Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Software Engineering
Master's Programme in Software Engineering and Digital Transformation

Grigorii Shestakov

B2B IT STARTUP MANAGEMENT BASED ON A DATA-DRIVEN DECISION-MAKING APPROACH

Examiners: Associate Professor Jussi Kasurinen
Associate Professor Sami Hyrynsalmi

Supervisors: Associate Professor Jussi Kasurinen

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Software Engineering
Master's Programme in Software Engineering and Digital Transformation

Grigorii Shestakov

B2B IT startup management based on a data-driven decision-making approach.

Master's Thesis

2021

89 pages, 26 figures, 21 tables,

Examiners: Associate Professor Jussi Kasurinen

Associate Professor Sami Hyrynsalmi

Keywords: IT startup, Data-driven decision-making, Data analysis, B2B market, mobile application

This master's thesis is devoted to the research of a data-driven decision-making approach. The research was carried out on the basis of a Russian IT startup called "Restik" which creates cloud solutions for automating the restaurant business. The main research question is how to apply principles and methods of data-driven decision-making approach to mobile applications operating in B2B market. General scientific research methods, literature analysis, case studies, and financial modeling were used during the research. As a result, a data-driven management approach was investigated, the approach was adapted to the studied startup and practically applied to decision-making within the studied startup. This paper considers 4 different practical cases of approach usage, which show how data could help in such complicated task as managing the start-up.

ACKNOWLEDGEMENTS

Thanks to my supervisor from university Lappeenranta-Lahti University of Technology LUT – Assist. Professor Jussi Kasurinen and also supervisor from university Peter the Great St. Petersburg Polytechnic University – Associate Professor Ilyashenko O. Y., for guidance and feedback throughout the project. Also, I would like to thank my startup team, who help me during my research and give me opportunity to test this approach on the real example.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

AGM	Average gross margin
API	Application Programming Interface
ARPU	Average revenue per user
AUL	Average user lifetime
B2B	Business to business
CAC	Client acuisition cost
CRM	Customer Relationship Management
DAU	Daily active users
ETL	Extract-Transform-Load
IRR	Inner rate of return
IT	Information technologies
KPI	Key performance indicators
LTV	Lifetime value
MAU	Monthly active users
MVP	Minimun valuable product
NPV	Net present value
POS	Point of sale
PR	Public relations
ROI	Return on invesnment
SDK	Software development kit
TR	Total revenue

1. INTRODUCTION

This introduction will set the scene for my master's thesis. Firstly background and motivation for the study will be described. Then goals and delimitations of the study will be set. Finally, overall thesis structure will be considered.

1.1. Background

On average, 9 out of 10 startups fail [1]. There are many reasons why a seemingly successful startup eventually goes bust, one of these reasons is the wrong management decisions [2]. The problem with such decisions is that the results of the decision are very difficult to evaluate immediately, and the consequences caused by these decisions may appear long after the decision is made.

In addition to the difficulty in making decisions, the startup market has also faltered during the COVID-19 pandemic. The consequences for many startups were devastating, 43 percent of European startups suspended the hiring process [3], and the volume of venture capital investments decreased by 38 % in the first 2 months of the beginning of the pandemic, compared with the previous period [4]. In such a difficult period for doing business, the management of startups has to look for new ways to effectively and timely solve emerging problems [5].

Thus, taking into account the state of the market, as well as the general complexity of managing startups, it is necessary to develop an approach that will help reduce the risks of making incorrect management decisions, since at the moment any mistake can be very expensive not only for the management of startups, but also for investors.

One of the approaches to help in decision-making is the data-driven decision-making approach [6]. The idea behind this approach is quite simple: "Use your actual startup data, instead of your feelings about how market works, for decision making". Despite the simplicity of this idea, the implementation of this approach can be difficult, which can ultimately lead to bankruptcy of the startup, so it is very important to understand why and how approach should be implemented.

My startup team decided to use this approach after several decisions which lead to bad consequences, so it is become clear that somehow, we should validate that management decisions will lead to startup to desired state. As a startup analyst I started to research which approaches could fit to our startup, and decided to adapt data-driven decision-making approach for our startup purposes.

1.2. Goals and delimitations

Within this research work, will be considered the basic principles of a data-driven management approach, the tools that startups can use to implement this approach, and also the cases of one of the startups creating an application on the B2B market.

Research work also has the following delimitations:

1. The thesis considers startups that creating apps on B2B market.
2. Data-driven decision-making approach is not a standardized framework, the description of the approach in this thesis is a consequence of the analysis of cases and literature and does not pretend to be an industry standard.
3. This work is not an action guide, but only a description of the approach and in practice shows how this approach can be applied.

The main research questions of this work are:

1. What is data-driven decision-making?
2. How can this approach be applied by a start-up operating on the B2B market?
3. What is the effect of using data-driven approach?

Main research methods are literature review, cases studies, basic general scientific methods and financial modelling.

1.3. Structure of the thesis

The study consists of 4 main parts, in addition to this, the study also contains an introduction, conclusion, and a bibliography. Logically, the narration is built from the study of theoretical aspects to the application of the knowledge gained in practice.

The first part examines the rationale for using the approach by analyzing the current state of the startup market in Europe. It also defines the term startup and the startup life cycle. One of the factors influencing the choice of approach was also the COVID-19 virus pandemic, the impact of which on the startup market was also discussed in the first part.

In the second part, by analyzing the literature and various cases, it is determined what a data-driven decision-making approach is. The principles, methods, and tools of the approach are investigated, a theoretical base is created with the help of which the approach will be implemented in an existing startup.

The third part explores a startup that plans to implement the approach. Due to the fact that the approach completely changes such an important aspect of management as decision-making, it is necessary to create in advance many tools and methods that will be used in the future to make decisions based on data.

In the fourth part, practical cases of application of the approach are considered, with specific descriptions of when the approach helped to make decisions, and how these decisions were ultimately reflected in the startup.

2. STARTUP PROCESSES OVERVIEW

This chapter will consider general picture of IT startups in Europe, define the term startup, consider the stages of development of startups, and also assess the impact of COVID-19 on the startup market.

2.1. Current startups state in Europe

At the moment, there is no single clear definition of the term startup. Since from different points of view, the term startup can be interpreted in different ways . We will take a definition from Eric R.'s book The lean startup, according to this book startup is "a human institution designed to create a new product or service under conditions of extreme uncertainty."[7] A startup is based on the assumption that a proposal (service or product) will be accepted by the market as it solves the customer's problem. Due to the high uncertainty, many startups go bust, but some of the startups become so-called “unicorns” (a private startup that has reached a market capitalization of $ 1 billion) [8]. According to CB Insights, as of May 2021, there are more than 600 unicorn startups in the world [9]. This is the reason why, despite all possible failures, startups are so attractive to investors. Every year, startups attract multibillion-dollar investments (31.313 million euros in Europe in 2019 [10]) and create many jobs.

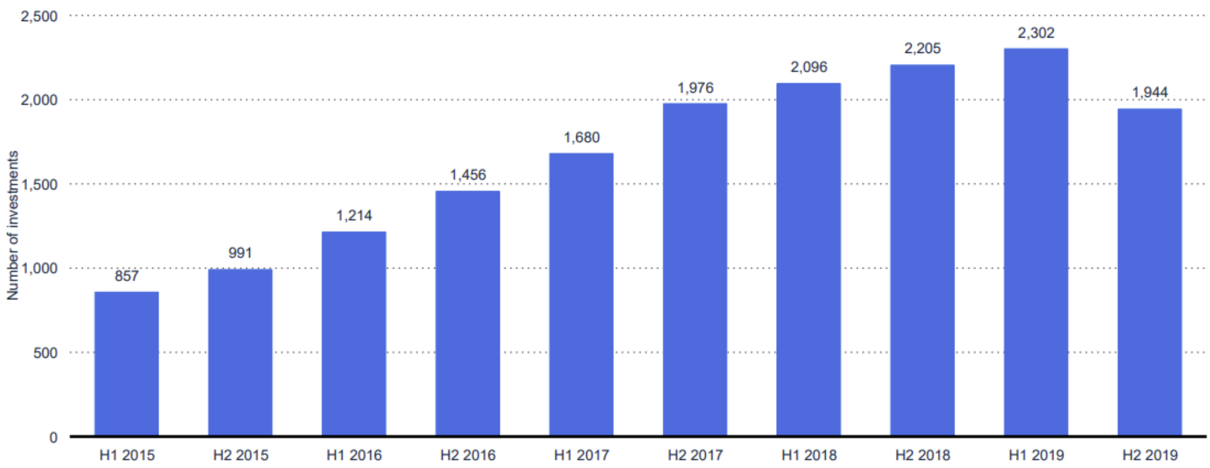


Figure 1. Amount of investments in startups in Europe since 2015 (in mln euros)

At the same time, in 2020, according to Tech Tour research, 61 % of super startups focused on the B2B market.

