Investigating Aha Moment Through Process Mining

Wan-Hsuan Chiang¹, Usman Ahmad², Shenghui Wang¹ and Faiza Allah Bukhsh¹

¹Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands
²Management Science Department, DHA Suffa University Karachi, Pakistan

Keywords: Aha Moment, Process Mining, Customer Satisfaction.

Abstract: Aha moment is when users realize the value of using the software product, which is a key to driving revenue, mainly for B2B SaaS vendors. According to the Acquisition-Activation-Retention-Referral-Revenue (AARRR) model, the aha moment can refer to activation, and the following customer phase is retention. This research aims to find customers’ “aha moment” through process mining. 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. Evaluation is performed through the data of the clustered customers from the time before they move to the retention phase. 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, the concept of human-computer interaction is focused on 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-stratigizing.

1 INTRODUCTION

Digital e-platforms, Amazon or Airbnb, have disrupted the fundamental structure of the industries through the change in essential customer interaction structure (Parker et al., 2016). Particularly, the digital platforms of the business-to-consumer (B2C) sector have already established their potential. However, in business-to-business (B2B) digital platforms are still work in the process. In B2B the digital platforms enable the smooth interaction between supply and demand sides through e-infrastructure (Rix et al., 2020). In B2B e-models, the managers face a real challenge of customer retention (Rix et al., 2020). For instance, in Indonesia, the customer retention rate of BIZGO is very low and customer relationship management is recommended to avoid the opportunities in digital platforms (Situmorang and Harmawan, 2022).

The role of digital marketing is increasing in the post-pandemic era and several B2B firms have started using digital marketing to increase customer acquisition (Assal, 2022). Digital marketing enables businesses to connect with specific customers through targeted marketing activities (Teixeira et al., 2023). Although digital marketing benefits almost every B2B firm, there is a dearth of literature on this topic (Shaltoni, 2017; Pandey et al., 2020). In fact, various businesses believe that digital platforms work for B2C organizations only (Teixeira et al., 2023; Lacka and Chong, 2016). However, the business models of Cisco and IBM are considered the success stories of the B2B digital platforms (Teixeira et al., 2023; Venkatesh et al., 2019). Today due to digital media, the customers of B2B have ample access to information of products and services and this information aids B2B customers to make better decisions. However, traditionally customer relationship management remains a problem for B2B organizations, including online B2B firms (Teixeira et al., 2023; Hochstein et al., 2020).

In order to cope with the challenges of customer relationship management in B2B platforms, the concept of user growth was developed by turning the “passengers on digital website” into customers. This practice helped the various start-ups to retain a large number of users in a short period of time with very low or even zero investment. Based on this practice
of increasing users, a few frameworks have gradually formed in the market that effectively grows the users on digital platforms. In these models, the AARRR model is the most frequently mentioned and used framework in user growth. AARRR is a set of indicators developed by the Dave McClure, the founder of 500 Start-ups, that refers to five stages: acquisition, activation, retention, referral, and revenue (McClure, 2022; Zhang, 2021; Lin and Chaomin, 2021). The activation begins with the first happy visit of the user which is also called as "Aha! moment" Initially, the term is defined as an insight that represents a sudden cognitive change and was adapted to product development, indicating the time when the users realize the value of the product (Carpenter, 2019). The following customer phase of activation is retention. In this phase, businesses consider whether users continuously engage with the product, and the frequency of usage is calculated to understand the repeated behaviors. This repeated behavior leads to the retention of customers and plays a vital role in customer relationship management. Through process mining, we can explore users’ repeated behavior patterns which may lead to improvement in customer relationship management.

Software companies, especially startups, are struggling to develop product strategies. Finding the aha moment is the key to growth, which guide the companies in defining strategies. The aha moment is a set of actions the users realize the value of the product (Stancil, 2015; Balboni, 2022). The idea is often bound with the customer lifecycle where the customer funnel consists of five steps: acquisition, activation, retention, referral, and revenue (McClure, 2022; Zhang, 2021; Lin and Chaomin, 2021). The aha moment usually associates with activation. Users who find the value of the products are more likely to be retained.

