Impact and Integration of information monitoring in the context of a software startup: how to increase user retention

Master thesis submitted to Delft University of Technologyin partial fulfilment of the requirements for the degree of **MASTER OF SCIENCE**

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Alberto Incisa della Rocchetta

Student number: 5383285

To be defended in public on June 12th 2023

Graduation committee

Chairperson	: Dr., R.M., Verburg, Economics of Technology and Innovation
First Supervisor	: Dr., N., Pachos-Fokialis, Technology Policy and Management
Second Supervisor	: Dr., A.M.G., Zuiderwijk- van Eijk, Engineering Systems & Services

Abstract

Software start-ups are important sources of innovation, new services and new software products. They play a vital role in the economy by driving innovation and creating new jobs as well as being a source of breakthrough technologies. However, most software startups fail, during the first two years. This is due to a lack of customers rather than for technological reasons. Specifically, companies struggle to retain the users that they are acquiring. In this way, startups struggle to build a loyal customer base that will help them grow and refine the value proposition.

This research explores how a modern startup is using information monitoring to increase user retention. User retention refers to the ability of a company to retain its existing customers over a period of time. Essentially, it measures how many customers continue to use the product or service after the initial purchase or sign-up. By constantly monitoring user behaviour and reacting quickly to any changes, startups can keep their users engaged and satisfied with their product or service and also better target their acquisition efforts. This thesis aims to explore the strategies and metrics, which are derived from product usage, that software startups use to achieve high user retention rates.

Literature concerning frameworks, characteristics and current use of business metrics in the context of software startups has been reviewed and integrated with the knowledge gained from a single case study on a modern startup that has been using product usage metrics to increase user retention. The research is conducted through a case study approach and includes a thorough literature review of previous studies on software startups, their strategies, reasons for failure, and the importance of monitoring and metrics for achieving success.

Case studies are a suitable methodology for investigating complex systems such as software start-ups, which are influenced by multiple factors that may impact user retention. This is due to the flexible nature of case studies and their focus on examining a particular phenomenon in a specific context. User retention can be influenced by various contextual factors that need to be understood an taken into account in a meaningful way. Through case studies, researchers can investigate the specific contextual factors that affect user retention in a particular start-up, allowing for a nuanced understanding of how these contextual factors influence user retention, and how product usage data monitoring can be used effectively to improve retention. To this end, the methodology adopted in this thesis involves a qualitative research design and a single case exploratory study approach, which allows for a detailed investigation of the specific startups practices and how they relate to user retention. Secondary data from the database of a software startup has been gathered and analyzed. The results of the case study should reveal the conditions under which users are more likely to retain. Assembling various literature pieces concerning startup frameworks with the current state of the art regarding information monitoring provides substantial contribution.

Findings have been formulated in a way that could be applied by most software startups. The study found that user eligibility, user traits, and collaboration all play a role in user retention. Eligibility refers to the user's ability to interact with specific features in the product, which is directly related to the technical infrastructure the user has in place. In this

case, user traits refer to the job roles of the users in the product, such as founders, product managers, and engineers. Lastly, collaboration refers to multiple users working together on a shared project or task within the same workspace.

The study found that user eligibility has a significant impact on user retention through its effect on user engagement. Higher levels of eligibility were found to lead to increased engagement, though not necessarily higher retention rates. Furthermore, user traits were found to be positively associated with user retention, with founders displaying higher retention rates compared to product managers and engineers. Additionally, collaboration was found to have a positive impact on retention across all user groups.

The final chapter of the case study discusses the practical implications of the findings and provides learnings for software startups to follow in order to increase user retention. Specifically software startups should identify factors that could influence retention. Then, with those in mind, product usage data should be gathered an analysed to understand under which conditions users are more likely to retain. With this knowledge in hand, organisations can then implement specific strategies to improve those areas and keep users coming back for more time. In this particular case, the main factors that impact user retention, such as eligibility and engagement, were identified and analyzed. Based on these findings, it is recommended that companies focus their resources on initiatives that increase eligibility and engagement, target marketing campaigns to acquire users with positive retention characteristics, and promote collaboration in the app.

Finally, information monitoring is found to be very well embeddable in the main startup strategies discussed in literature. This further strengthens the value of the strategies to follow because it is not replacing the ones that worked well up until now, but rather it is fostering them.

The study acknowledges its limitations, including its non-generalizability and small sample size. Suggestions for future research include expanding the sample size and using a mixed-methods approach to obtain a more comprehensive understanding of user retention in software startups.

In summary, this thesis contributes to the understanding of the strategies and metrics that software startups can use to achieve high user retention rates. The findings provide valuable insights for software startups looking to increase their user retention and establish a sustainable business model.

Acknowledgements

This thesis is the result of the passion for software startups I developed during the last 12 months. Since September 2021 I have been collaborating with an early stage software startup where I learned about the importance of achieving good user retention. This has stimulated me to undertake a literature review about this topic, at first, and this thesis afterwards, to understand how the practices we were following to achieve high user retention could be generalised and integrated with the current frameworks. Seeing, in first person, how a newly created technology company is managed, has been an incredible opportunity for me to understand how the concepts learned during the MOT course are put in practice. Moreover, I feel like the knowledge gathered for this thesis will definitely have a positive contribution to the work I will do at June.so.

I would like to thank Dr Nikos Pachos Fokialis for giving me all the support and guidance on how to approach such a work. You have always been available, and reactive, answering all my questions. I am also thankful that you have been believing in this project from the really beginning.

I want to thank Professor Robert Verburg and Dr Anneke Zuiderwijk - van Eijk for their guidance and precious feedback that they delivered to me throughout this project.

I am extremely grateful to Enzo Avigo, Ferruccio Balestreri and the rest of the June team for being so generous with me. You have been sharing with me all your tacit and explicit knowledge, captured in years of work, to make me grow. You also managed to get me passionate about building something new, taking many challenges and collaborating all together. I also appreciate a lot that you allowed me to use June as the object of the case study which is the backbone of this thesis. Without you this would have not been possible. Collaborating with you all has, and still is, a privilege for me.

Finally I want to express my gratitude to my family, girlfriend and friends for supporting me during this work. Nice, still challenging, experiences such as this one would not be the same without you all.

. . .

Contents

1 Introduction

Software start-ups are important sources of innovation, new services and new software products [Klotins et al., 2021c]. These kinds of new ventures are considered crucial drivers of economic development [Rafiq and Wang, 2020].

In the software industry, startups are newly created companies that are typically operating with resource scarcity, time pressure and high level of uncertainty [Paternoster et al., 2014]. The goal of such organisations is to achieve repeatable growth after having found a viable business model [Ries, 2011].

Software start-ups are based, especially during the early stages, on engineering practices and skills. Start-ups need an even higher level of engineering practices given the scarcity of time and resources [Paternoster et al., 2014].

According to Giardino et al. [2014a] due to a vast adoption of Internet and lower costs of starting ventures, the number of newly created software start-ups is increasing more and more; however, a vast majority of these fail during the first three years [Bednár and Tarišková, 2017]. Most of the times, failure is not due to lack of technology or engineering capabilities but because of lack of customers [Crowne, 2002].

Given that most of these companies operate under severe resource constraint, the lack of customers cannot be solved just by allocating more financial resources to marketing but rather focusing on keeping satisfied the ones that are already using the service. This because acquiring a new customer is said to be five times more expensive that retaining one [Pfeifer, 2005]. It is therefore crucial to understand how to keep having customers over time while keeping the customer acquisition cost (CAC) at an acceptable level.

1.1 Problem Statement

The lack of customers, which is ultimately the main reason of software startup failure according to Crowne [2002], is mainly an effect of lack of *problem-solution fit*. Giardino et al. [2014a] state that this *fit* is achieved when the product continuously manages to solve a problem that users encounter.

Still, measuring such a fact is not straightforward. According to Sayyed-Alikhani et al. [2021], user retention is a good indicator of *problem-solution fit*. This because a user only continues to use a product if it manages to deliver value by solving one or multiple problems, over and over again. Moreover loyal customers (also known as retained customers) are an excellent source of referrals, which is a cost-effective customer acquisition strategy [Nasir, 2015]. Pursuing an increase of retention could be particularly suitable given that it is said to help companies increase customization and improve communication which leads to better relationship with users and higher customer satisfaction[Zhang et al., 2020]. User retention policies are also said to be preferable in spite of focusing on user acquisition ones. Without an increase of retention, acquisition policies will not increase the total number of users appreciably [Sayyed-Alikhani et al., 2021]. This further reaffirms the crucial role of user retention during software startups' early stages.

However, it is unclear how software startups can effectively increase user retention in order to achieve a better *problem-solution fit* and ultimately retain more customers. This is particularly important given the constraints on resources, the higher cost of acquiring new customers compared to retaining existing ones [Pfeifer, 2005], and the need for software startups to maintain customer satisfaction [Zhang et al., 2020]. The challenge lies in identifying the factors that influence user retention and improving them to optimize it and enhance the *problem-solution fit* in software startups. The main challenge for software startups is therefore to identify the factors that influence user retention and improve them. By doing so, startups can enhance the *problem-solution fit*, leading to increased customer loyalty and a higher likelihood of success Giardino et al. [2014a]. However, the specific factors that impact user retention may vary depending on the product or service offered and the industry in which the startup operates. These elements could be captured and measured by means of product usage data monitoring.

With standard web technologies it is possible to monitor, in a detailed way, how users interact with a product [Atterer et al., 2006]. Therefore, product usage data monitoring refers to the process of collecting and analyzing data related to how customers are using a particular product. This can include data such as how frequently the product is utilised, which features are being used most often, how long customers are spending on certain tasks, and other relevant metrics.

1.2 Research Questions

To tackle the problem(s) mentioned in 1.1 some questions are formulated:

RQ: *How can software start-ups increase their user retention by means of product usage data monitoring?*

User retention plays an important, indeed crucial role during start-ups' early stages. However, no studies have been undertaken on how tracking and measuring user actions and characteristics could help increase user retention over time.

The above mentioned main Research Question is decided into three sub-questions:

SQ1: *How can user eligibility influence user retention through an effect of user engagement?*

Product value proposition can be perceived only if users connect their own data source to the platform. However, there are various possible states of eligibility. This variable could also influence the engagement with certain features of the product which, in turn, is assumed to impact user retention.

SQ2: *How can user traits and collaboration affect the relationship between user engagement and user retention?*

User retention represents perceived ongoing value so we assume that high engagement with the features of the product, measured by the number of track events triggered, could positively impact user retention. This relationship could also be affected by factors such as user traits or collaboration between colleagues.

SQ3: *How does continuous information monitoring integrate with the current software start-up strategies?*

While there is existing literature that outlines best practices for software startup strategies, it doesn't explicitly address how continuous information monitoring could fit into those strategies. This is an important consideration, as leveraging real-time data and analysis can provide startups with a competitive advantage. Therefore, the question seeks to explore how startups can successfully incorporate continuous information monitoring into their already well-established operational strategies.

By addressing all sub-questions first, we will be able to answer the main one.

Specifically SQ1 and SQ2 will be tackled with the learnings gathered while conducting the case study. SQ3 instead will instead be answered by analysing and assembling the current literature with the current technology capabilities.

1.3 Contribution

This work aims to have both a scientific and societal contribution.

Firstly, the objective is to start filling the current knowledge gap around how software startups could increase their user retention, providing scientific contribution. Literature currently provides us with many resources about which strategies startup should pursue [Ries, 2011] to achieve success as well as which metrics should be tracked Kemell et al. [2018]. However there is no knowledge about how to combine both together.

This thesis is focused on software startups and specifically on how to make them successful. Managing and taking strategic decisions in such a type of venture is definitely aligned with the courses of the Master in Management of Technology (MOT). In particular, observing and studying how strategic topics are tackled with a scientific methodology is the application of what the MOT course is all about. This study, as well as the MOT course, is about using modern technologies to address complex questions and improve current processes.

This study could also have a positive societal impact. Specifically, it could help increase the number of customers available to startups, increasing the chances that more and more of them could remain in business and therefore maintaining, and also generating, more jobs. This societal impact is not negligible given that, according to Bednár and Tarišková [2017], the number of new projects started every year is in the order millions, therefore having an important contribution to the labour market [Pantiuchina et al., 2017].

1.4 Organisation

This study is structured as follows.

Chapter 2, Literature review: Literature study concerning typical startup strategies and frameworks along with reasons of failure. A review of the common metrics used and data monitoring is also presented.

Chapter 3, Methodology: Overview of the research methods used and reasons for such a choice. Presentation of the conceptual model, variables and context of the case study. Data collection.

Chapter 4, Results: The outcome of the case study is presented and dicussed.

Chapter 5, Conclusion: research questions are answered, reflections are discussed together with contributions and future research.

2 Literature review

To gain context about how and why software startups fail to have customers and how this could be tackled, a thorough literature review has been performed.

In first place, in Subsection 2.2.1, the main startup strategies found in literature are reviewed to comprehend how these type of software companies work and which type of frameworks and strategies are predominantly followed by the founding team.

Secondly, in Subsection 2.2.3, reasons of failure are reviewed an analysed to get all the context needed to be able to derive possible solutions.

After this, in Section 2.3, an in-depth study of the current state of the art, implementation and usage of information monitoring is presented to ground our knowledge about what this technology is and how it is currently used by startups. The last Subsection 2.3.1 of this chapter is dedicated to better understand User Retention.

2.1 Method

A literature review offers an extensive summary of the existing literature relevant to a particular topic, theory, or method, and it combines previous research in order to solidify the knowledge base Paul and Criado [2020]. A literature review is a necessary step before writing a thesis or an article. This is because, if well conducted, it will help to find a literature gap, which is the main motivation for scholars to write a new paper. A literature review should be conducted following the scientific method, i.e should be repeatable by peers. For this reason hereafter, the process pursued to write this work is documented in detail Brocke et al. [2009]. As a first step of the search, a general assessment of the articles was done. Answering questions like: Does this article relate to user retention in the context of software startups? Is the content of this article connected to the operational tactics employed by software startups? Does the article discuss the possible causes of a startup's failures? Does the article discuss the use of data analytics in the context of a startup?