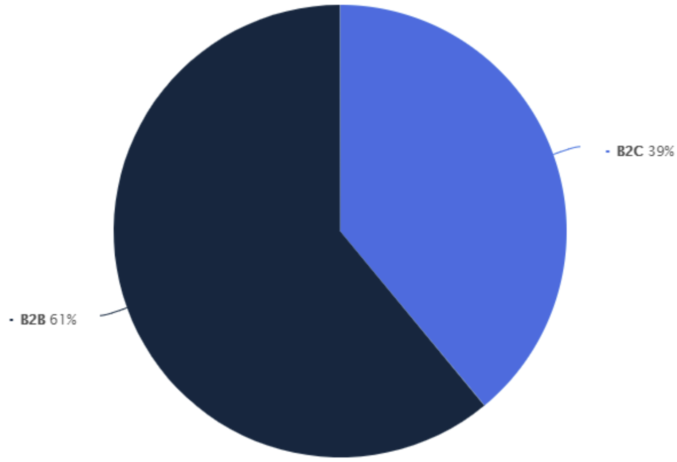


Figure 2. Distribution of super start-up companies offering B2B and B2C solutions worldwide in 2020

The most popular technology area for startups in Europe in 2018 was IT / software development (19.1 % of the total as shown on fig 3.).

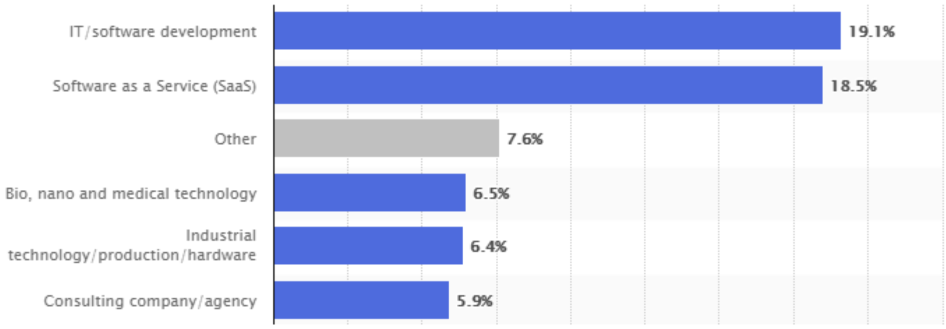


Figure 3. Distribution of startups according to industry sector in Europe in 2018

Thus, it can be argued that for the period up to 2018, the creation of a B2B startup in the IT industry was the most popular direction for startups in Europe. For a better understanding of the market, it is also necessary to understand what stages startups go through, so further startup development stages, in general, will be considered.

2.2. Startup development stages

This section will consider the stages of a startup from a development point of view, stages can also be considered from a funding point of view (Seed stage, startup stage, growth stage), but in this case we are only interested in the development stages. There are many interpretations and formulations about what stages a startup goes through [11-14] in the development process, in this work will be used the following 5 stages: Concept and research, MVP, Traction, Scaling, Exit. Take a closer look at each of these stages.

1.2.1 Concept and research

The entrepreneurial challenge of the startup founder is, to identify and validate the key value (business concept) that underlies the startup. To do this, it is necessary to understand the market opportunities, supply in this market and form an initial business model. If the initial value hypothesis is found and confirmed, it is also necessary to find a distribution channel through which the value will be conveyed to the end-user. Thus, the initial success for a startup is the product itself (the value hypothesis that underlies it), the distribution channel through which the product will be distributed, and the market in which the startup will operate. When these factors are found, a startup will not necessarily be successful, but all successful startups have found these components (perhaps not all startup owners specifically looked for these 3 factors, but when considering any successful startup, it is possible to understand what the product, market, and channel was). Consider each of these components.

The value hypothesis essentially is an answer to the question “Why will people use a startup?” [15]. The startup creator must understand why his startup will be attractive to the market, what is an urgent need and how the new startup will effectively solve itit. As American businessman and entrepreneur Ben Horowitz wrote in his book "The Hard Thing About Hard Things: Building a Business When There Are No Easy Answers": “The primary thing that any technology startup must do is build a product that's at least ten times better at doing something than the current prevailing way of doing that thing. Two or three times better will not be good enough to get people to switch to the new thing fast enough or in large enough volume to matter” [16]. And while no studies have been cited to support this phrase to date, many market participants agree with it. The startup creator needs to understand how to create a product that will be 10 times better than its predecessors.

Market - before creating a startup, it is also necessary to evaluate the market in which the startup is planned to operate. To understand what kind of players are present in this market, what strengths and weaknesses they have, what user tasks, and how the players in this market solve them. It is important to understand that it is necessary to consider not only direct analogs but also all possible technologies that solve a similar problem. For example, when WhatsApp entered the messenger market, its main competitor was not other messengers, but SMS messages, because they solved the same problem as WhatsApp - the quick transfer of information from one person to another [17]. Therefore, when assessing the market, it is important to understand exactly what task the planned startup is performing and in what other ways users are solving this task.

The distribution channel is the channel through which the startup owners deliver value to the market. The channel is also very important. For example, if a startup creator has a great startup that effectively solves some problems, and a market segment that needs this startup, but at the same time has not found a channel with which he can convey value to the segment, most likely the startup will fail. For example, people studying programming is a market, creator could interact with them through advertising on the Internet, using articles on thematic resources, and advertising on billboards. However, most likely, if the creator choose advertising on billboards as a distribution channel, most of the target audience simply will not recognize your company. (This is just an example, in reality, it is needed to study what channels your target audience uses).

1.2.2 MVP

After determining the value, market segment, and distribution channel, a Minimal valuable product (MVP) is created - this is a minimum viable product that delivers the main value to the consumer [18]. An MVP is created to test that the value hypothesis underlying the product is correct, the target audience needs the product, and the distribution channels are working. Before the mass release, company could conduct the so-called Soft Launch - a trial launch for a limited audience. Soft launch has several goals:

1. Collecting feedback from real users.
2. Collecting data on key metrics (if the value hypothesis is not correct, such metrics will show poor results, then it is better to change the product before the mass release so that it is in demand in the market).
3. Find all bugs and problems that were not identified as a result of testing.
4. Check marketing channels

After analyzing the results and confirming the main hypothesis of value, the startup moves from the MVP stage to the Traction stage.

1.2.3 Traction

At this stage, the startup begins to attract its first users. If everything was determined correctly at the previous stages, a competent marketing strategy becomes the main task of startup management [15]. It is important to competently present product to the world, for example, partnerships with various information resources, company PR, and performances at thematic exhibitions can help.

The communication strategy becomes extremely important, since at this stage a core of the most loyal users is formed around the product, so it is extremely important to quickly respond to feedback and make changes in accordance with market demands. Also, at this stage, it is possible to notice unusual user behavior, information about which can be used to search for further points of growth.

1.2.4 Scaling

The startup owner already has a working product that attracts new users, but the product is not yet mass-produced, so at this stage it is necessary to find new ways to expand the client base, the offer and the company itself. Development becomes iterative, at each iteration it is necessary to analyze what is happening with the product and change it for the better.

1.2.5 Exit

At this stage, a startup leaves the status of a "startup". Depending on the goals of the management, a startup can be sold to a global campaign, go public, or remain a private company. The most important thing is that a startup at this stage already has a solution to a specific task for a certain market segment, and in addition to the task of attracting users, the task of retaining users also appears.

These stages are given in general terms and can vary or depending on the specific situation, however, any startup in one way or another goes through these 5 stages or is at one of them. At each stage, management has to make decisions about how the startup should develop. Despite the abundance of cases, each new startup is a separate case, managment can't just copy existing solutions and hopes that they will bring success. It is necessary to constantly develop a startup, find new points of growth and new ways to create value for the end user. In the conditions of uncertainty in which startups are created, all these tasks become a real challenge for entrepreneurs.

To complete the picture of the state of startups in Europe, it is also necessary to analyze the influence of another factor - the COVID-19 virus, the spread of which was recognized by WHO as a pandemic in March 2020.

2.3. COVID-19 impact on startups

The outbreak of the COVID-19 virus began with the report of an unknown virus in December

2019. In the next few months, the virus spread throughout the world, which is why WHO declared an outbreak of the virus a pandemic in March 2020. The virus caused the global crisis, which affected both the social and economic spheres of life of the world's population. The lack of a quick medical response to the virus (vaccine or treatment) has necessitated lockdowns in many countries. Like normal business, startups were hit just as hard during this time. Consider how the startup market has changed since December 2019.

1.3.1 Private Equity & Venture Capital

Private and venture capital investments are one of the main sources of income for startups. Private equity investments are medium to long term investments by individuals in unlisted companies in exchange for stakes in those campaigns. Private equity firms focus on established campaigns that require capital gains and reorganization in order to be sold at a profit. Venture capital is money, which is money that most often helps to start a business from scratch. A venture capital firm invests in a company early in its development and provides it with the critical capital it needs to get started and, if lucky, grow actively. Venture capital firms are betting on growth, often neglecting profitability, which is why they are more likely to invest in companies with high growth potential. Such investors orient startups towards rapid, not always sustainable, growth. Venture capitalists are playing the long game by investing early in companies that can generate huge returns. Many unicorn startups rely on venture capital, for example 82 % of 190 European startups that have achieved unicorn status were funded by venture capital investments [19]. In previous global economic crises, the number of private and venture capital investments initially declined, but then peaked, as shown in Table 1.