The objective of this paper is to study the case of K, a B2B company with a Software-as-a-Service (SaaS) deployment model, and investigate the ”Aha! moment”. The study intends to use the event logs generated by clients of K to find the ”Aha! moment,” based on customers’ repeated behavior. Therefore, the objectives of this study(OS) are as follows:

OS1 - Distinguish the repeated behavior pattern from retained customers.

OS2 - Finding the aha moment to understand customer’s delight.

The research sheds light on applying process mining to customer lifecycle analysis. Applying process mining techniques to discover the aha moment is an innovative approach among previous studies. Introducing human-computer interaction to data cleaning provides a brand-new perspective to re-organize event logs. The results provide the business with a guideline for future product development. The research framework can be further applied to other B2B SaaS customer analyses.

This paper first introduces the background knowledge and methodology used in this research. Section 3 introduces the research questions associated with the objective statement, as well as the framework used in the following experiment. Section 4 demonstrates the process models of the empirical study. We wrap up the paper in Section 6 and Section 7 focusing on conclusions and future work.

2 BACKGROUND

2.1 The Aha! Moment

Companies, especially software service providers, struggle to define the business strategies that customers love, which can be referred to as the aha moment (Aha!, 2022). The Aha! moment is a set of actions the users realize the product’s value (Stancil, 2015; Balboni, 2022). Service providers, especially start-ups, change their product strategies dynamically to test the product market fit or value propositions. Discovering the aha moment guides companies to develop product roadmaps based on customers’ experiences rather than companies’ claims.

2.2 Process Mining

Businesses pay attention to analyzing their customers in order to improve their products and services. However, the existing solutions cannot solve the problems of developing product strategies in companies without clear value propositions because the vendors are still discovering.

There are many commercial analytic tools, such as Google Analytics, Mixpanel, and Tableau, are adopted for customer analysis. However, researchers mentioned the limitation (Vinod et al., 2013; Terragni and Hassani, 2018). Using these tool to analyze customer journeys require a clear understanding of the primary processes, which is not suitable for discovering customer patterns.

Why Process Mining: Quantitative research is powerful during the discovery phase. However, user researchers still need to capture the primary scenarios before recruiting users. However, quantitative methodologies are often costly and time-consuming, which is not practical for small businesses.
Thanks to the widely spread data-driven concept, most companies are used to storing and leveraging data for decision-making. Therefore, process mining discovery techniques become the most sufficient tools for discovering customer behaviors. Process mining adopts event logs to analyze sequential patterns (van der Aalst, 2016). Process discovery is part of process mining that can build models without a priori knowledge, making it the most powerful and commonly-used technique in process mining research (Zerbino et al., 2021; Guzzo et al., 2022).

Previous process mining papers, however, mainly focus on workflow and operational management (Corallo et al., 2020; Zerbino et al., 2021). The limited papers discovering customer behaviors mainly focus on navigation and learning behaviors (Vinod et al., 2013; Zaim et al., 2018; Husin and Ismail, 2021; Taub et al., 2022). None of the studies adopt process mining to discover an aha moment for developing product strategies.

Meanwhile, process mining also shows the potential to integrate with other customer research disciplines, such as customer journey mapping (Chan et al., 2021; Bernard and Andritsos, 2017; Terragni and Hassan, 2018; Terragni and Hassan, 2019) and human-computer interaction (Theis and Darabi, 2019; Else et al., 2019; Liu et al., 2018; Cerone, 2015). Previous papers show the possibility of process mining to support analyzing customer behavior in the evidence base.

Process Discovery Techniques: This research applies the miners that are capable of addressing noisiness because the real world is full of noise, as seen in Figure 1. Heuristic miner (Weijters et al., 2006) and fuzzy miner (Günther and van der Aalst, 2007) are the most widely used miners among previous experimental process mining papers (Zerbino et al., 2021; Corallo et al., 2020). In this study, fuzzy miner is used to produce the test models in order to consult stakeholders due to the intuitive output process map; heuristic miner becomes the primary algorithm because the miner takes frequency into account to capture the primary behaviors.