The research was carried out in an iterative way using keywords in ScienceDirect and Google Scholar. The research question of this study is to understand how startup companies (especially software-based ones) can increase user retention by using product usage data monitoring. To find relevant literature, a variety of keywords related to startup and software development were used. The first attempt to use the keyword "user retention" AND "startups" only gave less than 10 results in ScienceDirect but over 300 in Google Scholar. To narrow down the focus, in the Google Scholar search the keywords "mobile" and "game" were excluded because of the low relevancy in regards to this study. Articles were selected keeping in mind the questions mentioned in the above section. Similarly, another search was performed "software startups" AND "framework" AND "early stage" in both databases. Additionally, case studies, research methods, metrics, value proposition, customer loyalty, customer retention, lean startup, and agile development were also used as keywords to find relevant literature.

It is important to note that a significant portion of the papers that were found during the literature research, seemed to concentrate on how venture capital financing affects the success of companies. Although this is a pertinent and significant subject, it might not always offer useful insights for developing a product that addresses a problem in the actual world. This is due to the fact that success in the startup world depends not only on financing but also on finding a genuine issue, creating a solution that appeals to the target market, and successfully marketing and scaling the product. So, even though financing is undoubtedly an essential component of beginning and expanding a company, it is not the only thing to take into account when developing a successful and sustainable startup. Similarly, while engineering practices and UX issues are important for creating a successful product, it is worth noting that several articles in the literature review appear to focus heavily on these topics, which may not provide a comprehensive solution to the specific challenge of increasing user retention. The key to improving user retention lies in understanding the target audience's needs and preferences, identifying pain points and frictions, and making data-driven decisions by continuously monitoring and analyzing usage data. While funding and engineering practices are crucial components of creating and growing a business, they may not necessarily provide constructive insights into understanding how to increase retention. Therefore, it is important to focus on the specific challenge at hand when reviewing the literature and to consider the most relevant and actionable insights to achieve the desired outcomes.

2.2 Software Startups

Software start-ups are newly created companies that typically operate under multiple constraints which can arguably be used as a definition for the term start-up. According to Paternoster et al. [2014], there is no clear definition of such a concept in literature, but the most common traits are: 1. resource scarcity; 2. high level of reactivity; 3. high level of innovation; 4. uncertainty and 5. time pressure.

It is worth noting that such characteristics are not exclusive to start-ups, but they are also shared by other software-focused organizations [Klotins, 2018]. As stated by Klotins et al. [2021a], startups typically go through a life cycle composed of these stages: inception, stabilization, growth, and maturity which differentiate them from the already established organizations.

To elaborate further, the first two stages are referred to as the "early stages" of the company, from the genesis of the idea to the preparation of the growth phase. This study is focused on this particular phase.

2.2.1 Software startups strategies to avoid failure

Software start-ups are based, especially during the early stages, on engineering practices and skills. Given the resource and time limitations mentioned above, start-ups need to be effective and highly reactive in building and marketing their product compared to an already established firm. For this reason, start-ups need an even higher level of engineering practices given the scarcity of time and resources, as Melegati et al. [2020] affirms. Moreover, this kind of venture cannot rely on traditional business planning as already established companies do, because there is no previous experience to use as a driver for strategic decisions according to Blank [2013] and also, as mentioned in the previous section, there are multiple stages to go through. In addition, software companies, especially startups with their severe resource and time constraints, should be as *user-centric* as possible. This notion was introduced by Debellis and Haapala [1995] referring to the procedure of involving end users in the development phase of a software. This can be done during the implementation of a whole business model, which is ultimately what a startup is. In the entrepreneurship field, many agree that startups should validate their business model by quick and numerous iterations [Baker and Nelson, 2005, Sarasvathy, 2001].

All the main startup strategies that will now be presented, have at the center the concept of involving users from the really beginning to gather feedback quickly as Yaman et al. [2020] states, given that the risk of failure lies in releasing a feature that users do not need.

Lean Startup movement

Ries [2011]'s book proposed a business model strategy based on quick iterations, that became almost the standard when starting a new software company according to Yang et al. [2019]. Still, it is worth noting that this approach was previously introduced by others like Baker and Nelson [2005], Blank [2007], Sarasvathy [2001] as Bortolini et al. [2021] point out. Ries [2011] contrived the term Lean Startup to refer to the strategy of quickly iterating to find a sustainable business model. The shortest the time of a loop the more iterations can be carried, the higher the probability of escalating.

The framework proposed in the Lean Startup is the so-called Build-Measure-Learn loop, represented in Figure 2.1.

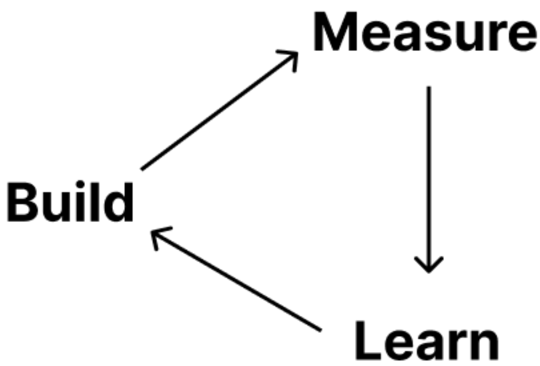

Figure 2.1: Build-Measure-Learn loop

Once the business ideation, or inception according to Paternoster et al. [2014], which is the reason why entrepreneurs start a company, is completed, the Lean build-measure-learn loops starts. Ideas and hypotheses related to the business model are formulated and then built in the form of experiments like minimum viable products (MVP), prototypes, a / b testing etc [Blank and Dorf, 2012]. It is worth noting that minimum viable product is among the most common ways of carrying out experiments in software start-ups Duc and

Abrahamsson [2016]. Then, results of these experiments are measured by means of data which is then assembled into learning.

According to Ries [2011], learnings can lead to different actions: iterating, escalating, pivoting or renouncing. This set of actions must be carried out until the outcome is either escalating or giving up. After discarding an hypothesis which led to a specific experiment, the outcome can be of different types. Iterating if the measurements revealed that just a minor adjustment is needed within the same business model. Pivoting if the measurements revealed that one or multiple changes should be made to the business model or giving up if the team does not believe anymore that a sustainable business model can be found. This process is shown in Figure 2.2. In Subsection 2.2.3, pivots and other types of failures will be discussed in detail.

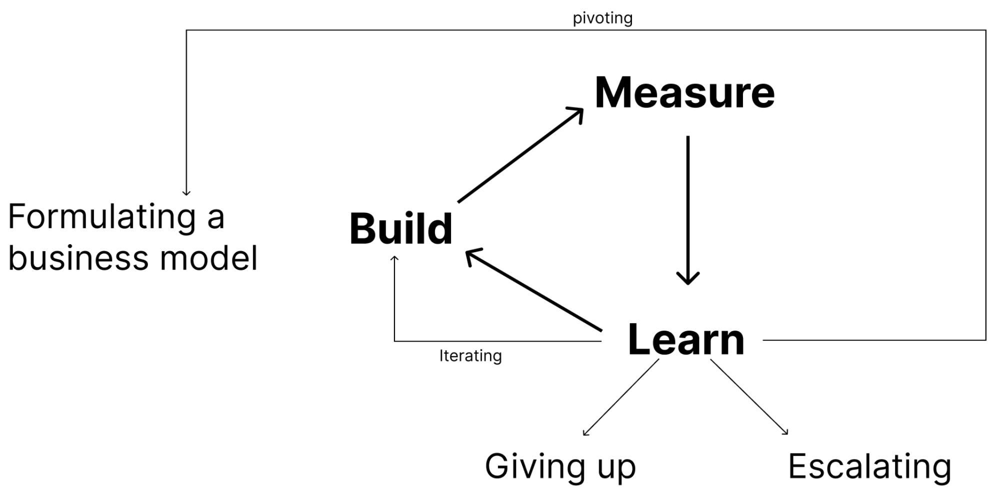

Figure 2.2: Lean Startup (LS) approach

It is worth mentioning that Olsen [2015] developed the Lean Product Playbook on the basis of the Lean Startup to provide a proper step-by-step framework to get make sure customers get enough value from the product. Similarly to Design Thinking, which will be presented in Subsection 2.2.1, prior to putting in place the Build-Measure-Learn loop from the Lean Startup, the target customer as well as its needs should be determined and validated.

Design Thinking

How designers think has been a topic of research in the design field since a long time [Georgiev, 2012]. Topic was introduced in the late '90s by authors like Goldschmidt [1994]. The underlying principle is to start from identifying and studying user needs to design suitable solutions. Design thinking utilises user research, iterations and feedback loops to come up with a product that is vauable to users [Brown, 2008]. There are multiple approaches to design thinking as Schallmo et al. [2018] states.

Basically this is the process that Design Thinking prescribes [Schallmo et al., 2018].

- **Determine design challenge**: once the area of the topic is clear the design challenge is formulated and the end users are defined
- **Understanding the challenge**: users are analysed and the effort is conveyed towards getting all the context needed. Means such as data analysis, interviews and surveys are used to gain the underlying knowledge.
- **Delineate perspectives**: specific use cases and situations are analysed to make sure the challenge is well scoped
- **Ideation**: ideas developed during the precious phases are described and assessed
- **Development**: prototypes that solves the identified challenge are developed
- **Testing**: different prototypes are tested with users and on the basis of the experience one or some of them are preferred
- **Integration**: prototypes are improved and integrated in a business model

Mind that the connection between each activity is not linear but iterative [Mueller and Thoring, 2012]. This approach can be effectively used in many fields Dorst [2011] and it is similar to the Lean Startup given its user-centric and iterative approach [Mueller and Thoring, 2012].

2.2.2 Technical Practices

Besides the need for continuous quick experiments and iterations explained in the previous Subsection 2.2.1, engineering practices should be optimised in order to be able to deliver good quality code. Product teams could benefit from *Agile* practices given that these are associated with better overall code quality, more control on the team's performances and better resource exploitation [Conboy, 2009]. Specifically, *Agile* engineering methodologies propose a more lightweight, flexible and user oriented approach compared to plan-driven software development which is less suitable for start-ups Blank and Dorf [2012].

However, Klotins et al. [2021b] argue that during early stages start-ups mostly rely on an *ad hoc* approach to engineering and operations given their focus on speed-related actions such as quick iterations rather than quality-related ones.

2.2.3 Reasons of failure

Acknowledging, studying and understanding failures is crucial for software startups Bajwa et al. [2017]. Given the constraints that are typical of startups, even one single failed project can put one of these companies out of business Giardino et al. [2014b]. It is not a surprise that over 90% of them fail, and mainly not because of competition. Blank [2007] states that most software startups fail because of a lack of customers and not for technological reasons. Among the possible reasons, the lack of *problem-solution fit*, prior to conveying energies towards marketing the product, stands out [Giardino et al., 2014b]. Giardino et al. [2014a] state that this *fit* is achieved when the product continuously manages to solve a problem that users encounter.

Having a quantitative measure of such a fit is not an easy thing though but following Sayyed-Alikhani et al. [2021] work, user retention is a good indicator of *problem-solution fit*. This because a user only continues to use (and pay) for a product if it manages to deliver value by solving one or multiple problems, over and over again.

When *problem-solution fit* is not found, a *pivot* should take place. This is not a simple iteration but more of a change of a concept or element of the business model currently adopted as mentioned in Subsection 2.2.1. *Pivots* can be of different types as reported in Ries [2011] but the most common ones according to Bajwa et al. [2017] are the *Customer need pivot* and *Customer segment* ones. These findings are coherent with the lack of *problem-solution fit*: the product is not useful for a particular set of customers so the startup can either change the product or the targeted audience.

Ries [2011] introduced in his book ten different types of pivots that can take place:

- *Customer Need Pivot*: occurs when the product is solving a need that the user does not experience in reality
- *Customer Segment Pivot*: changing from one customer segment to a different one, e.g. app for sales teams switches to customer support
- *Engine of Growth Pivot*: the growth strategy changes. For example abandoning paid ads in favor of content market ot search engine optimization (SEO)
- *Channel Pivot*: it is a recognition that a startup company has identified a way to reach their customers more effective than their previous one, e.g., switching from relying on a sales division to a self-served approach
- *Technology Pivot*: a startup delivers a similar solution by using totally different technology, e.g. transitioning from web application to native application
- *Zoom-in Pivot*: a feature may become a whole product. For example, sharing feature of an photo editing app can become the core of the app
- *Zoom-out Pivot*: opposite to zoom-in pivot, a product becomes a single feature of a bigger product. For example the API that allows to send information become a single feature of a CRM
- *Platform Pivot*: transitioning from an application to its underlying platform or the other way around, e.g., shifting from an online store to a platform that hosts online stores
- *Business Architecture Pivot*: in this pivot, a company switches business model e.g. going for many logos but low LTV, to few logos with a high LTV

- *Value Capture Pivot*: changing the way to capture the value. Switching from commissions on transactions to a fixed monthly rate

To summarize, a major cause of failure in startups is the lack of a structured process to discover and understand their markets, identify their customers and validate their hypotheses in the early stages of design [Trimi and Berbegal-Mirabent, 2012]. In short, the main reason of failure of young software startups is the fact of not being able to solve an on-going problem for users. This is further confirmed by Marmer [2011] that, after analysing 3200 high growth software startups, claimed that more than 70% failed due to premature scaling which essentially means that these companies have allocated too many resources to solve problems that are not that relevant for users.