Table 1 - Funds raised during previous financial crises [19]

	Privatre equity		Venture capital	
Year of final close	Amount of funds	Aggregated capital raised, bn $	Amout of funds	Aggregated capital raised, bn $
Dot com era				
1999	225	92,9	293	42,2
2000	285	130,4	488	76,4
2001	256	91	359	43,5
Global financial crisis				
2007	659	366,2	424	46,5
2008	655	357,5	445	52,7
2009	447	185,5	360	26,9
Pre-coronavirus pandemic				

2018	833	537,9	1105	108,5
2019	733	548,8	872	95,7
2020 Q1	211	102,6	225	37,7

There is still no exact information on the impact of COVID-19 on the investment market in 2021, but it can be said for sure that the amount of attracted investment and private capital decreased in Q1 2020, which will also affect the development of startups.

1.3.2 Employment during COVID-19

As a result of financial difficulties, startups are faced with the problem of hiring employees. 43 % of European startups have frozen their hiring process. A poll conducted by LocalGlobe and Dealroom among 140 startups, mainly from France and Germany, created over the past 5 years, showed that more than a third of the startups-respondents made decisions to downsize staff on a permanent basis, 17 percent of respondents said they could lay off 10 % of the current state workers [3]. The experts consider that problems caused by COVID 19 pandemic will have far more serious implications than those companies are used to dealing with, thus, in order to overcome these challenges, companies should adopt an operational model that accounts for extreme uncertainty.

Thus, in addition to the standard difficulties, startup owners also had to cope with the consequences of the pandemic. In such conditions, any mistake in the decisions made can lead to undesirable consequences. Therefore, entrepreneurs were faced with the task of finding a method with which they could increase the likelihood of making the right decision, one of such methods is the data-driven decision-making approach.

3. DATA-DRIVEN DECISION-MAKING APPROACH

The decision-making process may be really confusing. When making decisions, it is necessary to take into account many factors (which often may not be obvious), and the consequences of the decisions made can reveal themselves years later, after their adoption. Each decision-maker has his own approach to how to make these decisions, someone uses personal experience, someone relies on the reports of reputable consulting agencies, and someone makes decisions based on their understanding of how a process must be arranged. These approaches can be used in various fields, but none of them provide assurance that the decision will actually change the process as originally intended.

In the modern world, information is one of the most valuable resources. Information is collected during many human interactions with the real world. This is how the idea that collected information could help in making decisions arose [20]. The data-driven approach can be applied in different areas to achieve different effects and vary depending on the specific area, however, the thought behind this approach remains the same - “Data-driven decision making is the process of making organizational decisions based on actual data rather than intuition or observations.".

First, it is needed to once again determine what impact the implementation of a data-driven environment could bring. As already mentioned, decisions can be made on various grounds (personal experience, observations, etc.) and, with some probability, achieve or not achieve the goal that was pursued when making the decision. A data-driven decision-making approach does not guarantee that the decision will accurately achieve the goal, but it increases the likelihood that the goal will be achieved.

At the same time, the use of this approach does not prohibit the use of personal experience and expert advice but complements the decision-making process with tools that can be used to validate whether a given decision is significant and can potentially achieve the goal that is set for it.

3.1. Data-driven approach principles

Since the data-driven approach is not a framework, but rather a way of thinking, in general, there are no fundamentally established principles for it, however, these principles can be established for each separate area, after analyzing the literature [21-25] with different cases, the following principles for B2B IT startup could be distinguished:

1.Trust facts, not hypotheses - a data-driven approach implies that a product that is being developed by an IT startup can only be evaluated using a model that is built on the basis of the product. Therefore, each hypothesis needs to be tested and supported by data. For example, the statement “If the marketing budget will be raised by 50 %, it will bring 1000 new users” is a hypothesis, and “after increasing the marketing budget by 50 %, 1000 new users came to product” is a fact. It is extremely important to draw the right conclusions from the facts, for example, from the fact of the above, it does not follow that every 50 % increase in the budget will bring 1000 users. It just means that for some reason, in this case, the increase in the budget worked. It is necessary to analyze why exactly it worked and what factors influenced it. This is the main idea of this principle, observed events are facts from which conclusions can and should be drawn, assumptions about how the market works are hypotheses that must be supported by facts and experiments.

2.Compare the Comparable - When analyzing data, the task of comparison often arises (comparing user groups, comparing application versions, comparing markets, etc.), in this case, the data-driven decision-making approach advises comparing what can be compared with each other. When analyzing the behavior of a product or users within a product, we only see the result, which was influenced by many factors. The task of comparison is to determine how the observed factor has changed the compared objects. In the context of startup management, it is necessary to clearly understand that compared objects should differ only in the factor by which they are compared. For example, the task is to evaluate 2 different versions of an application, while it is known that in the first version mainly users from one country are represented and in the second version from another country. In this case, the comparison will be incorrect, since, in addition to the factor of the new application, the factor of different geography of users also appears. That is, the essence of the principle is that before making a comparison, it is necessary to make sure that the observed groups differ from each other only by the factor whose influence we are evaluating. In practice, it is not always possible to create 2 identical groups, however, it is worth striving to ensure that the compared groups are similar in all respects. If the influence of a certain factor on one of the groups is known, this influence must be removed or minimized.

3.You interact with the product model, not the product itself - the data-driven approach is characterized by the idea that the team that produces the product does not know how the product actually works (meaning not the technical aspect, but how product works in the market). All the conclusions a team can make about a product are based on some product model that team members form in their heads. For a data-driven environment, this model must be the same for every team member. At the start, the model is built only on the basis of the team's theoretical ideas: hypotheses about the value of the product, its distribution channels and possible ways of monetization. Then the team begins to test them through experiments, communicate with users, and develop expertise. Learning from the lessons, the team changes their understanding of the product and, as a result, its model. The basic model can be the so-called bucket model [15]. At the start, the product is a leaky bucket into which the product team brings new users (pour water into a bucket), some of these users leave, because they do not find value for themselves (water is poured out through the holes in a bucket), and some users remain (water remains in a bucket). Then it is necessary to constantly clarify this model, how exactly users come to the product, why they leave, why they stay, how to make sure that as few users leave as possible, etc. all these questions are necessary in order to understand how the product actually works, and not how it works in the mind of the creators.

3.2. Data-driven approach process

Based on the principles of the approach, it is possible to build a data-driven management process. Before creating a startup, the approach advises the same steps as described in the startup development stages section. It is necessary to conduct market research and formulate a value hypothesis. Already at this stage, even without creating an MVP, the correctness of the value hypothesis could be evaluated, for example, using the "Splitmetrics" service, which allows creating an application page in popular application stores, without creating the application itself. The approach advises testing any value hypothesis experimentally and doing so with minimal investment. The main idea is to check that the value hypothesis is correct and the product will really be in demand while spending the minimum amount of time and money on this verification.

After confirming the value hypothesis and creating an MVP based on this hypothesis, the process of further product development begins, Figure 4 shows the process for IT startups (this scheme can also vary depending on the specific task, but in general the process will remain the same).

The first step is to determine what problems the startup has (low download rate compared to the market, key indicators not achieved, etc.). At the same time, it is necessary to identify problems constantly, and not only when the startup does not go in accordance with the KPI. “Problem” - in this case, refers to a part of the startup that can be improved.

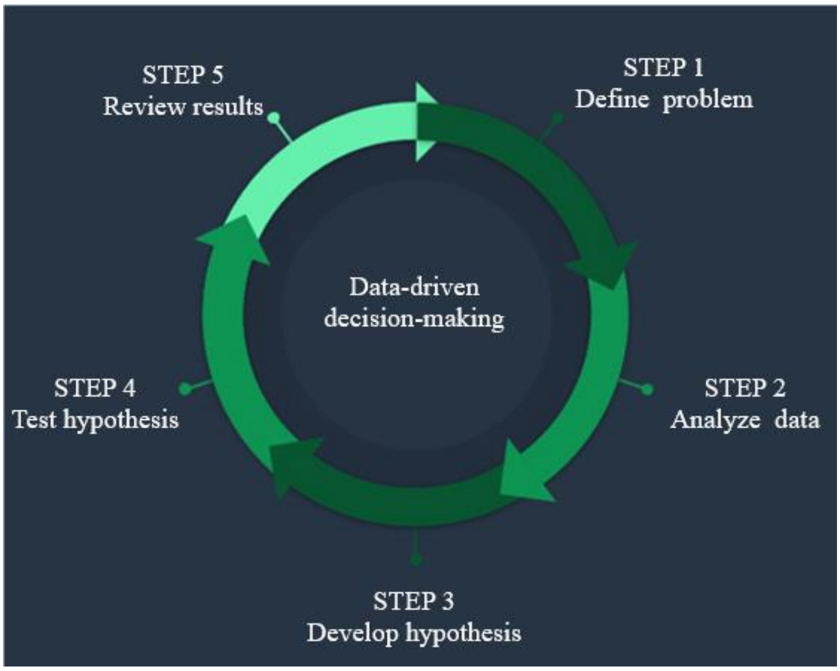


Figure 4. Data-Driven approach process

In the second step, data on how customers use the application need to be analyzed (customers, in this case, can be users, business, etc., depending on the market and area in which the IT startup operates). In order to analyze this data, various metrics are used. Metrics - the numerical value of some property of the developed software to control the creation process. In addition to analyzing metrics, this step also includes the analysis of quantitative and qualitative data obtained during interviews, surveys, and other methods of collecting information from the end-user.