We use ProM Lite 1.3 to build process models; Python and Google BigQuery to process data.

2.3 Domain Object

In this study, we adapt the fundamental concept of human-computer interaction to restructure the UI logs. The domain object, from the set of potential objects of interest for the user of a given application, forms the basis of the interaction and its purpose (Beaudouin-Lafon, 2000). Each domain object can be viewed as a virtual object on the software for users to create, edit, and change attributes.

The domain object is derived from the object of interest, which is associated with the first proposed visual interaction style, direct manipulation, to replace the command-line interface in the 1980s. The object of interest is a virtual representation that can be directly manipulated (Shneiderman, 1982; Shneiderman, 1983).

The concept of virtual objects and actions plays a vital role in data cleaning. We adapt the concept to reduce the number of events, preserving the semantic meanings. More details can be seen in Section 4.2.

3 METHODOLOGY

In order to discover the aha moment, our research questions are as follows:

- **RQ1** - What are the behavior patterns of the clients in the retention phase for B2B platforms?
- **RQ2** - What is the “Aha! moment” of clients for B2B platforms, leading to retention?

Adapted from CRISP-DM (Wirth and Hipp, 2000). Figure 2 shows the research plan of this paper. There are five steps:

**Step 1** - Understanding: This step contains business and data understanding in CRISP-DM.
Step 2 - Initial Data Preparation: This step extracts and cleans the data from the data warehouse.

Step 3 - Retention Clustering: The retention clustering answers RQ1 as discussed in Section 4.3.

Step 4 - Discover Aha! Moment: The output from the previous step is the input of this step to build a process model for discovering the user activation.

Step 5 - Stakeholder Feedback: Aligned with the evaluation phase in CRISP-DM, is the final step of this project, discussed in Section 5.

Step 2, Step 3, and Step 4 represents the loops data preparation and modeling. Each step contains more than 10 small and big iterations. The following discussions only write down the big iterations.

Step 5 includes stakeholders’ views on the models. However, the involvement of the domain knowledge does not limit to the final stage.

This study adopts the business context from K, a B2B SaaS vendor providing performance tracking service for customer support (CRM) teams. The company actively collected customer behavior data in the past. We select the customers that installed the software between December 27, 2021, and June 26, 2022. The records start from December 27, 2021, to July 3, 2022. We make sure the records cover the accounts’ actions for more than a week. The accounts for internal use or uninstalled within a week are filtered out.

4 ANALYSIS

4.1 Dimensions

We recognize several dimensions to structure the analysis during the business understanding phase in Step 1. These are the main factors we consider during the research. Besides the time dimension, the study can be discussed under 6 scenarios, with 3 pillars and 2 role types. Each scenario has its own process model as it has its particular behavior pattern.

4.1.1 Pillar and Domain Object

The study can be divided into 3 pillars, representing the solutions provided by the product: quality assurance, missions, and performance coaching. Each solution aims to solve a particular pain point from target customers. We also recognize the domain object associated with each pillar.

4.1.2 Role Type

The role in B2B software products for operational needs often refers to permission, reflecting the real-world position in the organization. In our study, the product has 5 different roles. The lowest is the agent, while the highest is the account owner. If we consider performance tracking as a game requiring two parties to interact, the roles can be classified into manager and agent, and the former supervises the latter.

4.1.3 Time

The business customers follow the weekly operation cycle from Monday to Sunday. Customers who have repeated behaviors week over a week are considered retention, or actively using such features.

4.2 Data Cleaning

The job in this subsection is done during the data preparation in Step 3, retention clustering. However, since the event classification influence most reshapes the implementation at the very beginning, this subsection can be viewed as part of Step 2, initial data preparation.

The diverse event naming leads to spaghetti-like models (Figure 1). The original event logs are named by UI components, design purpose, user actions, and domain objects. The UI components change over time and versions; the design purpose frequently changes in order to test different user flows.