2.3 Information Monitoring

In this section we present what information monitoring is, how it works and what companies can achieve by using it. With standard web technologies it is possible to monitor, in a detailed way, user interactions [Atterer et al., 2006]. Specifically, an HTTP proxy changes the HTML pages by inserting JavaScript code before displaying them to the client. This JavaScript tracking code gathers data about the various inputs that users give. This data is a record of behaviours performed by entities that is very useful to describe trends and habits. Event data is formed by an action, a timestamp and the state. This kind of information can be used to evaluate usability, interactions or user tests Atterer et al. [2006].

According to Münch [2014], software companies should evolve their development by deploying continuously new integration on the basis of customer feedback, which ultimately is the same of the *BML* loop [Brecht et al., 2021]. This continuous experimentation through constant deployments needs an effective infrastructure that allows the team to measure performance of each experiment, and learn from it. This can be done by gathering data through embedded systems [Münch, 2014]. In this way it is extremely easier to involve users and customers in the implementation of a software. Latency of customer feedback is now reduced and therefore assumptions can be evaluated more easily [Yaman et al., 2020]. This data can also be leveraged as a basis for a pivot. In line with this, data-driven decisions are said to be particularly suitable to prove a chosen direction or hypothesis is wrong rather than for coming up with a new one [Hirvikoski, 2014].

Companies with data-driven culture can support management in their decisions (DDDM), making them more objective, and therefore achieving higher effectiveness according to Troisi et al. [2020]. Still, during early stages, the volume of data is pretty low so analytics could be misleading. For this reason intuitions and data should be combined according to Hirvikoski [2014]. Data analytics helps companies increase customization and improve communication which leads to better relationship with users and higher customer satisfaction[Zhang et al., 2020]. Moreover it can also be helpful to understand possible profitable opportunities. However, from the customer side, concerns regarding ethics and privacy are rising as Kopalle et al. [2022] states.

2.3.1 Metrics

Metrics about user behaviour in the context of software startups, is not a well covered topic in academic literature. With metrics we refer to quantitative measurements of a phenomenon or object.

Still, Kemell et al. [2020] found out that 41% of 4700 software start-ups are not yet using metrics to steer their business because they feel it is too early. Moreover, about 10% does not track metrics because they believe it would not benefit them or because of a lack of resources. This finding is quite controversial given that many studies underline the importance of measuring (within the Build-Measure-Learn loop) the impact of new releases and of experiments [Blank and Dorf, 2012, Duc and Abrahamsson, 2016, Klotins et al., 2021c]. Kemell et al. [2018] by presenting an extensive list of metrics for software start-ups, argued that utilising such forms of product usage data could help start-ups to take strategic decisions.

Software startups can combine different types of metrics. Traditional business metrics, startup-specific business metrics, and software-related metrics including website metrics can be tracked [Kemell et al., 2018]. Software engineering metrics can be divided into product metrics and process metrics [Kupiainen et al., 2015]. Process metrics are metrics related to the process of creating software or maintaining it during its lifetime, which is typically done in the context of the *Agile* methodologies discussed in Subsection 2.2.2. On the other hand, product metrics are metrics related to product usage. Product metrics may also include usability-related metrics. Process metrics, on the other hand, take into account practice-specific metrics. The types of metrics measure and taken into considerations varies depending on the current stage the startup is in [Wang et al., 2016].

In this study we will focus mainly on product metrics, the ones related to understand user behaviour and product usage, rather than on process metrics. This because these kind of metrics suit more the research questions of this work.

Traditional business metrics such as Net Promoter Score (NPS) have been extesively studied [Lee], but these are not particularly useful for early-stage companies given that the product, most of the times is still in development and there are only a few users.

Kemell et al. [2018] after carrying out an extensive, still not comprehensive, literature review, states that practitioners agree that the most interesting metrics for software startups are:

- User retention and churn metrics
- User activity and engagement metrics
- Financial metrics about cash burn and development
- User-focused financial metrics such as lifetime value (LTV)

The retention rate describes the percentage of users that stay and use the product again over a certain period of time [Münch, 2014]. During early stages Pfeifer [2005], Sayyed-Alikhani et al. [2021] argue that it is better to focus on user acquisition only when user retention is at a good level. This is because high user retention is a good indicator of good problem/solution fit as discussed in Subsection 2.2.3. Ries [2011] also states the importance for young software ventures of measuring and improving user retention.

Nasir [2017] claims that customer retention is crucial since the most valuable asset of a company are loyal customers. Given the need of quickly acquiring users and run experiments, as discussed in Subsection 2.2.1, many software startups adopt a *Freemium* business model. This business model basically gives away some services for free and some other value-adding features for a premium price [Josimovski et al., 2019]. In this context, which is very different from one-off transactions, user retention becomes a fundamental indicator of performace[Ross, 2018]. Retention is also crucial when companies start making revenue as Murphy [2002] states. In his book Murphy [2002] shows some impressive statistics:

- Profitability on customers increases over the lifetime of a retained customer given that their trust increases
- 80% of a company future profits will come from only 20% of its existing customers
- A 2% increase in customer retention has the same effect of lowering costs by 10%

2.4 Literature gap

From the above literature review a clear overview of the following topics is provided:

- Main software startup strategies that are currently pursued
- An overview of the main reasons of failure
- How information monitoring works and how this is related to the main metrics and in particular user retention

These aspects are the foundation of this study and the outcomes will need to properly fit in this current state of things.

As the above literature review shows us, user retention is a key indicator of the success of a software product, as it reflects the extent to which users continue to use and engage with the product over time. High user retention rates can lead to increased revenue and growth for a software startup, as retained users are more likely to make purchases and recommend the product to others. On the other hand, low user retention rates can be a sign of problems with the product, and can ultimately lead to the failure of the startup if not addressed.

Given the importance of user retention, it is essential for software startups to identify and address any issues that may be causing users to leave the product. One way to do this is through the use of product usage data monitoring, which can provide insight into how users are interacting with the product and help to identify any potential issues. However, there is currently a lack of guidance on how to effectively use product metrics to improve user retention in software startups. While there are many different product metrics that can be tracked, it is unclear which ones are most relevant and how to use them to identify and address retention issues.

This knowledge gap represents a significant opportunity for further research and investigation. By exploring the use of product metrics to improve user retention in software startups, this work could start to fill this gap in the knowledge base and provide valuable guidance to startups looking to improve their chances of success.

Understanding how to use product metrics to improve user retention in software startups is an important step towards increasing the chances of success for these startups. By integrating the learning into existing startup strategies such as the Build-Measure-Learn cycle and Design Thinking, startups can more effectively identify and address retention issues as they arise, and improve their chances of success without having to change their main modus operandi.

3 Methodology

The methodology chapter is an essential part of any research project as it lays out the procedures and techniques used to conduct the study. In this chapter, we will outline the research design, data collection, and data analysis methods used in this study. The research design chosen for this study is a single-case exploratory study, which is suitable for the goal of understanding how software startups can increase user retention by means of product usage data monitoring. The study will be based on secondary data collected from a modern B2B SaaS startup, which is currently using information monitoring to increase user retention. This approach allows us to have access to exclusive information that is normally difficult to find in literature.

3.1 Research

Research methods and approaches available to undertake studies in management disciplines are multiple. Some came to light quite recently, and therefore have been utilised less, whereas others were accepted by the academia a long time ago. To make sure the approach used for this research is clear, the framework introduced by Ricciardi and Rossignoli [2015] is adopted to describe this work.

Contribution	**Research design**	**Information Elaboration**	**Information Gathering**
Context-Specific Description	*Case study*	*Direct data Analysis*	*Secondary data collection*
In-depth description of a case in its context. The interest can come from: (i) newness of the described event and/or (ii) the event hits with existing theories.	The focus is on the nature of the case(s) considered. Multiple information and studies can be carried out	Analysis of simple tables and data sets without statistics tools performing intuitive calculations	Structured data is utilised (e.g company databases, official statistics, pubblications)

Table 3.1: Categorisation following Ricciardi and Rossignoli [2015] framework

In Section 1.4 of the previous Chapter, the structure of the paper is presented. In this Chapter the methods used for the research are presented in detail. Specifically the framework presented in Section 3.1 will be discussed in detail.

3.2 Case Study

Case studies are a suitable methodology for investigating complex systems such as software start-ups, which are influenced by multiple factors that may impact user retention. This is due to the flexible nature of case studies and their focus on examining a particular phenomenon in a specific context Cruzes et al. [2015]. Given that retention could be influenced by many different factors depending on contextual envirnmoent, it seemed a rational choice for such a topic. This type of methodology is not based on a specific sampling logic to ensure representativeness but rather cases are chosen for their specific attributes, such as being representative, crucial, illuminating, or uncommon in some aspect[Cruzes et al., 2015]. Moreover, in line with the previous section, data exploration in a case study is conducted in its real context of use and not in an artificial environment [Zainal, 2007].

There are multiple typologies of case studies [Runeson and Höst, 2008, Zainal, 2007].

- *Explanatory:* approach aimed at explaining a situation or a problem
- *Descriptive:* approach aimed at reporting a phenomenon or a situation
- *Exploratory:* approach aimed at discovering new insights that could lead to the formulation of an hypothesis
- *Improving:* approach aimed at enhancing a certain aspect of a known phenomenon

An *exploratory* case study is the best choice for this thesis as it aims to explore and understand a new phenomenon, which is what the thesis aims to do. The use of a single-case exploratory study allows the researcher to investigate the topic in depth and to gain an in-depth understanding of the phenomenon under investigation in this specific context of the B2B SaaS startup.

It is indeed important to underline that a case study will not provide statistical significant conclusions but rather different kinds of evidence connected together to support a relevant conclusion [Runeson and Höst, 2008]. This study could be the precursor to a formal, large-scale research project. In general the goal of this kind of exploratory study is to prove that further investigation is necessary.

According to Yin [2011], case studies can be either holistic or embedded, depending on the context and goals of the study. In this work a single-case holistic exploratory study analysis is chosen given that an in-depth examination of a complex phenomena in its completeness is needed. Following Mir and Jain [2017]'s work, the rationale to legitimate a single case study should meet at least one of the following conditions:

- The case is an uncommon phenomenon - extreme
- Researchers could not access the case before - revelatory
- The case can be observed longitudinally - longitudinal

In this study the condition that allows us to undertake a single case study is the second one. Specifically, during the last months I have been working in a software start-up which is specialized in product analytics. This gives me the unique opportunity to access and use, thanks to the founder's kindness, all the data regarding product usage of a software start-up that is currently operating in the market.

3.2.1 Validity

The contribution of a study to the academic knowledge lies in its validity. According to Solingen and Berghout [1999], Yin [2011], there are mainly four aspects of validity:

- *Internal validity*: this aspect should be considered especially when causal relationships between variables are investigated. In this particular case the relationship investigated is not a causal one. Instead, third factors such as mediating and moderating variables are taken into account while investigating a correlational relationship between the dependent and independent variables and therefore internal validity is ensured.
- *External validity*: this aspect is about the degree of generalisability of the findings. Being a single case study this aspect is more concerning for this research. Further details will be discussed in Section 5.2
- *Reliability*: this aspect is about how the data and the analisys are dependent on the specific researchers. Basically the research should be easily reproducible. This aspect is ensured by a precise and well documented description of the research method used

3.2.2 Process

The process of conducting a case study is composed by steps that are similar to other research methods. However, the main difference is that case studies are flexible in nature so iterations between steps is common [Runeson and Höst, 2008]. Still, between iterations the objectives (understand how monitoring metrics could help to increase retention) should remain the fixed. If not, a different study is started.

Five are the main steps that should be followed when conducting a case study [Runeson and Höst, 2008]:

- *Design of the case study*: objectives and plan of the case are defined
- *Preparation for collection of data*: methods and protocols for gathering data are defined
- *Data collection*: execution, data gathering from the selected case(s)
- *Data analysis*: data is studied following a certain model
- *Reporting*: findings are organised and displayed

3.2.3 Case description

The focus of this case study is a B2B SaaS startup, established in 2021, that delivers product analytics solutions to other businesses. Founded by a skilled software engineer and an experienced product manager, this seed-stage startup has recently introduced its software product to the market. The platform offers businesses a comprehensive solution to gather, analyze, and visualize data related to their product usage, enabling them to make informed decisions based on real-time insights.

The company's journey began when the founders managed to secure funding from business angels, which allowed them to recruit an additional software engineer to help develop their minimum viable product (MVP). This crucial milestone led to their acceptance into Y

Combinator, a prestigious startup accelerator known for fostering the growth of numerous successful companies. Participation in the Y Combinator program not only provided the startup with valuable guidance and resources but also served as a springboard to attract further investment.

Capitalizing on this momentum, the company successfully raised seed-stage funding from investors, which fueled their growth and expansion. Subsequently, the team added another software engineer and welcomed a part-time member, who joined alongside an additional engineer to further strengthen their technical capabilities. Six months later, the startup expanded its team once more, bringing on board a talented product designer to enhance the platform's user experience and aesthetics.

Today, the company boasts a diverse and dedicated team of seven professionals, working collectively towards the shared goal of revolutionizing the product analytics landscape. With their innovative platform, incredible ease of use, and commitment to delivering valuable insights to businesses of all sizes, this startup is well-positioned for success in the rapidly evolving world of data-driven decision-making.

At the heart of this analytics platform is its extraordinary ease of use, designed to enable businesses to explore their product analytics without the need for long, expensive training sessions. This user-friendly approach sets the platform apart from traditional product analytics tools, which often require extensive training and significant technical expertise to be fully exploited.

The primary objective of this platform is to democratize product usage data monitoring, making it accessible and user-friendly for every team member, regardless of their technical background or experience. This inclusive approach enables all team members to gain a comprehensive understanding of crucial priorities, which in turn facilitates collaboration and data-driven decision-making. To achieve this, the platform offers a versatile range of metrics at both the user and company level, allowing businesses to evaluate their products with varying degrees of granularity.