In the third step, based on the data obtained, it is necessary to develop a hypothesis - some assumption about your product, put forward to explain any phenomena. The main property of the hypothesis is that it is not supported by any facts, but is only a certain conclusion from the analyzed data.

The fourth step is to test the hypothesis using an experiment. This is done using various techniques as experiments, A/B testing, user research. Essentially using technique here is just instrument, using which startup management could understand if initial hypothesis was right or wrong.

At the last step of the cycle, the results are checked, based on the results of the experiment, it is necessary to make a decision about the startup. If hypothesis was right new features, based on this hypothesis should be implemented. The cycle then starts over. In general terms, a data-driven approach looks like this, before you implement any change, you need to make sure that it has a fruitful effect on the IT startup.

3.3. Data driven approach methods

This section will describe the techniques that apply to data-driven decision-making. This list was compiled upon analysis of the current state of affairs and is subject to change over time. It is also worth noting that this list is typical for IT startups and may vary depending on the area.

2.3.1 Experiment.

An experiment is a procedure performed to support, disprove, or confirm a hypothesis or theory. Experiment is the only way to transform hypothesis into a knowledge in a data-driven approach, any experiment starts with a hypothesis [26]. Then it is needed to determine a specific way to test the hypothesis, to understand what exactly changes in the product in order to test the hypothesis, select users, and the moment when they become part of the experiment. At this point, users are randomly assigned to test and control groups, the fairness of the experiment ensures random distribution and a similar experience for users who are assigned to groups.

Next, metrics are defined that will measure the impact of the change. Metrics should always be predefined. It is necessary to think in advance about the expected effect of the change or determine the level of effect that the company is interested in.

Knowing the expected effect, the required sample of users could be determined, as well as time and resources estimation required to conduct the experiment. For example, if it takes six months to get enough data, and the results of the experiment do not significantly affect the startup, it may be worth postponing the experiment.

At the end of the experiment design, it is necessary to evaluate what decision will be made based on the experiment results. If the experiment does not significantly affect the future of an IT startup, there may be no point in conducting it. In general view experiment template displayed in table 2:

Table 2 - Experiment template

Step	Description
Hypothesis	What and how shoud be achieved
What will be changed in product	How experience of test and control group will look like
Test group	Which users and then become a part of experiment
Key metrics	What should change, how we measure it
Effect	What result, do we expect
Action plan	Depending of result if If event A has occurred, do B, if event C has occurred, do D

2.3.2 Working with data

The data-driven approach is characterized by careful work with data. Since data is the primary source of product knowledge, the data itself must be carefully cleaned and the sources must be known. Data cleansing deals with the detection and removal of errors and data inconsistencies in order to improve data quality [27]. The data cleansing approach must satisfy several requirements. First of all, it must detect and eliminate all major errors and inconsistencies both in individual data sources and when integrating multiple sources. The approach should be supported by tools, constraining and extensible to easily cover additional sources. The factors that reduce data quality and require cleaning are:

1. Inconsistency
2. Missing values
3. Duplicates
4. Outliers and abnormal values
5. Data entry errors

Cleaning can be performed both at the time of loading data into the storage and within the analytics system.

2.3.3 Statistical approach

A statistical approach is required to ensure that results of the experiment wasn't just accident [28,29]. All results obtained in the experiment should be statistically significant. Statistically significant is an estimated measure of confidence that a result is not random. The result can be the difference in the distribution of two samples, the degree of difference or closeness of some statistical distribution from the normal, the value of the regression coefficients that makes the model "useful", and so on. Any result can be valid, i.e. really reflect the patterns of the problem under study or be the result of the influence of random factors, errors, etc. Therefore, to evaluate the results, statistical criteria are used that allow to assess the likelihood that the results are random. Results are considered statistically significant if the probability of their accidental occurrence does not exceed some generally accepted level. In many cases, this level is taken as 5 % (or less) of the probability of an accidental result. This means that if a given study will be repeated 100 times, then the random occurrence of the results would be expected in less than 5 cases.

In addition to statistical significance in individual cases, it is also important to use other statistical tools, remember to evaluate the sample, the distribution of this sample, and calculate anomalies. A statistical approach is important in order not to make mistakes in the conclusions, for example, if estimating the conversion from site visit to registration in two groups, with 1000 users in one group, and 10 users in the other group, the results of such a check are likely to be statistically insignificant. since it is cannot be said for sure how a group of 10 users will behave if it is expanded to 1000 users.

Therefore, it is necessary to use the confidence interval, sample size, mean and sample variance, and other statistical tools to evaluate the results obtained and check their reliability.

2.3.4 User research

In order to create a product for users, it is needed to know users. This could be achieved using various user research techniques [30]. User research helps understand the behavior, needs, and motivations of those for whom products are created. The type of research depends on many parameters, such as type of application, market, market size, etc. In each case, it is necessary to separately draw up a user research plan. To obtain information about what users want, the following user research methods are characteristic of B2B startups:

Table 3 - User research methods

Method	Description
Contextual interview	In such interviews, how the user uses the application in everyday life is observed. These are usually informal interviews that lack a pre-written list of questions. The main task of such an interview is to understand exactly how the user works with the application, where the user has difficulties, and how he or she copes with them.
Individual interview	Individual interviews with the user, which allow to understand the motivation of the user, the real reasons for certain actions, or to reveal deep problems. For such interviews, a list of questions is drawn up in advance. When compiling a list, it is important to remember that questions should be open-ended (yes / no questions may not give useful information), as well as about the so-called social desirability bias [31] - people most often tend to give answers that are expected from them, rather than answers that correspond to reality.
Surveys	Questionnaires include several questions with predefined answers or open-ended questions. Used to obtain quantitative data on interesting issues. With the help of such questionnaires, it could understand which user segments are present in the application and how often users use a particular function of the application.
User personas	Creation of a reference user based on available data and interviews. This profile of the user will reflect the main motivation of their segment, what pains this segment has, and how the user is configured to interact with the application. In each individual case, each person has their own motivation, but this tool allows to generally look at the average portrait of the user.
Customer journey map	This is the story of the client's interaction with the startup from the moment of realizing the need for repeated communication. It is compiled on behalf of the user, taking into account their goals, feelings, emotions, fears, values. This tool allows understanding who the target audience of the startup is and why this audience remains or leaves for competitors.

User segmentation	Dividing users into groups depending on their characteristics (for example, a month of registration, age, region, device, etc.). Segmentation can be used for many tasks: assessing user behavior, assessing the contribution of some users to the overall picture, etc. This approach allows making a more customer-oriented product. In some cases, the customer segment may be referred to as a cohort.

This is not a complete list, there are many more tools, the most commonly used tools are presented here, but for each startup, it is needed to separately understand and choose which tools will bring the greatest benefit.

This chapter gives a general view of the basic principles of the processes and methods used in the data-driven decision-making approach. The approach can be applied not only in startups but also in many other areas, for example, in education [32], manufacturing [33], or in large corporations [34]. The approach does not provide a ready-to-use guide but simply shows how decisions need to be made. Using this approach does not negate the use of other methodologies or frameworks, but only complements, making it possible to increase the likelihood that the consequences of decisions made will be exactly as they were originally intended.

After describing the essence of the approach, it is necessary to understand exactly how the principles, processes, and methods of the approach are implemented in the B2B startups. In the next chapter, we will look at metrics that can be used to track exactly how a startup is performing.

3.4. An overview of metrics for b2b apps startup

Metrics - the numerical value of a measured indicator. For each startup, depending on the desired goals, its own set of metrics should be chosen. There are many different classifications of metrics and their meanings, in any case, the main idea of any metric is to provide information about exactly how a certain process takes place inside a startup [35].

This chapter will look at metrics suitable for a B2B startup that builds an application, and suggest different classifications depending on the purpose of the metric. Since the data-driven approach assumes that the developer is not interacting with the product itself, but with its model, a way is needed with which the operation of this model can be refined. Each change in certain metrics allows to understand what exactly happened in the product, and then the management needs to find an explanation why this or that metric has changed, and what needs to be done to save or change this change. In fact, metrics are the language through which the product communicates information to its owner, so it is extremely important to choose those metrics that will answer the questions posed, and not just display unnecessary data. First, consider the general properties of metrics, the so-called growth metrics and product metrics [15].

2.4.1 Growth metrics and product metrics

Most metrics can be divided into 2 types: product metrics and growth metrics. Product metrics characterize the product itself, and show how well it performs its task. Growth metrics show the current state of the business. To understand these concepts, draw the following analogy. A startup is a black box that turns one resource into another resource (in this case, a startup turns customers into money), how well a startup turns customers into money is a product metric, and how much money leaves the startup and how many users it includes - growth metrics. Both growth metrics and product metrics are important for a startup, as growth metrics answer such general questions as:

1. Are incomes rising or falling?
2. What is the monthly audience of the product?
3. Amount of receipts created by establishments per month?

