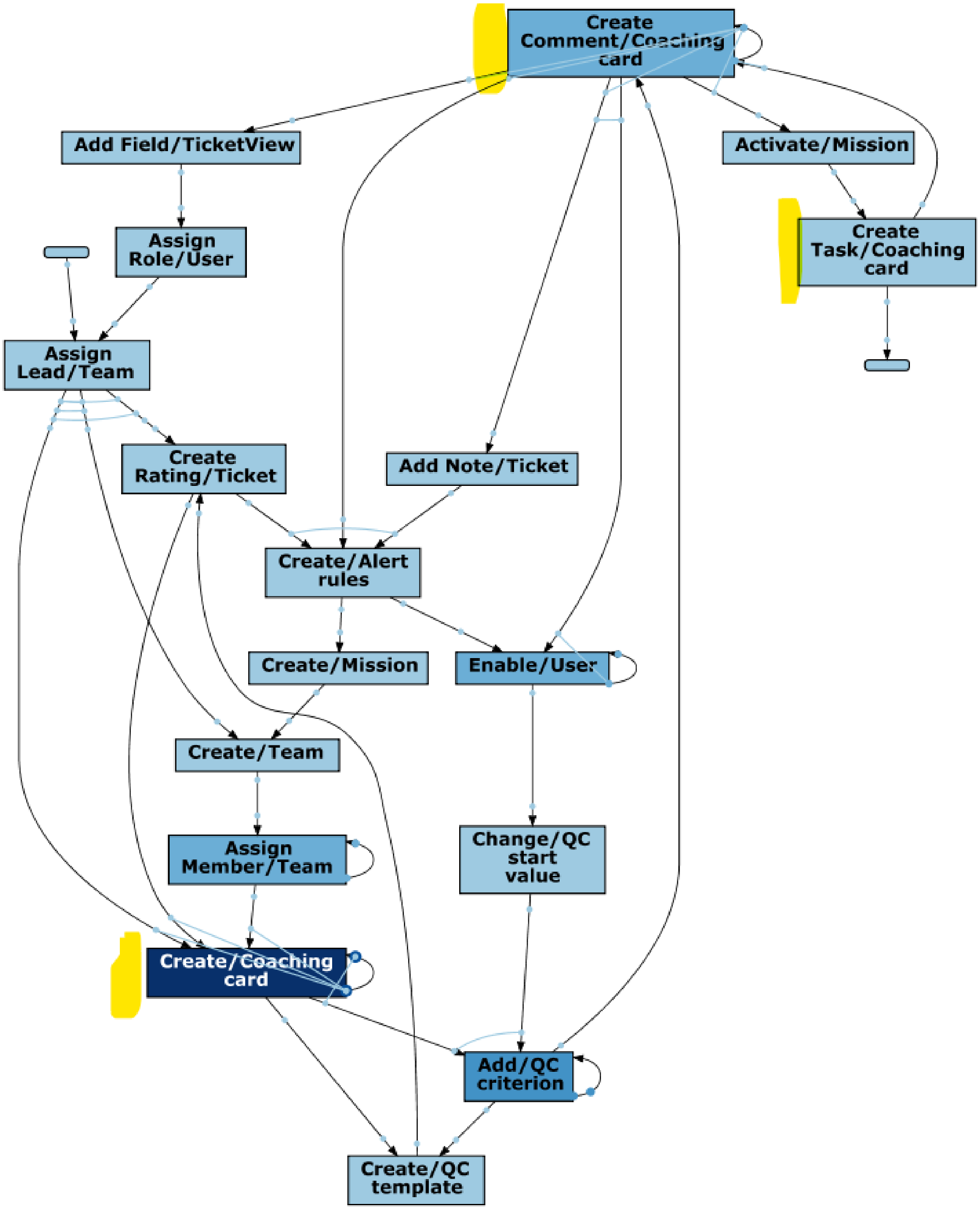


Figure E.13: **Performance Coaching**-Retention Account Activities