Unlike many other analytics tools on the market that require users to have considerable knowledge and expertise to build custom metrics, this platform simplifies the process by providing ready-made templates. These templates act as a guiding framework, enabling users to extract valuable insights with minimal experience or technical know-how. Traditional analytics tools often involve a steep learning curve, requiring users to master not only the software but also the underlying concepts and techniques to create relevant and meaningful metrics. This can pose a significant challenge for businesses, particularly those with limited resources or expertise.

Addressing these challenges, the platform offers a set of ready-made templates designed to cater to a wide range of common business scenarios and objectives. These user-centric templates ensure that even those with minimal experience in product usage data monitoring can effectively leverage the platform's capabilities and derive meaningful insights from their data. By utilizing these templates, users can efficiently identify key performance indicators (KPIs), track trends, and monitor the impact of various initiatives without being bogged down by the complexities typically associated with creating custom metrics from scratch.

By removing barriers to entry for product analytics and fostering a more inclusive and collaborative environment, this platform revolutionizes the way businesses approach data-driven decision-making. With the ability to analyze performance at multiple levels, companies can make well-informed decisions that drive growth, optimize user experience, and ensure their products remain competitive in a rapidly evolving market.

The platform relies on event data to function effectively, and various degrees of data connection are required to unleash its full potential. The process begins with identifying users and companies to track events, allowing for a comprehensive understanding of product usage and user behavior. Although the platform does not offer real-time data, it excels in processing data efficiently and generating valuable insights for businesses.

Data is updated regularly, ensuring that businesses have access to the most recent performance metrics and can make informed, data-driven decisions accordingly. The platform's ability to process vast amounts of data quickly and accurately is crucial for businesses looking to adapt and respond to constantly changing market conditions.

The different levels of data connection start with basic identification of users and companies, enabling the platform to monitor user interactions with the product. As more data is collected, the platform can track a wide variety of events, such as user engagement, feature usage, and user retention. This granular data helps businesses understand how different aspects of their product or service are performing, allowing them to optimize their offerings and enhance the user experience.

Moreover, the platform's versatility extends to handling more complex data connections, which may include integrating with other data sources or third-party applications. By facilitating seamless data integration, the platform ensures that businesses have a holistic view of their product performance and can identify areas for improvement more effectively.

The quality of the data source plays a crucial role in unlocking the full potential of this analytics platform, allowing users to benefit from a more extensive set of features. A better data source not only improves the accuracy and relevance of insights generated but also enhances the platform's ability to deliver deeper and more comprehensive analyses.

When businesses provide high-quality, structured data, they enable the platform to perform advanced analysis, such as segmentation, cohort analysis and audience comparisons. These features empower businesses to identify trends and patterns that may not be apparent with less sophisticated data sources, ultimately helping them make better decisions.

In addition, a robust data source allows the platform to integrate seamlessly with other tools and systems, such as customer relationship management (CRM) software or marketing automation tools. This integration enables businesses to leverage the platform's analytics capabilities in tandem with other tools, further enhancing the effectiveness of their data-driven decision-making processes.

In the first section of the Appendix 2, a detailed explanation of what behavioural data is and how it is gathered, is given. This is indeed a critical step, without connecting a data source, users are not eligible to properly use the service and get value from it.

Product analytics plays a vital role in various aspects of a company, with different team members leveraging it to make informed decisions and optimize product performance. In early-stage startups, founders and product managers are often the primary stakeholders using product analytics to drive strategy and product development. For this reason the company object of the study focuses on these profiles. Also, as seen in Chapter 2, while trying to achieve *problem-solution fit*, which is the goal of the company that is being studied, early-stage startups should focus on a small set of problems which, in most cases, are interest only of a limited set of people. As mentioned in this case, the product is supposed to be actively used primarily by Founders and Product Managers of early-stage (from seed stage to series B for venture-backed companies) B2B SaaS companies with less than 100 employees.

Founders rely on product analytics to gain a complete understanding of the product's performance, user behavior and engagement. These insights help them make informed decisions about the company's direction, resource allocation, and overall strategy. Additionally, founders can use product analytics to monitor the product-market fit and validate their hypotheses about customer needs and preferences.

Product managers utilize product analytics to guide the product roadmap, prioritize features, and identify areas for improvement. By analyzing user behavior and engagement, product managers can better understand what users value most in the product and make data-driven decisions on what to build, enhance, or remove. This enables them to create a more user-centric product and align the development process with the company's strategic goals.

Software engineers also benefit from product analytics by measuring the adoption and impact of their latest deployments. By monitoring key performance indicators (KPIs) and user engagement metrics, engineers can identify and address any issues or bottlenecks in the product's functionality or performance. This ensures that the product remains stable, efficient, and meets user expectations.

In early-stage startups, professional data analysts may not be feasible due to limited resources and a smaller user base. Instead, founders, product managers, and engineers often take on the responsibility of analyzing product data and making data-driven decisions. As the startup grows and the user base expands, the need for more specialized roles, such as data analysts or data scientists, may arise to handle the increasing complexity and volume of data.

Collaboration within the analytics platform is crucial for businesses, as it fosters a unified approach to decision-making and helps teams align their strategies and goals. For this reason, the company object of this study has built numerous features around this concept, ensuring that teams can effectively work together to derive insights from the data and make informed decisions.

One of the key reasons collaboration within the tool is essential is that it encourages a shared understanding of product performance and user behavior. By providing a common platform for team members to access, analyze, and discuss data, the analytics tool helps bridge gaps in knowledge and expertise. This collective understanding enables teams to identify trends, patterns, and areas for improvement, ultimately leading to better-informed decisions and a more cohesive product strategy. Moreover, collaboration within the platform promotes transparency and accountability among team members. When everyone has access to the same data and insights, it becomes easier to track progress, measure the impact of initiatives, and hold each other accountable for the outcomes. This shared responsibility fosters a sense of ownership and encourages teams to work together towards common objectives.

The analytics platform offers a variety of collaborative features designed to enhance teamwork and streamline communication among team members. Two notable examples include Slack integration and the ability to write and share queries within the workspace.

- Slack Integration: The platform's integration with Slack, a popular team communication tool, allows users to set up daily, weekly, or monthly metrics digests. This feature enables team members to receive regular updates on key performance indicators (KPIs) and other important metrics directly within their Slack channels. By bringing these insights into a familiar communication space, the platform promotes a more informed and data-driven team culture. Additionally, this integration encourages discussions around the data, making it easier for teams to share their observations, ask questions, and collaborate on strategies for improvement.

- Query Sharing in the Workspace: The platform allows users to write custom queries to dive deeper into their data and extract specific insights relevant to their goals. These queries can be easily shared within the workspace, allowing team members to collaborate on their analyses and build upon each other's work. By providing a shared space for creating and refining queries, the platform fosters collective learning and helps teams unlock the full potential of their data. This feature also promotes knowledge transfer and empowers team members to learn from each other's expertise, ultimately enhancing the overall effectiveness of the analytics process.

At last, every startup ideally wants its users to perform a certain set of actions. For the startups which is object of this case study, the value proposition is around reports on metrics so it is very important that users create new reports and check the ones that have already been created.

Opening reports is therefore a crucial aspect of the startup's value proposition, as it allows users to access and review the insights and metrics generated by the analytics platform. By encouraging users to regularly open and examine reports, the startup ensures that its users are leveraging the full potential of the platform and making informed, data-driven decisions. Constantly checking reports enables the product team to stay up to date with the latest trends, patterns, and performance indicators. By reviewing these reports, users can gain a deeper understanding of user behavior, product engagement, and overall performance. This knowledge allows them to identify areas of improvement, prioritize features, and make strategic decisions that lead to a better product experience for their users.

Furthermore, the act of opening reports fosters collaboration among the team members. As users share and discuss the insights from these reports, they can collectively brainstorm solutions, test hypotheses, and refine their understanding of the data. This collaborative approach not only promotes knowledge sharing but also encourages a sense of shared ownership and responsibility for the product's success.

Objectives

The objective of this case study is to examine, within the context of a modern startup, the impact of data connection quality (user eligibility) on user retention, primarily through the effect of user engagement (SQ1), and to investigate the influence of user roles (user traits) and collaboration among teammates on metrics (collaboration) on user retention (SQ2).

In the context of this analytics startup, user eligibility refers to the quality of data connection. A higher quality data connection allows users to access more accurate and comprehensive insights, which in turn, could lead to increased engagement with the platform.

SQ1: User engagement could be a key factor in driving user retention, as engaged users are more likely to continue using the platform and derive value from the analytics insights. Improved data connection quality enhances user eligibility, which in turn should enable users to engage more effectively with the platform and potentially positively influencing user retention. The case study aims to explore this relationship, analysing how data connection quality can impact engagement and, consequently, retention.

SQ2: User roles, such as product managers, engineers, or founders, can also play a role in regards to user retention. Users with different roles may engage with the platform differently, which could influence retention rates. For example, product managers might focus on identifying trends and prioritizing features, while engineers may use the platform to measure the adoption of recent deployments. The case study aims to examine the effect of user roles on retention. Collaboration among teammates on metrics is another factor that can influence user retention. As the platform emphasizes collaboration, users who actively share reports and insights, and work together with their team members to analyze and act on the data, should be more likely to retain. The case study aims to investigate how collaboration on metrics can strengthen retention, highlighting the importance of fostering a collaborative environment within the organization.

By exploring these relationships in a real-world context, the case study provides valuable insights into the complex interplay between data connection quality, user engagement, user roles, and collaboration on metrics. These insights can inform the startup's strategies for enhancing the user experience, optimizing the platform, and ultimately driving long-term success in the market.

3.2.4 Preparation for data collection

User retention

In this subsection, the concept of user retention is presented.

User retention definition: percentage of users that are active at a certain point in time after they have signed up or installed the software.

`Retention = active users in a certain time / total number of registered users`

The second term is easy to determine whereas the first one is a bit more complex. Typically a software product has an ideal frequency of use. Airbnb, for example, is used ideally only a couple of times a year. Instead, a tool to calculate taxes has a monthly ideal frequency, Uber has a weekly one and WhatsApp is supposed to be used daily. This means that Airbnb considers a user retained even if a user was active only once in a year. For this reason the retention calculation should take into account the ideal frequency of usage. The second factor is how to define a user active. In the SaaS (Software as a Service) industry a user is considered active if the main value proposition of the software is exploited. For Airbnb this value could be either registering a new apartment / room or book a stay. For WhatsApp it would be sending a message.

For the company which is object of the case study the ideal frequency of usage is weekly and a user is considered active if a report is opened. Moreover, in most cases retention is calculated on single cohorts and not overall. In this case the formula becomes the following:

`Retention = active users in a certain time / total number of registered users in the same period of time`

Mind that the measure of retention on a single user can only assume two values: 0 (not retained) and 1 (retained).

DATE	USERS	NUMBER OF WEEKS LATER YOUR USERS WERE RETAINED				
		0	1	2	3	4
May 30-05 2022	34	100%				
May 23-29 2022	40	100%	38%			
May 16-22 2022	40	100%	35%	20%		
May 09-15 2022	36	100%	42%	39%	31%	
May 02-08 2022	33	100%	27%	39%	33%	24%

Figure 3.1: Example of weekly cohorts retention

In this table, each row represents a weekly cohort of users who started using the product during a specific week (e.g., 2022-05-30). The first number (34 for example) is the number of users that registered for the product in that given week. The columns represent the subsequent calendar weeks after the users joined the product. The first column (Week 0) always shows 100% retention, because it represents the starting point for each cohort.

The percentage values in the table indicate the proportion of users from the respective cohort who are still active in the corresponding week. For example, in the second row, 38% of users were active during the second week after joining.

When the retention rate is 100%, it means that all users in that cohort are still active or using the product during the corresponding week. However, this percentage is typically seen only in the first column (Week 0) since it is the starting point for each cohort. As time progresses, it is common to see a decrease in the retention rate, as some users might drop off or become less active. The goal for businesses is to improve the retention rates over time, ensuring that users stay engaged with the product or service for a longer period.

Eligibility

Some products require some preliminary conditions to be used. For example a car requires a driving license to be used on the road. Similarly, users have to connect a data source to be eligible to get value from the software produced by the company which is object of the case. Data sources can be of different types and not all of them are compatible with the platform. Assuming that the source is compatible there are still different possible setup configurations. Users can have a very sophisticated system (tracking plan) that allows them to track any sort of information or a very simple one that only allows them to understand the web pages viewed by users. We, therefore, assume that eligibility is an important term to be considered while trying to increase user retention.

Moreover, users with a poor data setup may not be eligible to use certain features of the product, impacting user engagement. The degree of eligibility state is automatically measured using group traits. Traits are pieces of information about a user or an account that are included in an "identify call." These traits can be demographics like age or gender, account-specific information like plan type, or data setup quality. More information about "identify calls" and traits can be found in Appendix 2.2.

The eligibility variable can assume a combination of "true" or "false" values for the following elements:

- Tracking events: this refers to the process of recording user interactions with the product, such as clicks, page views, or specific actions. A "true" value indicates that the user's data setup allows for tracking events, while a "false" value indicates that tracking events are not properly set up. When the value is equal to "true" users are then able to measure interactions that user have with their software product and therefore understand how engaged they are with it.
- Identifying users: this involves assigning unique identifiers to each user, allowing the platform to distinguish between individual users and track their behavior over time. A "true" value means that the user's data setup is able to identify individual users, while a "false" value means that user identification is not properly configured. When the value is equal to "true" users are then able to recognise their customers and assign a certain behavior to the person that has performed it.