And product metrics allow to evaluate the product itself and answer questions such as:

1. How much money does each new user bring?
2. What percentage of registered users pay for a subscription?
3. How many solutions does one user use on average?

This classification is important because building a startup only on growth metrics could hide the fundamental deterioration of the startup. In essence, growth metrics are a function of the number of users and product metrics. For example, a metric such as LTV (Life time value) multiplied by the number of paying users allows determining total revenue over time. However, if LTV starts to fall and the number of users grows (for example due to marketing investments), tracking only revenue will not reflect this situation, and when the number of users stops growing, the startup management will see a sharp drop in revenue, although it was possible to predict this fall and take some action much earlier.

The opposite situation can occur if startup management focuses only on product metrics. The product can show excellent conversions and LTV, but the number of users can be very low, and therefore the profit will also be very low (this situation is possible when a very good product is created, but no distribution channel is found through which the value of the product will reach users). Therefore, when compiling a list of metrics, it is necessary to understand what type the metric belongs to and what the change in this metric indicates.

Most often, product metrics are relative and growth metrics are absolute. Not all metrics can be assigned to one or another group, but it is necessary to clearly understand at what point the metrics of growth are used, and at what point the metrics of the product. Next, the classification of metrics will be created.

2.4.2 Metrics classification

To create consistency, classify the metrics. It should be noted that this classification is not fundamental, and was created on the basis of an analysis of various literature and cases [15,20-25]. So, metrics can be divided into 4 groups:

1. Behavioral - metrics that characterize the general behavior of users within the product
2. Finance - metrics that somehow reflect the financial flows of a startup
3. Feature-based - metrics created to track specific functionality within a product.
4. Custom - other metrics that cannot be assigned to any of the above groups.

Consider examples of each of the groups starting with behavioral metrics. As already mentioned, such metrics reflect user behavior within the product. Consider the main metrics that belong to this category.

2.4.3 Conversion

Conversion (or conversion rate) - the ratio of users who performed the target action to the total number of users in the group. Conversion refers to product metrics, that is, it shows how well the product performs its task, regardless of how many users were attracted to the product. Conversions can be used to build the so-called conversion funnel - the change in conversions when going through all stages of the process.

For example, in a conditional startup that creates a messenger, one of the targeted actions is to send a message. At the same time, the management knows that the conversion from registration to sending a message is 35 %. It's just that it's difficult to draw any conclusions from this fact as to why the conversion is exactly like this. However, it is possible to compose a conversion funnel (fig.5) of the entire onboarding process - the period of acquaintance with the product and its main functions, in this case, onboarding is the registration of a user and sending the first message, since sending a message is the main function of the messenger. The main losses of users occur at the stages of providing access to contacts (17 pp) and at the stage of uploading a profile photo (30 pp).

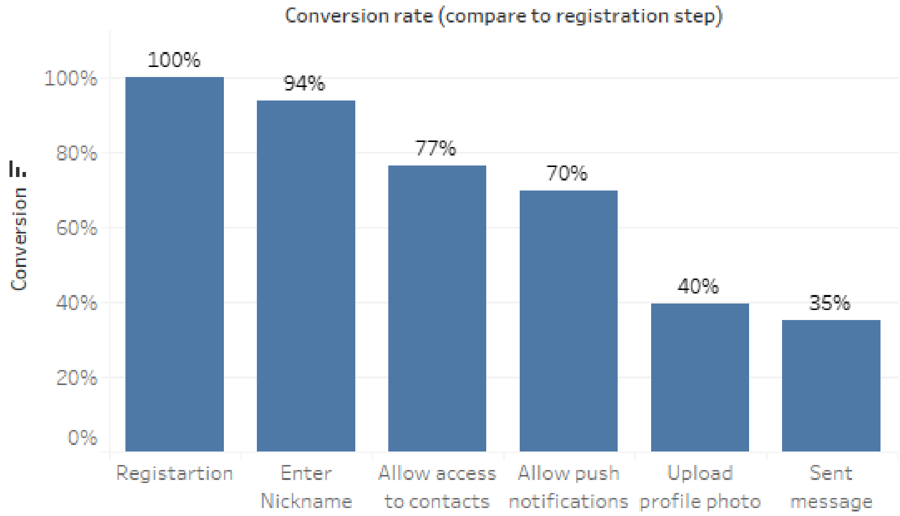


Figure 5. Example of conversion funnel

At the same time, without access to contacts, the application will not be able to perform its main function, and without downloading a profile photo, it will be able to. In this case, the management can experimentally check how much the conversion will change if the photo upload stage is removed. Thus, the conversion helps not only to understand how the product works, but also explains why something is happening in the product in a certain way. For each individual case, it is necessary to determine what is the target action of the group and what is the general action of the group.

Based on the conversion, various marketing indicators are built, for example, CTR -click through rate, in fact, it is a conversion from showing an advertising banner to clicking on an advertising banner. In any situation in which there is some sequence of actions, the conversion helps to understand exactly how users go through this the sequence of actions, the metrics themselves can be called differently, but the meaning will always come down to conversion. Based on this indicator, the startup management can build hypotheses about how the product works and test them.

2.4.4 Retention rate

This metric shows how new users come back to the application. There are 2 methods for calculating this metric, rolling retention and n-day retention [36]. The difference between these methods is shown in Table 4. In this case and hereinafter, the term app means any startup product (website, service, application, etc.).

Table 4 - Retention counting method

N-day retention	Rolling retention
$\frac{\text{Number of users who open the app the N}^{\text{th}}\text{ day after day 0}}{\text{Number of users who first used the app on day 0}}$	$\frac{\text{Number of users whose last day of activity is } on\ or\ past\ day\ N}{\text{Number of users who first used the app on day 0}}$

In fact, n-day retention is a special case of conversion (conversion from users who opened the application on day 0 to users who opened the application on day N). The main difference is that when calculating n-day retention on every day N, only users who opened the application on day N are taken into account, and in rolling retention if a user arrives on day N, it is counted on all days up to day N. Specifics of rolling retention lies in the fact that it is constantly changing (for example, if a user entered the 90th day, he will raise the retention rate of all previous days). For the vast majority of tasks and products, N-day Retention is better suited than Rolling Retention, but there are rare cases when Rolling Retention is more convenient. It is usually used for products that involve a fairly rare use (for example, booking hotels or buying equipment).At the same time, depending on the specifics of the startup, the granularity may change (retention is calculated not by day, but by hours, weeks, months, quarters, etc.).

It is also necessary to figure out what 0 and N day mean. These metrics are not associated with specific dates. 0 day means that on this day the user launched the application for the first time, and the subsequent retention is calculated for this particular user. For simplicity's, create an example: Consider user A and user B. User A first launched the application on April 1, then launched it on April 2, April 5, and April 7. User B first launched the app on April 5th and then launched it on April 6th, 7th, 8th, and 10th. Table 5 lists the specific days that users logged into the app. That is, for user A, day 0 (the day the application was first launched) is April 1, and for user B, April 5. For example, April 7th is the 6th day of interaction with the application for user A (user A opened the application 6 days after the first opening), and for user B it is the 2nd day. Despite the fact that a specific date is the same, for different users this date is different in a row of days of interaction with the application.

Table 5 - Retention rate calculation example 1

	Day of launch						
	0 day	1 day	2 day	3 day	4 day	5 day	6 day
User A	1 April	2 April			5 April		7 April
User B	5 April	6 April	7 April	8 April		10 April	

Using this table, it is possible to demonstrate exactly how the retention rate is measured, add 3 more users, C, D and E to table 4. To simplify the table, short date format will be used, result shown in table 6.

Table 6 - Retention rate calculation example 2

	Day of launch						
	0 day	1 day	2 day	3 day	4 day	5 day	6 day
User A	01.04	02.04			05.04		07.04
User B	05.04	06.04	07.04	08.04		10.04	
User C	10.04		12.04	13.04		15.04	
User D	15.04		17.04		19.04		
User E	28.04	29.04		01.05			

In order to get retention on the 1st day, it is needed to calculate how many users launched the application on the 1st day after the first launch and divide this number by the total number of users. In this case, the retention of the 1st day is 3/5 = 0.6 or 60 %. It doesn't matter that the specific launch dates on the first day were different for different users, the main thing is that these dates were the launch of the application 1 day after the first launch. Also, based on table 6, it is possible can clearly show the difference between rolling retention and n-day retention calculation. To do this, calculate the retention rate of each day in both ways and display it on the graph (fig.6). Calculating the retention rate of the 4th day, is equal to 40 % (2/5 = 0.4), since out of all 5 users on the 4th day, only two users (A and D) opened the application, but rolling retention of the 4th day is 80 %, since users B and C opened the application on the 5th day after the first opening and also got into the calculation of the 4th day (4/5 = 0.8).