Therefore, we adopt the concept of domain object as discussed in Section 2.3 to diminish the interference of user interface and design purpose. By focusing on the domain object and the actions that the user act on, the event naming reflects the motivation behind it. For instance, performance_journey_coachable_moment_comment_created event can be divided into Create Comment and Coaching card. The latter is the associated domain object of one of the pillars. The former reflects that users act on the attribute Comment of the virtual object.

We further identified the duplicated events and nodes without semantic meanings. The number of nodes, in the end, is reduced by 62%.

4.3 Retention Clustering

We first consult stakeholders from customer success team about which behaviors can be recognized as retained customers. The findings can be extracted in Figure 4 that retained customers in a particular pillar should contain milestone actions from both parties, inspired by identifying key actions of business goals.
for preprocessing (Peng and Cheng, 2009). In addition, the behaviors should repeat week over week, according to the definition of retention and business understanding. However, the frequency is not clearly identified within the business.

After the interview, we extracted 4 scenarios with clear milestone actions. We select the users’ working days with milestone actions as cases to build the process models.

The remaining 2 scenarios consider the domain object as cases. The node combines action and role type that who acts on the objects can be clearly shown. The process models uncover the actions on each domain object, as well as the interaction between managers and agents. Figure 3a shows the process model of the domain object of performance coaching. The nodes highlighted in red are the actions from managers; the nodes highlighted in green are the actions from agents. The process starts from Create by managers and then Open by agents. Before the end, managers Open the domain object again for reviewing.

Based on the retention modeling, milestone actions of all scenarios are verified and recognized. Figure 5 are the dotted charts in terms of pillar. The logs are colored by the milestone actions of both parties. The y-axis is the business customers, and the x-axis is week to capture the weekly patterns. The milestone actions, however, filtered out the test actions, such as self-creating domain objects. The companies containing the milestone actions from both role types week over week are considered as retained customers. As a result, three sets of retained customers are clearly identified.

4.4 Discover Aha! Moment

This subsection addresses RQ2 what the aha moment looks like in terms of pillar and user role? Based on the previous retention clustering, the retained customers are selected in terms of pillars.

The event sequences of the business customers are sliced into two periods: activation and retention. We select the time period before the clustered customers start the retention patterns in order to associate the activation period. That is, the output models start from Installation and end at Start Retention. Meanwhile, the events that are not "constructive" are filtered out to simplify the model. Such events...
include the visit events that occur too frequently and the events not triggered by users. Furthermore, the follow-up actions such as Edit, Disable, and Delete are filtered out. Since we are capturing the primary interest, for instance, Create, which is the prerequisite of the follow-up actions, can represent the interest in the models.

Figure 3b shows the retained companies using quality assurance in activation phase. The process highlighted with the orange circle shows the account’s interest in the associated feature configuration; the process with a yellow circle indicates the nodes with the interest of team management.

The models successfully show the prerequisites of adopting the features, including feature configurations. Apart from feature settings, retained customers of quality assurance actively modify the qualify standard and criterion; missions model shows strong interest in monitoring performance since the function is meant to provide goals for agents to improve. The models for performance coaching contain a mixed behavior pattern, positioning the function as an assisted solution among the other two features.

Abstracting the aha moment from the models involves graphical analysis and domain knowledge. The understanding of each event log and the relationship with the primary pillar functions plays an essential role.

5 EVALUATION

During the experiments, we continuously involve domain knowledge and stakeholders’ feedback on improving the modeling. The small model adjustments rely on authors’ work experience and domain knowledge in K; the iterations in larger scale, involving the change of dataset request feedback from sales, customer success, and analyst team. The feedbacks can be viewed as the evaluation of the results, which is the last step, Step 5.

Half of the stakeholders with business and strategy mindsets involves in the evaluation or consulting. Employees with the functions of marketing, sales, customer support, product planning, and analysis are selected because they have the potential to get in touch with customers. In the whole project, the colleague from the customer support team in the case company is highly involved due to their deep understanding of customer behavior.