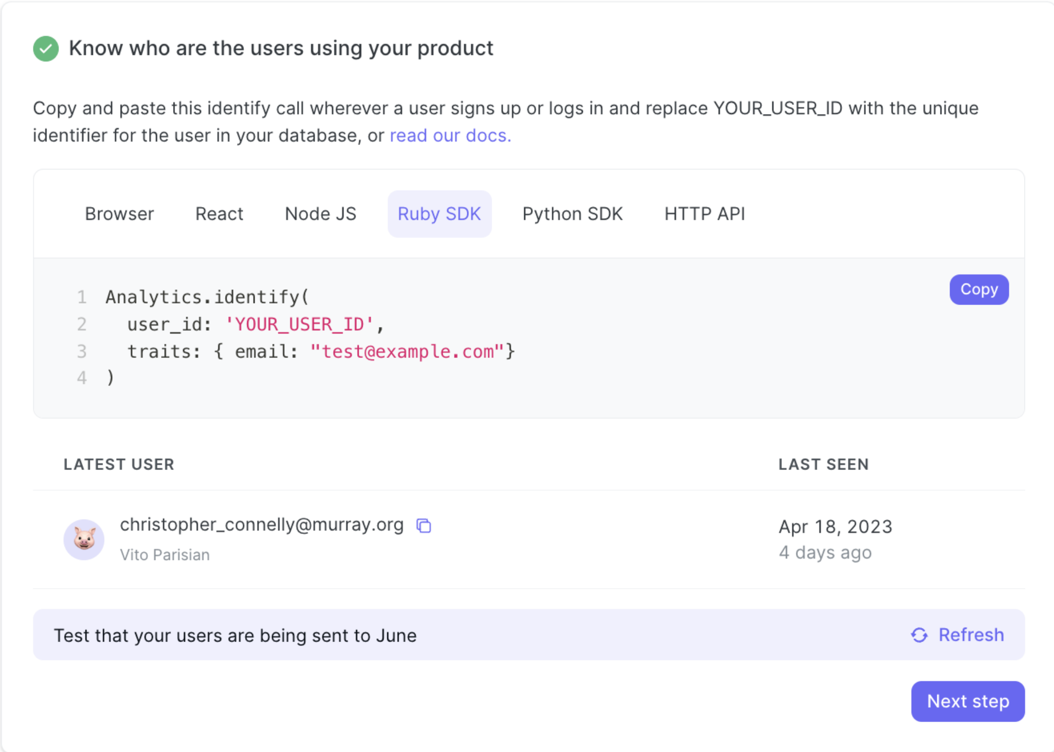

Figure 3.2: Eligibility check on tracking identifying users

- Identifying groups: this pertains to the process of categorizing users into groups or accounts, which can be useful for understanding user behavior at a more aggregate level. A "true" value signifies that the user's data setup allows for grouping users, while a "false" value indicates that group identification is not correctly set up. When the value is equal to "true" users are then able to associate customers into groups or accounts and evaluate behaviour comprehensively.

By evaluating the combination of these elements, the platform can determine the degree of eligibility for each user, which in turn influences user engagement and the overall effectiveness of the product.

User Engagement

To get value from the product, users should engage with its main features. With engagement, we mean actively using a certain feature. This behaviour can be measured using a track event as it was explained in the section above. There are multiple actions users can do within the product ranging from simply clicking a button to creating a proper report.

We therefore assume that user engagement is influenced by the eligibility variable given that some features can be used only with a certain state of setup.

This variable can be measured by simply assessing the raw count of track events performed by a certain user. The platform that is used to collect data also allows us to break down this number looking at single event which maps to a single feature and therefore behaviour. This enables us to choose the desired level of granularity.

User Traits and Collaboration

Still there are other factors to take into account when it comes to understand what drives retention. As mentioned above, the product of this company mainly gives to users an understanding of the overall health of the company and how their product is being used. In the tech industry these data points are mainly of interest for Product Managers and Founders of early stage companies. It is therefore reasonable to assume that retention will vary in the long run, depending on the kind of title. Moreover, metrics in a modern startup are shared across the organisation so we expect collaboration to play its role as well.

The relationship between engagement and retention could therefore be influenced by other variables such as the job title of the user and collaboration. Similarly to what Siddiqui [2016] argues, different user characteristics can have different impacts on engagement.

Traits are information regarding users such as the Role in the company. This can assume the following values: Product Manager, Founder, Engineer. The reasoning for including only these types of roles is explained in the Case Description 3.2.3.

3.3 Data Collection

3.3.1 Data sources

As previously mentioned, this case study utilizes secondary data derived from the Customer Data Platform (CDP), which collects behavioral data on users. To effectively visualize and analyze this data, we employ the software developed by the company being studied, a specialist in product analytics. The product analytics tool is designed to examine user behavior and interactions with digital products, such as websites and mobile apps, providing crucial insights into user behaviour.

The product analytics software offered by the company allows for the creation of customized metrics and the ability to narrow down the audience to specific groups of users. This powerful feature enables us to isolate and analyze behaviors and measure engagement and retention for different user groups, each with varying levels of eligibility. By segmenting users based on eligibility and other characteristics, we can identify patterns and trends that contribute to user retention.

To thoroughly investigate user retention, we chose to examine user behavior across month after signing up for the platform. This time frame serves as a strong indicator of user satisfaction and platform efficacy because users who continue to engage with the product after a month are more likely to be receiving value from the service. This one-month window also allows us to observe engagement and relate how different types of users interact with the platform as well as how collaboration relates to it.

The sample used for this study consists of all users who signed up between June 1st, 2022, and August 31st, 2022, totaling 1,305 individuals. By selecting the most recent and largest available sample, we increase the reliability and relevance of our analysis. This comprehensive dataset provides a solid foundation for identifying factors that influence user retention and engagement, as well as insights into the overall effectiveness of the product analytics tool.

To better understand the data, we have classified the sample users into three main interest groups based on their eligibility states, which are determined by the quality of their data connection. These groups are as follows:

- Group 1: Users with no data tracking or identification (tracking events = false, identifying users = false, identifying groups = false)
- Group 2: Users with data tracking and user identification, but no group identification (tracking events = true, identifying users = true, identifying groups = false)
- Group 3: Users with full data tracking and identification (tracking events = true, identifying users = true, identifying groups = true)

We have excluded the other three possible eligibility states from our analysis, as they are not applicable in real-world situations. Specifically, none of the 1,305 users in our sample have any of these ineligible states.

To collect data using the product analytics software, we followed these steps:

- Adjust eligibility parameters to cover the possible groups (Groups 1, 2, and 3).
- Measure user engagement and retention for each group.
- Introduce user traits and collaboration factors, and assess their impact on the retention of the three different interest groups.

By breaking down the sample into these distinct groups and analyzing their engagement and retention, we can better understand how different levels of data connection quality, user traits, and collaboration affect user retention.

3.3.2 User engagement measures

User engagement is measured by counting the number of a core action performed by users in the product in a fixed period of time (one month). In this case this action is captured by a track event called *loaded report*. The higher the number the higher the engagement. Measuring engagement taking into account only this action is conservative given that users can also perform a range of other different actions. The reasoning behind the choice of measuring engagement in this way can be found in the Case Description 3.2.3. Mind that the engagement will be measured only for the users that comply to the eligibility state.

Group 1

This is the scenario without any data source connected. Eligibility state is the lowest possible. Engagement, both cumulative and average, as well as retention are then measured using the software application and displayed in the table in the Appendix 4.1. The expectation is to have low user engagement compared to better eligibility states.

Group 2

This is the scenario with an intermediate setup where users are identified and actions are tracked through track events. However companies are not identified. This kind of eligibility state allow users to use most of the features available in the product. The expectation is to have good engagement and retention. Data points can be found in the Appendix 4.1.

Group 3

This is the scenario with the best setup where users and companies are identifying and also events are tracked. This eligibility state allows users to benefit from the full feature set the product offers. For this reason the expectation is to have excellent engagement and retention. Again, engagement and retention is calculated on the users that comply to the eligibility state as it can be seen in Table 4.1

3.3.3 User traits and collaboration measures

Taking into account the same three Groups mentioned earlier, we will now gather and present data related to user retention. This data will be analyzed based on whether or not users collaborate and the three possible user trait states discussed in the previous subsection 3.2.4.

To measure collaboration, we focus on a specific event within a workspace. We consider collaboration to have occurred if the variable "invited_a_teammate" has a value greater than zero, indicating that at least one teammate was invited to join the workspace.

Meanwhile, to assess how the role of the user influences retention, we apply a user trait filter on the audience. This allows us to isolate the effect of different user roles (such as founders, product managers, or product engineers) on retention rates.

By examining retention rates across varying levels of collaboration and different user roles, we can gain insights into the factors that contribute to long-term engagement with the product analytics tool. These insights can help the company identify areas for improvement and tailor its strategies to better address the needs of its diverse user base.

Group 1, 2 and 3 - Collaboration and user traits

In this section, we examine the same target audience as Group 1, 2 and 3 from the previous sections, which means the eligibility state remains consistent. However, here we investigate the effects of collaboration and user traits on retention.

The first tables in each Group in the Appendix 4.2 explores the impact of collaboration on retention. The table is divided into two rows: the first row represents users who collaborated, while the second row represents users who did not collaborate.

Retention is calculated for sub-group, allowing for comparison both within the group and across other groups. It is important to note that the total number of eligible users is equal to the total eligible users of the same Group in the previous sections.

The second table within the same in the Appendix investigates the influence of user traits on retention. The table consists of three rows, each representing a specific user trait within Group 1, 2 and 3:

- The first row considers users who are founders.
- The second row considers users who are product managers.
- The third row considers users who are engineers.

Retention is calculated for each subgroup to determine which user trait is associated with the highest retention. Again, the sum of eligible users corresponds to the total eligible users of each group of interest in the previous section.

By analyzing how different attribute influence user retention, we can better understand the factors that contribute to user retention and tailor strategies to meet the needs of different user groups.

4 Results

In this chapter, the gathered data is analysed, results are gathered and discussed.

4.1 Data analysis

4.1.1 Impact of Eligibility on Engagement and Retention

Based on the research design, the quantitative findings obtained in the earlier part of the study will be analysed. Based on the data provided in the Tables of the Appendix 4.1, it appears that the level of user engagement and user retention is positively influenced by the level of tracking and identification of events, users, and groups (eligibility).

In Group 1, where there is no data source connected, the user engagement is low, with only 776 eligible users and an average of 7.37 loaded reports per user. The user retention rate is also low, at 7.99%. This result resonates with the context provided about this particular case because without connecting any data there is not much users can do within the platform. Still the ones that retained are probably consuming the content that the company is publishing about metrics. It is fairly common that early stage startups do not start tracking user behavior before making any revenue which explains why there are so many eligible users in this group.

In Group 2, where events are tracked and users are identified but groups are not, there is an improvement in user engagement with 272 eligible users and an average of 17.64 loaded reports per user. The user retention rate is also higher, at 18.8%. This eligibility state is the most common among the users that do connect a data source. This setup can be used by both B2B (Business-to-Business) and B2C (Business-to-Consumer) platforms to get an understanding of how their users engage with the product. This eligibility state allows them to use most of the feature set of the platform and the technical setup is not that hard. This means that getting started with analytics is easier compared to Group 3.

In Group 3, with the best setup of tracking events, identifying users, and groups, user engagement is even higher, with 255 eligible users and an average of 18.1 loaded reports per user. However, user retention rate is slightly lower, at 16.8%. As already mentioned this is the most sophisticated eligibility state possible which is only relevant for B2B businesses who are selling their services to an organisation with multiple end users. The engagement level appears to be higher compared to Group 2 and this could be due to the fact that this eligibility state unlocks more features within the platform and therefore users can open more types of reports. The small drop in retention could be due to the fact that the features unlocked by group tracking are still under development and are not as consolidated as the user level ones.

This suggests that data quality, eligibility in this case, has a positive impact on user engagement, but there may be other factors at play that also impact user retention. On the basis of the data, it can be observed that in Group 3, both the number of loaded reports and average loaded reports per user are higher compared to Group 2, which indicates higher user engagement. However, the user retention in Group 3 (16.8%) is lower compared to Group 2 (18.8%). This means that while more users are actively engaging with the product in Group 3, a smaller proportion of those users are continuing to use the product in the long term. This could be due to various factors such as difficulty in using certain features of the product, short term value deriving from group level analytics or other factors.

Still, the difference in the retention values for Group 2 and Group 3 is not that big, whereas Group 1 is quite far below. This shows that a good eligibility state indeed leads to high engagement and retention. However, the company should not only pursue an increase in engagement, given that it does not always ensure better long-term retention. For this reason it is worth taking into account other factors to better understand how it possible to increase retention.

4.1.2 Impact of Traits and Collaboration on Retention

Examining the data from the case study, it becomes clear that collaboration and user traits play a crucial role in influencing user retention across the three distinct groups. These findings provide valuable insights into the dynamics of user behavior within the startup ecosystem and offer potential learnings for improving user retention.

In Group 1, where no data source is connected, collaborating users exhibit a notably higher retention rate of 68% compared to non-collaborating users with a meager 0.6% retention rate. Within this group, founders display the highest retention rate at 68.9%, followed by product managers at 12.6%, and engineers with no retention at all. This suggests that founders, who typically have a deeper interest in the success of the startup, are more likely to remain engaged with the product even in the absence of a comprehensive data tracking system. This might be due to the fact that the analytics platform provides a lot of valuable content around how to improve the main product usage metrics. Furthermore, the significant disparity in retention rates between collaborating and non-collaborating users underscores the importance of fostering a collaborative environment to increase user retention.

Similarly, in Group 2, which tracks events and identifies users but not groups, collaborating users again demonstrate a higher retention rate of 45% compared to their non-collaborating counterparts at 6.87%. In terms of user traits, founders maintain the highest retention rate of 67%, while product managers and engineers follow with retention rates of 7.35% and 5.83%, respectively. The data suggests that even with the addition of event tracking and user identification, collaboration remains a key factor in promoting user retention. Similarly across different roles, user retention patterns remain similar to Group 1.