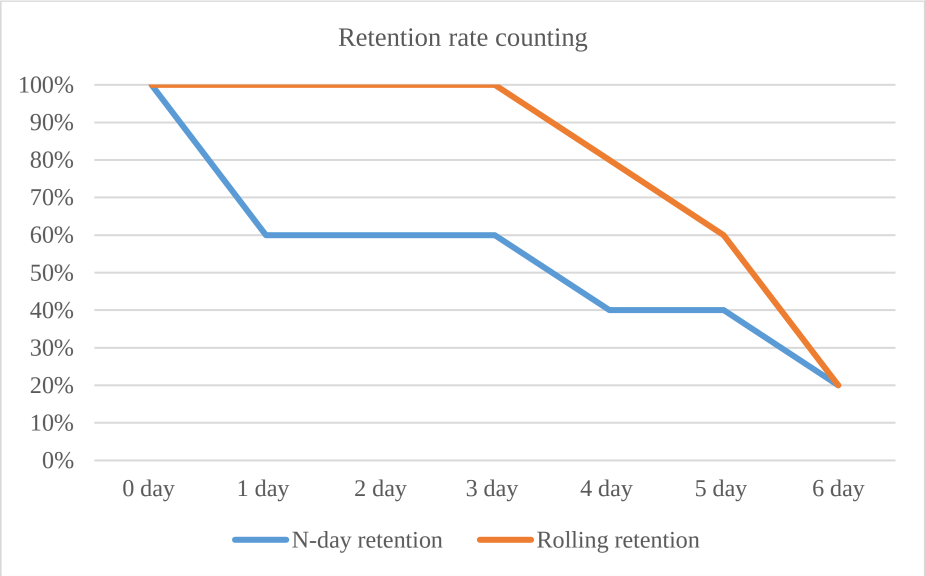


Figure 6 Retention rate methods counting comparison

In reality, there are usually many more users, but the counting method remains the same. The total number of users depends on the selected time period, for example, if calculating the retention rate for 1 month, then the total number of users will be equal to all users who first launched the application during this month. Consider user E from Table 6, despite the fact that he mainly opened the application in May, the first launch was on April 30th, so this user will fall into the segment of users for whom the retention rate for April is calculated.

There is also one more feature of the retention rate calculation - time windows. It is necessary to understand when for the user day N-1 ends and day N comes. For this, 2 different methods of calculating "the 24-hour window" and the "strict calendar day" are used. When calculated based on 24-hour window, retention for each user will be calculated based on individual time intervals. Day 0 of a specific user starts from the first start and ends 24 hours later, day 1 starts 24 hours after the first start and ends after 48 hours, etc. When calculated based on the strict calendar day, the user's day ends at the same time the calendar day ends. At first glance, the difference between these methods is insignificant but consider the following example: 100 users first entered the application at 23.50 on April 13, and then at 00.10 on April 14. If the calculation is made using calendar dates, then all 100 users will be counted in the 1-day retention rate, but when calculating based on the 24-hour window, such users will still only be counted in the 0-day retention rate. Calculation based on the 24-hour window gives a more honest answer, but this method of calculation will take more time, since the final value of the metric on day N can be obtained only after 2 calendar days (48 hours from the moment the last user arrived on day N). The date-based value for day N will be available the next day. If, when starting a startup, management focuses on market benchmarks, it is necessary to clearly understand how the retention rate is calculated in a startup, and how it is calculated in a benchmark.

Such a detailed explanation of the retention rate is necessary, since, in fact, this metric is one of the most important for startups. Using this metric, it is possible to determine if the value hypothesis on which the product is based is correct. For example, if the retention rate by 14 days becomes 0, this means that after 2 weeks of use, users do not find anything of value in your product and leave it. In this case, it makes no sense to spend money on marketing and further product development, since the product still cannot interest the user. On the other hand, if the retention rate is high enough, it means that the product retains the users and brings them some value. In this case, company can start scaling the product. Like conversion, retention refers to product metrics

2.4.5 Active users

This metric calculates the number of unique active users for a certain period. For example, DAU (Daily active users) counts the number of unique users who launched the application at least once a day, and MAU (monthly active users) calculates the number of unique users who launched the application at least once a month. These metrics also include average indicators, for example, the average daily audience - the arithmetic average of the daily audience for a certain period. These metrics refer to growth metrics and depend on the number of users who come to the application. For example, the MAU may be 10,000, and all those 10,000 launched the application in 1 day and never opened it again. In this case, the DAU of this day will be 10,000, and 0 on all other days. The average DAU will be 333 (10000/30 = 333). The ratio of the average DAU to MAU shows the regularity of the use of the product, the closer this ratio is to 1, the more regularly the product is used, if this ratio is closer to 1/30, this means that the product is used no more than once a month. Activity metrics help to quickly assess how many users are using a product. Based on this information, as well as other metrics, startup management can draw advisory conclusions.

2.4.6 Churn rate

Customer churn indicator is the percentage of startup users who stop being users within a certain period, this indicator in general is calculated by the formula [37]:

$$\text{Churn rate} = \frac{C_1 + C_3 - C_2}{C_1} * 100\ \% \qquad (1)$$

Where:

C_1 – amount of users at the beginning of the period

C_2 – amount of users at the end of the period

C_3 – amount of new users for the period

In each case, the formula can change, but the general meaning remains the same - to calculate how many users have stopped using the product for a certain period. Startup management also needs to understand what it means to "stop being a user", it can be unsubscribing, unsubscribing from email newsletters, or stopping the launch of an application. Churn rate is important for a startup, as the potential target market is not endless, and the more users have already tried the product and abandoned it, the more difficult and expensive it will be to attract new users.

The most popular behavioral metrics were indicated here, but the list can also be supplemented and modified depending on a particular startup. Next, consider a list of the main financial metrics. Financial metrics take into account all the cash flows that a startup has. it is difficult to distinguish such metrics separately since many metrics for their calculation require the calculation of other metrics. Therefore, start with the main financial metric Lifetime value, during the calculation of which other possible financial metrics will be considered.

Lifetime value (LTV) is the total value that the user will bring over the entire period of using the product [35]. Using this metric, it is possible to predict the future earnings of a startup. There are many formulas for calculating LTV, depending on the business model and the area in which the startup operates. In general terms, LTV is calculated using the following formula:

$$LTV = ARPU * AUL \qquad (2)$$

Where:

ARPU – average revenue per user

AUL – average user lifetime

However, this formula can be modified depending on specific needs and available data. At the same time, the calculation of ARPU and AUL indicators can also vary depending on the specific business model and market. Therefore, to begin with, consider the situations in which LTV is used. To begin with, consider a situation in which it is needed to understand how much value one user brought during their use in the past. This requires historical data and, in this case, LTV is calculated using the formula:

$$LTV = \frac{TR}{CQ} \tag{3}$$

Where:

TR – total revenue per period

CQ – customer quantity per period

Also, historical LTV can be calculated using the churn rate, for this, it is needed to divide the average income from the client by the churn rate. The advantage of historical LTV is that it is easy to calculate, but this approach only works if customers have similar preferences and remain in the product for the same period.

For modeling customer behavior, a predictive approach to calculating LTV can be used. There are many different ways to calculate predictive LTV, take a look at the most suitable for a B2B application. The calculation formula for a certain period, in this case, looks like this:

$$LTV = \frac{T * AOV * AGM * ALT}{CQ} \tag{4}$$

Where:

T- average amount of transactions

AOV - average order value

AGM – average gross margin

ALT – average lifetime

CQ – customer quantity

Consider each of the indicators. Starting with the average amount of transactions. A transaction is any in-app purchase; it can be a payment for a subscription, product or service. The average number of transactions is calculated using the formula:

$$T = \frac{OT}{P} \tag{5}$$

Where:

OT – total transactions amount

P – period

Knowing the average number of transactions, it is possible to calculate the average order value, this indicator displays how much was spent on average in one transaction and is calculated using the formula:

$$AOV = \frac{TR}{T} \qquad (6)$$

Where:

TR – total revenue

Next, it is needed to calculate the average gross margin (AGM), which shows how much of each sale is the actual profit, and how much is the cost. This indicator is expressed as a percentage, the formula for this indicator is:

$$AGM = \left(\frac{TR - CS}{TR}\right) * 100 \qquad (7)$$

Where

TR – total revenue

CS – cost price

The last indicator that needs to be calculated is the average lifetime (ALT), this indicator displays how much, on average, one user uses the application. Calculate this indicator, using churn rate (formula 1), while the churn rate must be calculated for the same period for which the LTV is calculated, and then calculate the indicator using the formula:

$$ALT = \frac{1}{Churn\ rate} \qquad (8)$$

The formula for predictive LTV (formula 4) takes into account more factors than the formula for historical LTV (formula 3). However, it should be borne in mind that this formula can also be misleading since it is only a forecast. For a more accurate result, the calculation should be adjusted according to the industry and business model. All indicators (metrics) mentioned in the formulas, if necessary, can be used as separate metrics.

This section did not consider such metrics as inner rate of return (IRR), return of investment (ROI), net present value (NPV), etc., since these metrics mainly serve to determine the investment attractiveness of a project, and not only depend on how well the product does its job, but these metrics can also be tracked if needed. Another source of financial metrics can be unit economics - a modeling method that helps to determine the profitability of a business by calculating the profitability of one business unit (this can be a product, a client, etc.). the previously discussed LTV metric also applies to the unit economy.