The results reflect the primary paths which are reasonable for stakeholders. The behavior patterns meet the company’s understanding of customers. The identified frequent behaviors, creating teams, even met the product development strategy at the same time, verifying the company’s observation of customers. However, the project failed to discuss the interaction between pillars, which was the current focus of the company at that time. The gap might due to the fact that the business strategy changes rapidly during the project conducted.

The representation is another factor that many researchers ignored. The process map from fuzzy miner is intuitive but not simple enough, due to the fact that the algorithm does not consider frequency to simplify the model. However, heuristic miner does not guarantee an understandable model. While applying all-tasks-connected heuristic, the models become more explainable, verifying the statement that many dependency relations can be hidden without records (Weijters and Ribeiro, 2011).

It is also important to know the stakeholders’ responsibility. It is natural that people only focus on what they are responsible for and care of. For instance, in this case, the opinions of stakeholders who focus on customer relationships are more valid than
the reviews from the sales team.

6 DISCUSSION

This study applies process mining to discover customer behavior on retention and activation, which is a brand-new field that no previous researchers have done. The involvement of customer funnel (McClure, 2022) provides a framework to analyze customers, especially for software industries. Companies can easily adapt the framework to their customer journey to discover more processes.

Understanding how customers experience the value of the product is the key to growth. Discovering the aha moment provides companies with a new approach to defining a bottom-up and evidence-based value proposition, which is more accurate than a top-down approach. During the discovery phase, decision-makers in the business are able to develop better business strategies and product roadmaps. This insight can help develop a better onboarding flow, reducing the churn rate at the beginning.

The integration of human-computer interaction to process mining is also another newly discovered field. Adopting the concept of domain object to classify the event logs is a practical method to clean the data because it reduces 62% of the nodes in this research. The concept transforms the event logs into actual incentives. This approach is especially suitable for cleaning weblog and UI logs because these logs are usually noisy and dynamically changing.

Although structuring the research by-product solutions is a practical framework, it failed to discover the interaction between each pillar since, in general, the users act on the same platform. The result shows that one of the pillars is the supportive function of the other two, and the retained customers of this pillar have more proportion of old installations, which is out of scope in this research. It would be interesting to see the interaction between these pillars.

7 CONCLUSION

During the business understanding, we structure the research into pillars and role type, with 6 scenarios. Meanwhile, time is also taken into account as it plays an important role on identifying repeated patterns.

Inspired by the domain object concept, the domain object in terms of pillar, and associated milestone actions by role type are recognized. Furthermore, the same concept is used to reduce the event nodes and the data size by diminishing the interference of dynamically changed user interfaces.

After data cleaning, we address RQ1 to cluster retained accounts. Customers with repeated milestone actions from each role type over three weeks are identified as retained accounts in terms of pillar. Scenarios with pre-recognized milestone actions are verified; scenarios missing milestone actions are modeled by the domain object to capture the actions from both parties.

We further slice the event sequences into activation and retention from retention clusters to address RQ2. The logs before the accounts start retention are added as an input to build the models. Based on the graphical analysis and the involvement of domain knowledge, a set of actions and the motivations behind them are recognized.

This research framework can be applied to B2B SaaS industry to analyze their customers. The study can also apply to a bigger dataset because the event classification can minimize the number of nodes. However, the data cleaning itself requires a lot of manual work, which should be re-considered in the future for scalability.

Some of the limitations should be noted. This project stopped at the discovery phase but failed to apply the findings to real business contexts. The evaluation only relies on domain knowledge and stakeholders’ eye. The limitation is due to the time limitation and project scale. It would be interesting to see the following influence on the product and the possible process enhancement.

The research only focuses on the specific time period and customer phases. We do not discuss the entire customer lifecycle, from acquisition to revenue. We only focus on the newly installed customers and do not discuss the longer customer relationship. It could be interesting to discover the aha moment not only from the retention but also from the revenue phase. Meanwhile, applying process mining to other customer phases remains a gap in the research field.

REFERENCES