Lastly, in Group 3, where events, users, and groups are all identified, collaborating users score a 54% retention rate, significantly outperforming non-collaborating users with a 5.86% retention rate. Here, product managers take the lead with a 57.14% retention rate, followed by founders at 50% and engineers at 6.15%. This indicates that when given the most comprehensive set of tools for tracking and identification, product managers, who are often responsible for driving product development and user experience, become the most engaged and retained users. Founders instead have a lower retention rates in this cae. This might be due to the fact that normally companies that identify groups are at a later stage and founders are not as hands on the product as they were during earlier stages.

In conclusion, the data analysis highlights that across all groups, user retention is positively influenced by collaboration and founders and product managers are by far more likely to retain compared to engineers. Encouraging collaboration and understanding the unique needs and motivations of different user roles, such as founders, product managers, and engineers, can contribute to the development of more effective strategies to enhance user retention for the startup in question.

4.1.3 Reporting

To summarize, the data shows that good tracking such as identification of events, users, and groups has a positive impact on user engagement and retention. The highest level of user engagement is seen in Group 3, where all three elements are tracked and identified, although the user retention rate is slightly lower compared to Group 2 which has a good, still not excellent eligibility state. This means that pursuing an increase of user engagement is not enough to achieve great user retention. Collaboration and the job role of the founder are positively associated with user retention across all three groups, with the highest retention rate for users who collaborate seen in Group 1. This suggests that having founders as users of the product can lead to higher user retention rates, compared to other job roles such as product managers and engineers.

4.2 Reflections on the findings

In the context of software startups, understanding and improving user retention is vital for business success. Our case study revealed that factors such as user eligibility, user traits, and collaboration significantly influence user engagement and retention. By examining these factors, companies can devise strategies to increase the likelihood of retaining users and promoting long-term success.

User eligibility, which refers to the quality of the data connected to the platform, plays a crucial role in user engagement and retention. If users are unable to connect a data source to the product, they may not derive much value from it and quickly lose interest. Consequently, low user engagement and user retention may hinder the business's success. To address this issue, the company should focus on simplifying the setup process and making the technical requirements more accessible. This would lead to better user eligibility states and potentially higher user engagement and retention.

User traits also have a significant impact on user engagement and retention. The case study found that founders consistently exhibited higher retention rates compared to product managers and engineers. This suggests that certain user traits, such as being a founder, are associated with higher user engagement and retention. Businesses can use this information to develop targeted go-to-market strategies and product iterations that will attract and retain users with these traits. By focusing on users who are more likely to be committed to using the product continuously, companies can increase the chances of building a strong and loyal user base.

Collaboration among users is another factor that can positively affect user engagement and retention. Users who collaborate with each other tend to have higher retention rates, possibly due to the sense of community and support collaboration fosters. A collaborative environment can lead to increased engagement with the product, as users interact, exchange ideas, and provide assistance to one another. This sense of community can encourage loyalty and increase the likelihood that users will continue to engage with the product over time, resulting in higher retention rates. To capitalize on the benefits of collaboration, the company should keep creating features that promote and facilitate interaction among users.

In conclusion, by addressing user eligibility, focusing on user traits, and promoting collaboration, this software startup was able to improve user retention. Understanding these factors and their impact on users is crucial for any company aiming to achieve long-term success in the competitive software startup landscape. By implementing strategies that cater to these factors, businesses can create a robust and loyal user base, driving growth and success over time.

5 Conclusion

In this study the literature regarding the main startup strategies and how those organisation currently use data has been reviewed. On this basis four *Research questions* were formulated. To conclude this study, we summarise the findings hereafter.

RQ: *How can software start-ups increase their user retention by means of product usage data monitoring?*

As highlighted in the literature review, increasing user retention is crucial for software startups. To achieve this by leveraging information monitoring, software startups should first identify the key variables that have the potential to impact user retention. These variables may include user engagement, eligibility, collaboration, or user traits, among others. Next, they should implement tracking and visualization methods for these identified variables to facilitate data-driven decision-making. By analyzing product usage data, startups can gain insights into the conditions under which users are more likely to be retained. This may involve identifying patterns in user behavior or exploring the relationship between specific variables and retention rates. Based on the analysis, software startups should prioritize and implement strategies that encourage actions and behaviors leading to higher retention. This could involve enhancing user onboarding, simplifying user experience, or offering personalized content and features. Lastly, software startups should tailor their acquisition and marketing efforts to attract the types of users that demonstrate the highest probability of retention. This targeted approach will contribute to the overall growth and sustainability of the startup.

SQ1: *How can user eligibility influence user retention through an effect of user engagement?*

The relationship between user eligibility, user engagement, and user retention was thoroughly examined in Section 3.3. As anticipated, user eligibility states have a significant influence on a user's ability to interact with the product, which in turn affects user engagement. The higher the eligibility state, the greater the level of user engagement. Interestingly, user retention does not follow the same trend in this case: users with the best eligibility state exhibit the highest engagement but not necessarily the highest retention. This observation suggests that there may be additional factors at play that influence user retention beyond eligibility and engagement. These findings highlight the importance of considering multiple aspects when trying to improve user retention in software startups. While eligibility and engagement are crucial components, it is necessary to evaluate and address other potential factors that may contribute to retaining users.

SQ2: *How can user traits and collaboration affect the relationship between user engagement and user retention?*

Data pertaining to user traits and collaboration was collected in Subsection 3.2.4 and subsequently analyzed. The analysis revealed that both factors significantly influence user retention. Specifically, one type of user was found to have a considerably higher likelihood of retention compared to others. Similarly, users who collaborated with each other demonstrated a higher retention rate. These findings contribute to a better understanding of the conditions under which user retention is more likely to be high. By examining the relationship between user traits, collaboration, and retention, companies can develop strategies to promote behaviors that lead to an increase in returning users. This includes tailoring their product offerings to the needs of specific user types or fostering a collaborative environment within the app, which ultimately serves to improve user experience and drive long-term customer loyalty. Additionally, the results suggest that a deeper understanding of user traits can help businesses identify which customer segments are more likely to retain and engage with the product. This information can be used to refine marketing efforts and target high-value customer segments, ultimately improving the effectiveness of customer acquisition strategies. Furthermore, the relationship between collaboration and user retention highlights the importance of promoting a sense of community and teamwork within the software. By creating features that facilitate collaboration, such as shared workspaces or communication tools, the company can encourage users to work together and stay engaged with the product. This not only enhances the user experience but also fosters a sense of loyalty and commitment to the platform.

SQ3: *How does continuous information monitoring integrate with the current software start-up strategies?*

After reviewing the primary startup strategies outlined in the literature and the state-of-the-art data collection techniques, it is evident that continuous information monitoring can be seamlessly integrated with these frameworks to gain deeper insights into customer behavior. By doing so, startups can effectively reduce the Build-Measure-Learn (BML) iteration loop time, which in turn allows them more time to identify a viable business model. Moreover, continuous information monitoring plays a crucial role in helping companies recognize when a pivot is necessary. By closely observing user engagement, retention, and other key metrics, startups can make informed decisions about whether to modify their product offering, target a different customer segment, or change their overall approach. This proactive approach, driven by real-time data, enables software startups to adapt quickly to the dynamic market landscape and better address their customers' needs, ultimately enhancing their chances of success.

5.1 Contribution

A context-specific description contribution, such as the one of this study, refers to research that is specifically designed and conducted within a particular context to yield more accurate and relevant findings. In this thesis, the context is a modern B2B SaaS startup, where the research question is centered on improving user retention through information monitoring. By studying the problem within its natural setting, the research can go deeper into the underlying issues and generate more precise insights.

The context-specific approach is valuable for several reasons:

- Real-world data: by conducting research in a B2B SaaS startup environment, the researcher can access exclusive and real-world data, which is typically unavailable in other studies. This enables a more comprehensive and authentic analysis of the problem.

- Practical recommendations: as the research is conducted within the specific context of a B2B SaaS startup, the recommendations derived from the study are tailored to address the unique challenges faced by such organizations. Consequently, these findings can be directly applied to similar settings, providing practical value to other B2B SaaS startups.
- Transferability: although the research is context-specific, the insights and recommendations obtained from this study may still be relevant and useful to other software startups or industries that face similar challenges in user retention and information monitoring.

5.1.1 Practical Implications and Societal Contribution

Our findings offer valuable insights for companies with similar characteristics to improve user retention. Based on the results, we recommend the following initiatives:

- Increase eligibility by simplifying the setup and making the technical requirements as easy as possible.
- Enhance in-app engagement by suggesting new workflows and follow-up actions.
- Adjust acquisition and go-to-market strategies to acquire the type of user that appears to be more successful.
- Boost collaboration within the app.

In our case, for example, increasing eligibility and collaboration is a safe bet to improve retention. Targeting marketing campaigns and user interviews to acquire more users with the characteristics positively associated with retention will also likely increase the number of returning users.

The practical implications of this study has considerable potential for generating positive societal impacts. By enhancing user retention for software startups, we can contribute to increasing the customer base available to these businesses. As a result, a higher number of startups could remain operational, leading to the sustenance and creation of new employment opportunities. This societal impact is of great significance, given the millions of new projects initiated every year, highlighting the vital role startups play in the labor market.

Moreover, higher user retention implies that the product consistently addresses a customer's needs and creates ongoing value for them. This not only leads to increased customer satisfaction but also promotes customer loyalty and long-term relationships between the business and its users. Such ongoing value creation is crucial for both the customers who benefit from effective solutions to their problems and the startups that can grow sustainably through a loyal customer base.

Furthermore, startups often drive innovation, leading to the development of new products, services, and technologies that can improve people's lives and well-being. By increasing user retention, software startups can continue to invest in research and development, leading to the creation of novel solutions that address pressing societal challenges. This, in turn, contributes to enhancing the overall quality of life for individuals and communities.

Additionally, thriving software startups can also contribute to the growth of local economies, as they often purchase goods and services from other businesses within their region. This stimulates economic activity and helps create a positive cycle of growth and prosperity.

In summary, our study not only benefits companies seeking to improve user retention but also plays a part in fostering broader economic growth, job creation, innovation, and ongoing value creation for customers. By understanding and addressing the factors that impact user retention, we can contribute to the success of software startups and, in turn, support their positive influence on society.

5.1.2 Academic Reflections and Scientific Contribution

This study contributes to the existing body of knowledge by addressing the current gap in understanding how software startups can increase their user retention. As an exploratory single case study, the unique access to data and the operational context of a B2B SaaS startup allows for a richer exploration of the topic. This research builds on and enriches the work of previous scholars who have explored startup strategies and product metrics.

By examining the relationships between data connection quality, user traits, collaboration, and user retention in software startups, we have shed light on the underlying dynamics that drive these key performance indicators. Furthermore, our findings confirm and expand upon the assertions made by previous researchers, such as the importance of user traits and engagement for user retention, as highlighted by Bansal and Pruthi [2021] and Siddiqui [2016].

In addition to validating existing theories, the research also extends the literature by applying these concepts to the specific context of software startups, where the nature of the product and the competitive landscape may differ from those in other industries. This allows for a more nuanced understanding of the factors that drive user retention in this unique setting.

Being an exploratory single case study, this thesis showcases the value of in-depth examination of a particular case, which can offer valuable insights and highlight trends that may not be as apparent in larger samples. By immersing ourselves in the operational context of the startup and leveraging the unique access to the data, a more comprehensive understanding of the factors affecting user retention in the B2B SaaS industry is provided.

In summary, this research contributes to the scientific community by advancing the understanding of user retention in software startups, validating and building upon existing theories, and showcasing innovative methodologies that can be further analysed in future studied. The exploratory nature of our study, combined with the unique access to data and operational context, allows for a more in-depth analysis of the subject, paving the way for further research in this area.

5.2 Limitations and Future Work

While our findings offer valuable insights, there are limitations to the generalizability and applicability of our results due to the nature of the single case study. However, it is important to note that the company at the center of this study shares several characteristics with many modern B2B SaaS early-stage startups. In this specific niche, our findings can still provide valuable insights. Our framework may be particularly relevant for:

- SaaS businesses built on top of another platform that needs to be connected to the service
- Products that require technical implementation
- Services that can be used by various roles within a company
- Products that can be used collaboratively by users

Another limitation of this study is its reliability. Data was gathered and analyzed directly using the company's software, which contains sensitive information. Access to this data was granted due to the relationship of trust that has developed between the founders and the researcher. Replicating this study would require full access to this data, which might not be easily attainable for external stakeholders. Therefore, a more extensive study with a larger sample and more accessible data would be beneficial for ensuring better external validity. Additionally, this study could be further enriched by incorporating a more quantitative approach to provide more comprehensive insights.

A further limitation lies in the small number of eligibility groups used in this study. This is mainly due to the fact that the relevant eligibility states for this particular company are limited to three. With this in mind, it would be interesting to replicate this study by conducting a single or multiple case studies focused on companies with a multitude of eligibility states. This would provide a more diverse range of insights and contribute to the robustness of the findings.

Building on the insights and constraints identified in this study, there are numerous opportunities for future research to further advance our understanding of user retention, specifically within the context of software startups.

1. Expanding the scope: Conduct research with a larger and more diverse sample of software startups to increase the generalizability of the findings. This may include startups from different industries, stages of growth, or with varying product offerings.
2. Longitudinal studies: Examine the long-term impact of the factors identified in this study on user retention, user engagement, customer lifetime value, and overall business performance. A longitudinal approach would provide insights into how these factors evolve over time and their cumulative effects on a company's success.
3. Multi-method approach: Employ a combination of qualitative and quantitative research methods to provide a more comprehensive understanding of user retention dynamics. This may include interviews, surveys, and advanced statistical analyses to validate and expand upon the findings of this study.
4. Comparative analysis: Conduct comparative analyses between successful and unsuccessful software startups to identify the factors that differentiate them in terms of user retention and overall performance. This would help determine best practices and common pitfalls for startups to consider when developing their retention strategies.
5. Impact of external factors: Explore the influence of external factors, such as market trends, technological advancements, and regulatory changes, on user retention dynamics in software startups. Understanding these factors could help companies adapt and respond to changes in their environment more effectively.