The unit economy also has many different metrics, and the calculation of these metrics varies depending on the type and business model of the business. The point of calculating the unit economy is to understand how much money is spent on attracting one client and how much this client ultimately brings in money. The amount of money that a user brings is the LTV metric already discussed earlier; the client acquisition cost (CAC) metric is used to calculate the cost of acquisition. This metric does not have a strict calculation formula and includes various costs of attracting a user (marketing expenses, salesperson's salary, expenses for conducting promotions, etc.).

At the stage of a large campaign, when all such expenses are taken into account in advance and included in the budget, it is better to use the “standard” metrics of NPV, ROI, etc. However, for startups, it is extremely important to control all costs, since startups are usually created with the money of investors, so the unit economy approach allows to determine in advance how much at what price and with what conversion it is necessary to attract users in order for a startup to achieve the desired level of profitability.

It is extremely difficult to list all the financial metrics, since it has already been said that each startup must be considered separately. The main task of such metrics is to give the startup management a clear picture of how the financial flows within the startup are arranged and based on this task, each startup leader should choose specific metrics and how they are calculated.

Now consider feature-based metrics group. Such metrics track the status of not the entire product, but some specific functionality within the product. This can be tracking the frequency of use of the functionality, and the time of use of the functionality, and any other measurements of functionality. Based on such metrics, behavioral and financial metrics can also be calculated. For example, if a startup operates on a subscription model, then tracking the number of payments for subscriptions will also allow estimating income. Also, such metrics are used in the construction of user scenarios - a visual schematic construction of how the user solves his problem using the product. The conversion example in Figure 5 is also based on the onboarding scenario. Startup management cannot record the screens of all users and track their reactions at the time of using the product, however, using user scenarios, it is possible to draw up behavioral patterns of users, and based on this information, draw conclusions about how and why users use the product. A special case of user scenario is the use case of a diagram in UML notation. Such diagrams allow to model user behavior within the product, to understand what steps the user takes depending on the goal he wants to achieve.

Feature metrics help in building use case diagrams since with their help the creator of the diagram can evaluate what actions the user performs and display them using the diagram, or vice versa, check that the steps described in the diagram correspond to reality and the user actually interacts with the product in the same way. as described in the diagram.

The last group of metrics is custom metrics. These metrics cannot be attributed to any of the above groups, since they are specific to each individual startup or task.

It must be remembered that metrics is just a tool to track what happens to the product. Tracking metrics alone is not enough to effectively manage a startup. It is necessary to clearly understand what this or that metric means, the reasons why it may change, and what to do in case of this or that change.

3.5. An overview of the technologies used for data-driven analytics

In order to track metrics and draw conclusions based on them, the startup management needs technologies with which this tracking can be implemented. For this, BI systems can be used - a set of tools and technologies for collecting analysis and processing data. Figure 7 shows the general concept of BI systems [38]:

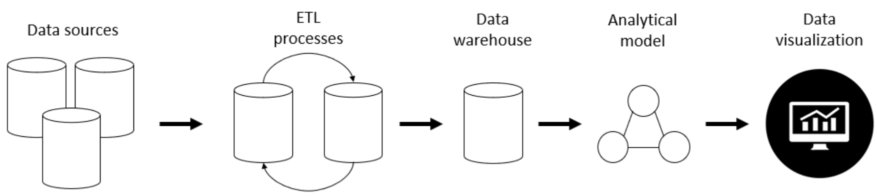


Figure 7 Business intelligence platform concept

Consider the main components of the system in the context of a startup developing an application in the B2B market:

1. Data sourses: Data sources can be different. For the application, the data source is the application itself, as well as sources such as CRM systems, marketing analytics systems and other systems that are integrated into the application and store data.
2. ETL processes: ETL is stand for Extract-Transform-Load. This term refers to the process of loading data from a source into a data warehouse from various data

sources.

3. Data warehouse: the final database that stores all the necessary information for further analysis.
4. Analytical model: a data model generated in a BI system, thanks to this model, the system understands how different data relate to each other.
5. Data visualization: transforming data from a set of values into human-readable tables and graphs.

Such systems can be used to track metrics. Because in order to draw conclusions from the data, firstly data should be somehow visualized. Currently there are many systems on the market that belong to the BI class. Figure 8 shows the ‘magic’ Gartner quadrant for 2021:

Figure 8. Magic quadrant for analytics and Business Intelligence platforms

The Gartner Quadrant allows to graphically display the situation on the market and evaluate the capabilities of products and their manufacturers. The quadrant is divided into 4 sectors: challengers, leaders, niches players, visionaries. The horizontal axis indicates the degree of completeness of the vision of the company’s business environment through the implementation of the BI system. Vertical displays the metric "ability to implement" in a particular business environment of the company [39]. Analyzing the Gartner, we can draw the following conclusions:

1. As of February 2021, the market leaders are Microsoft (power BI), Tableau and Qlik.
2. Domo, MicroStrategy and Looker are common in the market, however, according to Gartner, they do not have a clear strategy for the further development of the platform.
3. TIBCO software, ToughtSpot, Oracle, SAP and SAS understand what the market needs, but are not widely used.
4. IBM, AWS and Alibaba cloud are niche players with a focus on a specific area.

The choice of a specific BI system depends on the purpose for which it is being implemented. In a data-driven decision-making approach, one such goal is to track metrics and change them. But also, BI systems can serve other purposes. Since the product is in the center, first it is needed to understand what data and how should be visualized, and then determine which of the visualization and analytics systems is best suited. Also, in addition to BI systems, platforms for analytics of mobile applications can be used. They are a subset of BI systems and essentially perform the same function - data visualization. However, such platforms have a number of peculiarities associated with their use in mobile applications.

When choosing an analytics system, firstly draw up a list of tasks that will be solved with its help. Systems and tasks can be divided into the following types:

1. Marketing analytics systems - tracking traffic sources, calculating the cost of acquisition.
2. Product analytics systems - analysis of user behavior in the product and the impact of product changes on the user.
3. Advanced data analysis systems - deep product analytics and predictive model building.
4. Monitoring systems - tracking the main indicators of the product at different levels of functioning.

Take a closer look at each of these types below:

2.5.1 Marketing analytics system

Marketing requires tracking traffic and understanding how it flows through the acquisition funnel. Tasks that are solved using marketing systems:

1. Determination of traffic source, registrations, installations and clients.
2. Customer Acquisition Cost Calculation (CAC).

3. Calculating the return of funds from the marketing budget.

Marketing analytics tools for web services and mobile apps are different. Start by looking at tools for web services analytics. The main tasks for web services analytics are solved using end-to-end analytics. End-to-end analytics - combines data on all touchpoints with a client (from clicking on an advertising banner to paying for an order) and shows the profitability of marketing channels, taking into account sales, and not just requests. End-to-end analytics tools work on the following principle:

1. Uploading cost data from advertising systems (Google ads, Yandex Direct, Facebook Ads, etc.).
2. Uploading data on targeted actions (registration, payment for orders, etc.) from a database or from a CRM system.
3. Consolidation of information on any common parameter in a single database.
4. Creation of the necessary reports based on the received data.

End-to-end analytics services include services such as Roistat, Rick.ai, Owox BI, etc.

In addition to ready-made solutions, startup management could also assemble their own, since the principle of operation for all end-to-end analytics services is approximately similar.

Next, services for marketing analytics of mobile applications will be analyzed. The main difference between such services is that in applications, between the user's click on the advertisement and the installation of the application, there is an intermediary - an application store. In this step, information about the traffic source is lost.

To solve the problem of determining the source of traffic for new users of mobile applications, they use special systems for analytics and attribution of mobile traffic, such as AppsFlyer, Adjust, Brunch.io. The task of determining the traffic source in mobile applications is more difficult than in web services. In a simplified form, application analytics services work as follows [40]:

1. An advertising link is created in the analytics system interface, which is then inserted into the advertising banner.
2. When a user clicks on a link, he initially goes to the analytics system's redirect service, and only then to the app store. For the user, this happens seamlessly, that is, it looks like a simple transition to the application store, but at this stage, the analytics system stores information about the user and the source of his traffic.

3. The user installs the application and launches it, at this moment the pre-integrated function sends an installation event to the analytics system server with information about the user.
4. The system finds a match between those who went through the redirect service and information about the new user and determines the traffic source.
5. The system also receives data on the costs of advertising campaigns from advertising systems.
6. The system links data about ad clicks, app installs, costs and revenues from in-app purchases.

Thus, application marketing analytics systems allow to determine from which traffic source the user came to the application, and then, based on this information, determine which marketing channel is the most effective in terms of investment per paying user. Startup management needs to know exactly how a particular system works in order to integrate the necessary functions into a startup in advance, as well as understand how certain data is generated.

2.5.2 Product analysis systems

Product analysis systems help the product team to solve such problems as:

1. Search for obstacles on the way of the user to the solution of the problem in the product.
2. Evaluation of the popularity of various product functionality.
3. Measuring the effects of changes made.
4. Evaluation of experimental results.

Such systems allow to automatically build a sequence of user actions, calculate the retention rate, and other necessary metrics for product analysis. These systems include Amplitude, Mixpanel, Woopra, Heap Analytics, and other systems. Typically, such systems are integrated into the product using the SDK - software development kit. An SDK is a development kit that allows software engineers to create applications for a specific software package, software underlying development tools, hardware platform, computer system, game consoles, operating systems, and other platforms [41]. To simplify this definition, the SDK is a package of libraries so that the client can easily and quickly get started with a particular system.