These suggestions for future research would contribute to a more comprehensive understanding of user retention in the context of software startups and provide valuable insights for practitioners aiming to improve their retention strategies.

Bibliography

Richard Atterer, Monika Wnuk, and Albrecht Schmidt. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web - WWW '06*, page 203, Edinburgh, Scotland, 2006. ACM Press. ISBN 978-1-59593-323-2. doi: 10.1145/1135777.1135811. URL http://portal.acm.org/citation.cfm?doid=1135777.1135811.

Sohaib Shahid Bajwa, Xiaofeng Wang, Anh Nguyen Duc, and Pekka Abrahamsson. "Failures" to be celebrated: an analysis of major pivots of software startups. *Empirical Software Engineering*, 22(5):2373–2408, October 2017. ISSN 1382-3256, 1573-7616. doi: 10.1007/s10664-016-9458-0. URL http://link.springer.com/10.1007/s10664-016-9458-0.

Ted Baker and Reed Nelson. Creating something from nothing: Resource construction through entrepreneurial bricolage. *Administrative Science Quarterly*, 50:329–366, 09 2005. doi: 10.2189/asqu.2005.50.3.329.

Rohit Bansal and Nishita Pruthi. Role of customer engagement in building customer loyalty-a literature review. *Samvakti Journal of Research in Business Management*, 2:1–7, 10 2021. doi: 10.46402/2021.01.9.

Richard Bednár and Natália Tarišková. Indicators of startup failure. 12 2017.

Steve Blank. *The four steps to the epiphany*. K & S Ranch, 5 edition, July 2007.

Steve Blank. Why the Lean Start-Up Changes Everything. page 12, 2013.

Steve Blank and Robert Dorf. *The startup owner's manual. 1: The step-by-step guide for building a great company*. K&S Ranch, Pescadero, Calif, 1. ed edition, 2012. ISBN 978-0-9849993-0-9.

Rafael Fazzi Bortolini, Marcelo Nogueira Cortimiglia, Angela de Moura Ferreira Danilevicz, and Antonio Ghezzi. Lean Startup: a comprehensive historical review. *Management Decision*, 59(8):1765–1783, August 2021. ISSN 0025-1747. doi: 10.1108/MD-07-2017-0663. URL https://www.emerald.com/insight/content/doi/10.1108/MD-07-2017-0663/full/html.

Patrick Brecht, Daniel Hendriks, Anja Stroebele, Carsten H. Hahn, and Ingmar Wolff. Discovery and validation of business models: How b2b startups can use business experiments. *Technology Innovation Management Review*, 11:17–31, 03/2021 2021. ISSN 1927-0321. doi: http://doi.org/10.22215/timreview/1426. URL timreview.ca/article/1426.

Jan vom Brocke, Alexander Simons, Björn Niehaves, Kai Riemer, Ralf Plattfaut, and Anne Cleven. Reconstructing the giant: On the importance of rigour in documenting the literature search process. 06 2009.

Tim Brown. Design thinking. *Harvard business review*, 86:84–92, 141, 07 2008.

Kieran Conboy. Agility from first principles: Reconstructing the concept of agility in information systems development. *Information Systems Research*, 20(3):329–354, September 2009. doi: 10.1287/isre.1090.0236. URL https://doi.org/10.1287/isre.1090.0236.

M. Crowne. Why software product startups fail and what to do about it. evolution of software product development in startup companies. volume 1, pages 338 – 343 vol.1, 02 2002. ISBN 0-7803-7385-5. doi: 10.1109/IEMC.2002.1038454.

Daniela S. Cruzes, Tore Dybå, Per Runeson, and Martin Höst. Case studies synthesis: a thematic, cross-case, and narrative synthesis worked example. *Empirical Software Engineering*, 20(6):1634–1665, December 2015. ISSN 1382-3256, 1573-7616. doi: 10.1007/s10664-014-9326-8. URL http://link.springer.com/10.1007/s10664-014-9326-8.

Michael Debellis and Christine Haapala. User-centric software engineering. *IEEE Expert*, 10: 34 – 41, 03 1995. doi: 10.1109/64.391959.

Kees Dorst. The core of 'design thinking' and its application. *Design Studies*, 32(6):521–532, 2011. ISSN 0142-694X. doi: https://doi.org/10.1016/j.destud.2011.07.006. URL https://www.sciencedirect.com/science/article/pii/S0142694X11000603. Interpreting Design Thinking.

Anh Nguyen Duc and Pekka Abrahamsson. Minimum Viable Product or Multiple Facet Product? The Role of MVP in Software Startups. In Helen Sharp and Tracy Hall, editors, *Agile Processes, in Software Engineering, and Extreme Programming*, volume 251, pages 118–130. Springer International Publishing, Cham, 2016. ISBN 978-3-319-33514-8 978-3-319-33515-5. doi: 10.1007/978-3-319-33515-5_10. URL http://link.springer.com/10.1007/978-3-319-33515-5_10. Series Title: Lecture Notes in Business Information Processing.

Georgi Georgiev. Design thinking: An overview. *Japanese Society for the Science of Design (Design Thinking - Special Issue)*, 20-1:70–77, 01 2012.

Carmine Giardino, Pekka Wang, Xiaofeng, and Abrahamsson. Why Early-Stage Software Startups Fail: A Behavioral Framework. In Wil van der Aalst, John Mylopoulos, Michael Rosemann, Michael J. Shaw, Clemens Szyperski, Casper Lassenius, and Kari Smolander, editors, *Software Business. Towards Continuous Value Delivery*, volume 182, pages 27–41. Springer International Publishing, Cham, 2014a. ISBN 978-3-319-08737-5 978-3-319-08738-2. doi: 10.1007/978-3-319-08738-2_3. URL http://link.springer.com/10.1007/978-3-319-08738-2_3. Series Title: Lecture Notes in Business Information Processing.

Carmine Giardino, Xiaofeng Wang, and Pekka Abrahamsson. Why early-stage software startups fail: A behavioral framework. pages 27–41, 06 2014b. ISBN 978-3-319-08737-5. doi: 10.1007/978-3-319-08738-2_3.

Gabriela Goldschmidt. On visual design thinking: the vis kids of architecture. *Design Studies*, 15:158–174, 04 1994. doi: 10.1016/0142-694X(94)90022-1.

Kasper Hirvikoski. Startups pivoting towards value. pages 1–7, 12 2014.

Saso Josimovski, Lidija Pulevska Ivanovska, and Martin Kiselicki. Implementing the freemium business model in the software industry: Key findings and implications. 10 2019.

Kai-Kristian Kemell, Xiaofeng Wang, Anh Nguyen-Duc, Jason Grendus, Tuure Tuunanen, and Pekka Abrahamsson. 100+ Metrics for Software Startups – A Multi-Vocal Literature Review. page 15, 2018.

Kai-Kristian Kemell, Xiaofeng Wang, Anh Nguyen-Duc, Jason Grendus, Tuure Tuunanen, and Pekka Abrahamsson. Startup Metrics That Tech Entrepreneurs Need to Know. In Anh Nguyen-Duc, Jürgen Münch, Rafael Prikladnicki, Xiaofeng Wang, and Pekka Abrahamsson, editors, *Fundamentals of Software Startups*, pages 111–127. Springer International Publishing, Cham, 2020. ISBN 978-3-030-35982-9 978-3-030-35983-6. doi: 10.1007/978-3-030-35983-6_7. URL http://link.springer.com/10.1007/978-3-030-35983-6_7.

Eriks Klotins. Software start-ups through an empirical lens: are start-ups snowflakes? page 14, 2018.

Eriks Klotins, Michael Unterkalmsteiner, Panagiota Chatzipetrou, Tony Gorschek, Rafael Prikladnicki, Nirnaya Tripathi, and Leandro Bento Pompermaier. A Progression Model of Software Engineering Goals, Challenges, and Practices in Start-Ups. *IEEE Transactions on Software Engineering*, 47(3):498–521, March 2021a. ISSN 0098-5589, 1939-3520, 2326-3881. doi: 10.1109/TSE.2019.2900213. URL https://ieeexplore.ieee.org/document/8643804/.

Eriks Klotins, Michael Unterkalmsteiner, Panagiota Chatzipetrou, Tony Gorschek, Rafael Prikladnicki, Nirnaya Tripathi, and Leandro Bento Pompermaier. SIoT Framework: Towards an Approach for Early Identification of Security Requirements for Internet-of-things Applications. *e-Informatica Software Engineering Journal*, 15(1), 2021b. ISSN 2084-4840. doi: 10.37190/e-Inf210103. URL http://www.e-informatyka.pl/attach/e-Informatica_-_Volume_15/eInformatica2021Art03.pdf.

Eriks Klotins, Michael Unterkalmsteiner, Panagiota Chatzipetrou, Tony Gorschek, Rafael Prikladnicki, Nirnaya Tripathi, and Leandro Bento Pompermaier. SIoT Framework: Towards an Approach for Early Identification of Security Requirements for Internet-of-things Applications. *e-Informatica Software Engineering Journal*, 15(1), 2021c. ISSN 2084-4840. doi: 10.37190/e-Inf210103. URL http://www.e-informatyka.pl/attach/e-Informatica_-_Volume_15/eInformatica2021Art03.pdf.

Praveen K. Kopalle, Manish Gangwar, Andreas Kaplan, Divya Ramachandran, Werner Reinartz, and Aric Rindfleisch. Examining artificial intelligence (ai) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, 39(2):522–540, 2022. ISSN 0167-8116. doi: https://doi.org/10.1016/j.ijresmar.2021.11.002. URL https://www.sciencedirect.com/science/article/pii/S016781162100094X.

Eetu Kupiainen, Mika V. Mäntylä, and Juha Itkonen. Using metrics in agile and lean software development – a systematic literature review of industrial studies. *Information and Software Technology*, 62:143–163, 2015. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2015.02.005. URL https://www.sciencedirect.com/science/article/pii/S095058491500035X.

Herrmann B.L. Dogrultan E. Berman R. Eesley C. Blank S Marmer, M. artup genome report extra: Premature scaling. 2011.

Jorge Melegati, Rafael Chanin, Afonso Sales, and Rafael Prikladnicki. Towards specific software engineering practices for early-stage startups. 09 2020. ISBN 978-3-030-58857-1. doi: 10.1007/978-3-030-58858-8_2.

Raza Mir and Sanjay Jain, editors. *The Routledge companion to qualitative research in organization studies*. Routledge Companions in Business, Management and Accounting. Routledge, London, England, September 2017.

Roland Mueller and Katja Thoring. Design thinking vs. lean startup: A comparison of two user-driven innovation strategies. 08 2012.

Murphy. *Leading on the edge of chaos:The 10 critical elements of success in volatile times*. Prentice Hall, Philadelphia, PA, October 2002.

Jürgen Münch. Data- and value-driven software engineering with deep customer insight. 12 2014.

Suphan Nasir. *Customer Acquisition Management Strategies*, pages 219–243. 01 2015. ISBN 9781466682320. doi: 10.4018/978-1-4666-8231-3.ch008.

Suphan Nasir. *Customer Retention Strategies and Customer Loyalty*, pages 1177–1201. 01 2017. ISBN 9781522517948. doi: 10.4018/978-1-5225-1793-1.ch054.

Dan Olsen. *The lean product playbook*. John Wiley & Sons, Nashville, TN, May 2015.

Jevgenija Pantiuchina, Marco Mondini, Dron Khanna, Xiaofeng Wang, and Pekka Abrahamsson. Are Software Startups Applying Agile Practices? The State of the Practice from a Large Survey. In Hubert Baumeister, Horst Lichter, and Matthias Riebisch, editors, *Agile Processes in Software Engineering and Extreme Programming*, volume 283, pages 167–183. Springer International Publishing, Cham, 2017. ISBN 978-3-319-57632-9 978-3-319-57633-6. doi: 10.1007/978-3-319-57633-6_11. URL http://link.springer.com/10.1007/978-3-319-57633-6_11. Series Title: Lecture Notes in Business Information Processing.

Nicolò Paternoster, Carmine Giardino, Michael Unterkalmsteiner, Tony Gorschek, and Pekka Abrahamsson. Software development in startup companies: A systematic mapping study. *Information and Software Technology*, 56(10):1200–1218, October 2014. ISSN 09505849. doi: 10.1016/j.infsof.2014.04.014. URL https://linkinghub.elsevier.com/retrieve/pii/S0950584914000950.

Justin Paul and Alex Rialp Criado. The art of writing literature review: What do we know and what do we need to know? *International Business Review*, 29(4):101717, August 2020. doi: 10.1016/j.ibusrev.2020.101717. URL https://doi.org/10.1016/j.ibusrev.2020.101717.

Phillip E Pfeifer. The optimal ratio of acquisition and retention costs. *Journal of Targeting, Measurement and Analysis for Marketing*, 13(2):179–188, January 2005. ISSN 1479-1862. doi: 10.1057/palgrave.jt.5740142. URL http://link.springer.com/10.1057/palgrave.jt.5740142.

Usman Rafiq and Xiaofeng Wang. Continuous Information Monitoring in Software Startups. In Maria Paasivaara and Philippe Kruchten, editors, *Agile Processes in Software Engineering and Extreme Programming – Workshops*, volume 396, pages 280–287. Springer International Publishing, Cham, 2020. ISBN 978-3-030-58857-1 978-3-030-58858-8. doi: 10.1007/978-3-030-58858-8_29. URL http://link.springer.com/10.1007/978-3-030-58858-8_29. Series Title: Lecture Notes in Business Information Processing.

Francesca Ricciardi and Cecilia Rossignoli. *Research Methods in the itAIS Community: Building a Classification Framework for Management and Information Systems Studies*, volume 11, pages 297–315. 01 2015. ISBN 978-3-319-23783-1. doi: 10.1007/978-3-319-23784-8_23.

Eric Ries. *The lean startup: how today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Business, New York, 1st ed edition, 2011. ISBN 978-0-307-88789-4. OCLC: ocn693809631.

Nicholas Ross. Customer retention in freemium applications. *Journal of Marketing Analytics*, 6, 12 2018. doi: 10.1057/s41270-018-0042-x.

Per Runeson and Martin Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164, December 2008. doi: 10.1007/s10664-008-9102-8. URL `https://doi.org/10.1007/s10664-008-9102-8`.

Saras Sarasvathy. Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *The Academy of Management Review*, 26, 04 2001. doi: 10.5465/AMR.2001.4378020.

Amir Sayyed-Alikhani, Manuel Chica, and Ali Mohammadi. An agent-based system for modeling users' acquisition and retention in startup apps. *Expert Systems with Applications*, 176:114861, August 2021. ISSN 09574174. doi: 10.1016/j.eswa.2021.114861. URL `https://linkinghub.elsevier.com/retrieve/pii/S095741742100302X`.

Daniel Schallmo, Christopher Williams, and Klaus Lang. An integrated design thinking approach-literature review, basic principles and roadmap for design thinking. 07 2018.

Kamran Siddiqui. Personality influences customer loyalty. *Sci.Int.*, 28:477–480, 01 2016.

Rini Solingen and Egon Berghout. The goal/question/metric method: A practical guide for quality improvement of software development. 01 1999.

Silvana Trimi and Jasmina Berbegal-Mirabent. Business model innovation in entrepreneurship. *International Entrepreneurship and Management Journal*, 8:449–465, 2012.

Orlando Troisi, Gennaro Maione, Mara Grimaldi, and Francesca Loia. Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management*, 90:538–557, October 2020. ISSN 00198501. doi: 10.1016/j.indmarman.2019.08.005. URL `https://linkinghub.elsevier.com/retrieve/pii/S0019850118308496`.

Xiaofeng Wang, Henry Edison, Sohaib Shahid Bajwa, Carmine Giardino, and Pekka Abrahamsson. Key challenges in software startups across life cycle stages. In Helen Sharp and Tracy Hall, editors, *Agile Processes, in Software Engineering, and Extreme Programming*, pages 169–182, Cham, 2016. Springer International Publishing. ISBN 978-3-319-33515-5.

Sezin Yaman, Fabian Fagerholm, Myriam Munezero, Tomi Männistö, and Tommi Mikkonen. Patterns of user involvement in experiment-driven software development. *Information and Software Technology*, 120:106244, 2020. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2019.106244. URL `https://www.sciencedirect.com/science/article/pii/S0950584919302629`.

Xiaoming Yang, Sunny Sun, and Xiangyang Zhao. Search and execution: examining the entrepreneurial cognitions behind the lean startup model. *Small Business Economics*, 52, 03 2019. doi: 10.1007/s11187-017-9978-z.

Robert K Yin. *Applications of case study research*. SAGE Publications, Thousand Oaks, CA, 3 edition, August 2011.

Zaidah Zainal. Case study as a research method. page 6, 2007.

Chubing Zhang, Xinchun Wang, Annie Peng Cui, and Shenghao Han. Linking big data analytical intelligence to customer relationship management performance. *Industrial Marketing Management*, 91:483–494, 2020. ISSN 0019-8501. doi: https://doi.org/10.1016/j.indmarman.2020.10.012. URL https://www.sciencedirect.com/science/article/pii/S0019850120308762.

Appendix

1 Customer Data Platform

A customer data platform (CDP) is a system that collects, consolidates, and manages data from various sources to provide a unified view of customer information. It enables businesses to understand their customers better and personalize their interactions with them across different channels and touchpoints.

CDPs typically gather data from a variety of sources, including website interactions, customer service interactions, email campaigns, and social media activity. This data is then cleaned, consolidated, and stored in a central location, where it can be segmented, analyzed, and used to create targeted marketing campaigns and personalized customer experiences.

CDPs can also integrate with other systems, such as marketing automation platforms, CRM systems, and analytics tools, to provide a seamless flow of data and insights. This enables businesses to make data-driven decisions and deliver personalized experiences across different channels, such as email, web, mobile, and in-store.

CDPs are different from traditional data management systems, such as data warehouses and data lakes, in that they are specifically designed to manage customer data, rather than operational or transactional data. They also provide a more complete and accurate view of the customer, by consolidating data from various sources and providing real-time insights.

2 Behavioural data

This thesis is focused on SaaS (Software as a Service) start-ups. Specifically the product those companies are marketing is a software that runs on a browser (web application). Unlike software programs that run locally on a machine, web apps only need internet connection to work properly. The client (front-end) sends requests to the server (back-end) which performs the calculations before sending back the result to the user interface.

2.1 Event and relational data

There are mainly two ways of gathering data from a web app: event data and relational data. Event data is a record of behaviours performed by entities that is very useful to describe trends and habits. Event data is formed by an action, a timestamp and the state. Data is not normalised (duplicated information) and rich of details. Relational data is stored in tables and describes the state of an entity in a particular moment. For this reason it is not particularly suitable to study the actions of users or the evolution of a variable. Entities can be products, users, levels etc. Every entity has is a row in the table and each column describes an attribute. Relational data is normalised and this means that there are no duplicated information and the storage is more efficient. For these reasons in this thesis the data used will be event data given that it's more suitable to investigate behaviour.

2.2 How to gather event data

There are mainly four types of interactions that will be used in this study to gather learnings about users: page event, track event, identify call and group call. These four types of information can be gathered by including a specific script in the codebase of the software in correspondence of the action to be tracked.

Page events

A page event records whenever sees a page of your website or app together with any additional properties about the page. However this type of event does not give a high level of detail of the actions of a user given that in most pages there are multiple actions that can be performed.

Track events

A track event records whenever a user performs a specific action like a click or a text input. This type of event is more specific and can be rich in properties. For example tracking the action of signing up to a website can be enriched with the information of the type of sign up performed (with Google, with Facebook, Twitter etc). This extra colour is called property.


```json
{
  "anonymousId": "23adfd82-aa0f-45a7-a756-24f2a7a4c895",
  "context": {
    "library": {
      "name": "analytics.js",
      "version": "2.11.1"
    },
    "page": {
      "path": "/academy/",
      "referrer": "",
      "search": "",
      "title": "Analytics Academy",
      "url": "https://segment.com/academy/"
    },
    "userAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X
    "ip": "108.0.78.21"
  },
  "event": "Course Clicked",
  "integrations": {},
  "messageId": "ajs-f8ca1e4de5024d9430b3928bd8ac6b96",
  "properties": {
    "title": "Intro to Analytics"
  },
  "receivedAt": "2015-12-12T19:11:01.266Z",
  "sentAt": "2015-12-12T19:11:01.169Z",
  "timestamp": "2015-12-12T19:11:01.249Z",
  "type": "track",
  "userId": "AiUGstSDIg",
  "originalTimestamp": "2015-12-12T19:11:01.152Z"
}
```


Figure 1: Example of the code needed for a track event

Identify calls

An identify call allows to tie a user to their actions (which will be recorded using the Track or Page events) and record traits about them. When an identify call happens a unique user ID is generated and assigned to that specific user. Traits such as name, email, phone number, role can be collected and attached to the user ID as traits.


```
{
  "anonymousId": "507f191e810c19729de860ea",
  "channel": "browser",
  "context": {
    "ip": "8.8.8.8",
    "userAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_5)
  },
  "integrations": {
    "All": false,
    "Mixpanel": true,
    "Salesforce": true
  },
  "messageId": "022bb90c-bbac-11e4-8dfc-aa07a5b093db",
  "receivedAt": "2015-02-23T22:28:55.387Z",
  "sentAt": "2015-02-23T22:28:55.111Z",
  "timestamp": "2015-02-23T22:28:55.111Z",
  "traits": {
    "name": "Peter Gibbons",
    "email": "peter@example.com",
    "plan": "premium",
    "logins": 5,
    "address": {
      "street": "6th St",
      "city": "San Francisco",
      "state": "CA",
      "postalCode": "94103",
      "country": "USA"
    }
  },
  "type": "identify",
  "userId": "97980cfea0067",
  "version": "1.1"
}
```


Figure 2: Example of the code needed for an identify call

Group calls

A group call allows to associate an individual user with a group (whether it's a company, an account or a project). When a group is identified for the first time a group ID is generated and then all the actions of the members are associated with it.

To be able to leverage behavioural data a visualisation tool is needed.

3 User Retention calculation

User retention definition: percentage of users that are active at a certain point in time after they have signed up or installed the software. Formula: (active users in a certain time) / (total number of registered users). The second term is easy to determine whereas the first one is a bit more complex. Typically a software product has an ideal frequency of use. Airbnb is used ideally only a couple of times a year. A tool to calculate taxes has a monthly ideal frequency, Uber has a weekly one and WhatsApp is supposed to be used daily. This means that Airbnb considers a user retained even if the a user was active only once in a year time. For this reason the retention calculation should take into account the ideal frequency of usage.

```
{
  "anonymousId": "507f191e810c19729de860ea",
  "channel": "browser",
  "context": {
    "ip": "8.8.8.8",
    "userAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_5)
  },
  "integrations": {
    "All": true,
    "Mixpanel": false,
    "Salesforce": false
  },
  "messageId": "022bb90c-bbac-11e4-8dfc-aa07a5b093db",
  "receivedAt": "2015-02-23T22:28:55.387Z",
  "sentAt": "2015-02-23T22:28:55.111Z",
  "timestamp": "2015-02-23T22:28:55.111Z",
  "traits": {
    "name": "Initech",
    "industry": "Technology",
    "employees": 329,
    "plan": "enterprise",
    "total billed": 830
  },
  "type": "group",
  "userId": "97980cfea0067",
  "groupId": "0e8c78ea9d97a7b8185e8632",
  "version": "1.1"
}
```


Figure 3: Example of the code needed for a group call

The second factor is how to define a user active. In the SaaS (Software as a Service) industry a user is considered active if the main value proposition of the software is exploited. For Airbnb this value could be either registering a new apartment / room or book a stay. For WhatsApp it would be sending a message.

For the company which is object of the case study the ideal frequency of usage is weekly and a user is considered active if a report is opened. This is an example of how an acquisition and retention cohort looks like.

DATE	USERS	NUMBER OF WEEKS LATER YOUR USERS WERE RETAINED				
		0	1	2	3	4
May 30-05 2022	34	100%				
May 23-29 2022	40	100%	38%			
May 16-22 2022	40	100%	35%	20%		
May 09-15 2022	36	100%	42%	39%	31%	
May 02-08 2022	33	100%	27%	39%	33%	24%

Figure 4: Example of weekly cohorts retention

The numbers in the first column (34, 40, 40, 36 and 33) are the number of new users acquired in the corresponding calendar week. The percentages on the right is the user retention on that particular cohort in the corresponding week.

4 Data set

The data set used for this thesis has been extracted from the database of the company object of the case study. Access to this information could be grated by the founders of the company after a legitimate request.

4.1 User engagement

Group 1

Eligibility status	User engagement	Sample	User retention
tracking events=false *identifying users=false* *identifying groups=false*	*Number of loaded reports = 5720* *Average loaded reports per user 7.37*	*Number of eligible users = 776*	*User retention = 7.99%*

Table 1

Group 2

Eligibility status	User engagement	Sample	User retention
tracking events=true *identifying users=true* *identifying groups=false*	*Number of loaded reports = 4818* *Average loaded reports per user 17.64*	*Number of eligible users = 272*	*User retention = 18.8%*

Table 2

Group 3

Eligibility status	User engagement	Sample	User retention
tracking events=true *identifying users=true* *identifying groups=true*	*Number of loaded reports = 4623* *Average loaded reports per user 18.1*	*Number of eligible users = 255*	*User retention = 16.8%*

Table 3

4.2 User traits and collaboration

Group 1 - Collaboration and user traits

Collaboration

Collaboration	Sample	User retention
invited_a_teammate >0	*Number of eligible users = 85*	*User retention = 68%*
invited_a_teammate <1	*Number of eligible users = 691*	*User retention = 0.6%*

Table 4: Group 1 with collaboration

User trait - role Here data is sliced based on the role of the user.

User trait	Sample	User retention
role = "founder"	*Number of eligible users = 63*	*User retention = 68.9%*
role = "product manager"	*Number of eligible users = 142*	*User retention = 12.6%*
role = "engineer"	*Number of eligible users = 571*	*User retention = 0%*

Table 5: Group 1 with role breakdown

Group 2 - Collaboration and user traits

Collaboration

Collaboration	Sample	User retention
invited_a_teammate >0	*Number of eligible users = 84*	*User retention = 45%*
invited_a_teammate <1	*Number of eligible users = 189*	*Average user retention = 6.87%*

Table 6

User trait - role

User trait	**Sample**	**User retention**
role = "founder"	*Number of eligible users = 64*	*User retention = 67%*
role = "product manager"	*Number of eligible users = 33*	*User retention = 24.2%*
role = "engineer"	*Number of eligible users = 176*	*User retention = 0%*

Table 7

Group 3 - Collaboration and user traits

Collaboration

Collaboration	Sample	User retention
invited_a_teammate >0	*Number of eligible users = 79*	*User retention = 41.7%*
invited_a_teammate <1	*Number of eligible users = 177*	*User retention = 5.65%*

Table 8

User trait - role

User trait	Sample	User retention
role = "founder"	*Number of eligible users = 60*	*User retention = 64%*
role = "product manager"	*Number of eligible users = 41*	*User retention = 12.2%*
role = "engineer"	*Number of eligible users = 155*	*User retention = 0%*

Table 9