2.5.3 Advanced analytics systems

Such systems are necessary for deeper analytics tasks:

1. Creation of custom reports.
2. Finding the correlation between actions in the product.
3. Building predictive models.

To solve such problems, BI systems are required, as well as knowledge of programming languages. The most popular programming languages for solving analytical problems are Python and R [42]. The decision of how the advanced analytics system should be arranged should be made by the management of the startup, depending on the area, business model, convenience and other factors, which are also determined in each case.

2.5.4 Monitoring system

Such systems solve the problem of monitoring the main metrics of a startup. As a monitoring system, both the systems mentioned above and any other systems can be used, depending on the decisions of the startup management.

In this section, an overview of the tasks that are solved using analytics systems was made, as well as examples of systems that can solve such tasks.

Thus, based on the information from the chapter "Data-driven decision-making approach", outline a general approach to managing a startup. First, it is needed to define a hypothesis for the value of the product. Build a product model based on the value hypothesis. Then, through experimentation, the product model is continually refined. Experimental results are tracked using a variety of metrics that are visualized using a variety of technologies. The next section discusses the practical application of the data-driven management approach. This requires:

1. Select and analyze a startup.
2. Create an environment where data-driven decisions can be made.
3. Start making data-driven decisions.

4. ANALYZED STARTUP DESCRIPTION

This master thesis will consider the Russian startup "Restik". The startup operates in the B2B market and offers ready-made solutions for automating the restaurant business. Also, since “Restik” could be also used by different types of leisure facilities such as time cafes, hookah bar, and other leisure facilities, which couldn't be strictly named as 'café', further term "establishment" for any type of café, which uses Restik, will be used.

The startup was launched in 2018. In this year, 54 federal law began to be changed in Russia. One of the points of the change was the requirement for all catering establishments, according to which, for any calculations, they must not only print a fiscal receipt but also send an electronic version of the receipt. Initially, the startup focused on small establishments that, since the law came into force, in addition to a fiscal receipt printer, needed a way to send electronic receipts The value- hypothesis was that a simple solution that automates this process would be in demand in the market. Experimentally, no one confirmed this hypothesis, only a market survey was carried out.

The startup offers 3 automated solutions for the restaurant business. Consider what was implemented in each of these solutions:

POS system: Point of sale (POS) system, consist of 2 components: web version and application version. Using web version manager could manage all cafe information and get statistics about working flow. The application will allow waiters to create orders on the device and automatically send them to the kitchen. Restik will also allow them to control product inventory and automate purchases. This application does not require the purchase of additional equipment and, unlike expensive market solutions, will allow cafes, regardless of size, to automate work processes at a low cost. In addition, the POS system has a built-in ability to create loyalty programs. For each visitor of the establishment, it allows to create own profile, and use various discounts and promotional codes for this visitor.

Delivery site: This system allows the owner of the establishment to create a ready-made online storefront for accepting online orders. With the help of the showcase, the visitor can select the necessary items, the delivery method, and also the payment method. The order can be sent to the web application, telegram bot, or by email, depending on the preference of the institution. Depending on the availability of the item, an owner can change the menu that the visitor sees. This solution also allows creating loyalty programs and promotions.

E-menu: This function allows the owner of the establishment to create an online menu with access by a link or QR code, the visitor of the establishment can use this menu, as well as the standard "paper" menu, the online payment function is absent in this solution, however, when the module "delivery site” is also connected, such a function appears. The owner of the establishment, just like on the delivery site, can change the menu depending on the available items. If the establishment already has any other module connected, the menu can be generated automatically.

Initially, the startup implemented only the POS system function, however, due to the COVID-19 pandemic in early 2020, the restaurant solutions market has rapidly changed, and business owners have a need to quickly change the familiar model. In the wake of this, the startup also had to change.

At the beginning of 2020, the number of customers dropped sharply, some establishments went bankrupt due to the lockdown introduced in the Russian Federation, and ceased to be Restik clients, other establishments no longer needed a POS system, as they began to work in delivery mode or “take-away”.

The management of the startup made a decision to change the current business model to meet the rapidly changing market conditions. A review of existing solutions for managing startups was made and the choice fell on data driven decision making. To implement this approach, management began to move towards creating an environment within which these principles could be implemented.

4.1. Current service and technologies schema

To create an environment for data-driven management, it needs to understand the internal scheme of Restik services, what data is already collected inside the startup. The scheme of Restik services is shown in Figure 9. There are only 3 entities that can interact with Restik: terminal (works only with devices on IOS), administrator's web application (admin), and users of the delivery site or electronic menu (clients). Restik uses both its own written services and third-party services integrated into the system.

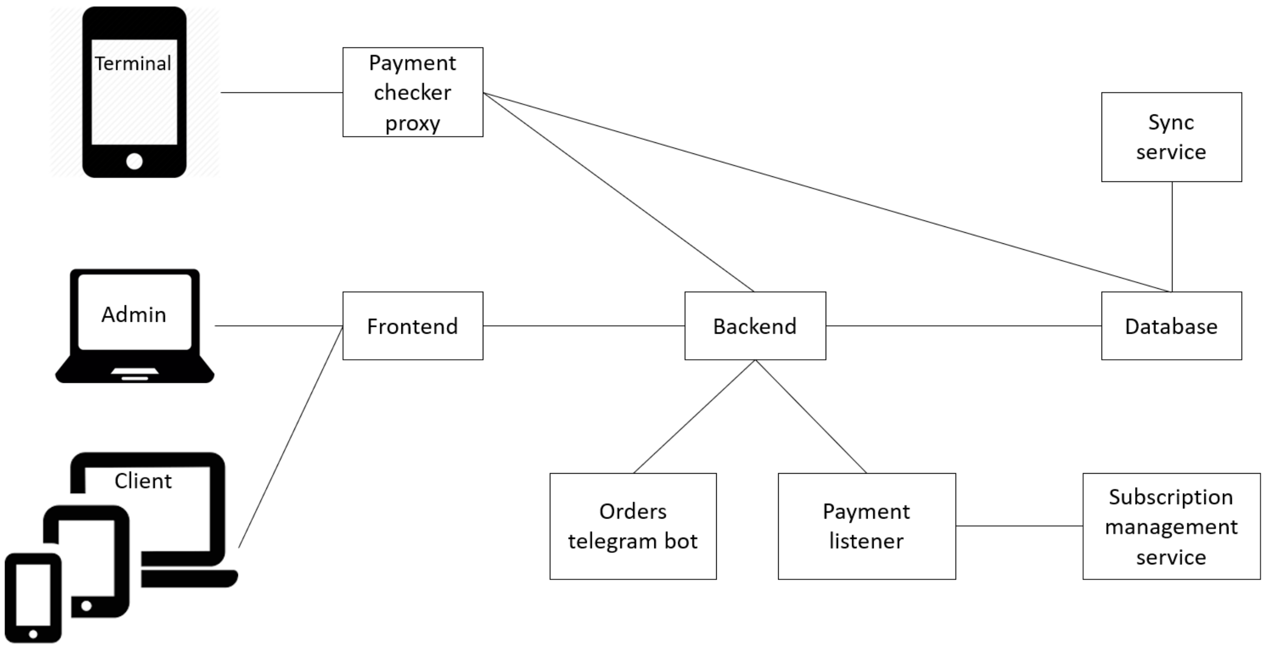


Figure 9. Restik services schema

The purpose of each of the services:

Table 7 - Services description

Service	Purpose
Frontend	The client side of the Restik UI with which the admin and client interact. Includes a marketing site with a description of the Restik features, a web application for the admin, an online showcase for the delivery site and an electronic menu.
Backend	Restik hardware and software part, which is responsible for the functioning of the internal part of the frontend and providing access to the database for the terminal. Implements such functions as creating a menu, creating a receipt, creating a warehouse, subscription management, management of loyalty programs, analytics and other functions provided to users of various Restik solutions
Database	The repository of all Restik data, including information about all payments by Restik users and payments by clients of institutions.
Sync service	A service that synchronizes data in a database, this service is needed because restik uses several DBMSs that need to be synchronized with each other (more about why this is needed in the section with a description of the technologies used)

Payment checker proxy	This service determines the status of a subscription using a database query. If the subscription is active, the service provides access to the function of creating new receipts. If the subscription has expired, the owner of the establishment can still see the analytics for past orders, add new items to the menu and change the warehouse, however, to create new receipts, establisment need to pay for a subscription.
Orders telegram bot	The service provides the necessary information to the telegram bot, with the help of which the owner of the establishment can find out about the orders received through the delivery site.
Payment listener	With the help of this service, information is transmitted to the backend that this or that client has paid for the subscription. More about the purpose of this service
Subscription management service	The service manages subscriptions, automatically charges money for a subscription, and also stores information about when a particular client's subscription expires.

Since the purpose of some services is not obvious, it is necessary to consider the services at the level of the technologies used in order to understand the reasons for the appearance of a particular service, taking into account the technologies and frameworks used, the Restik scheme shown on figure 10:

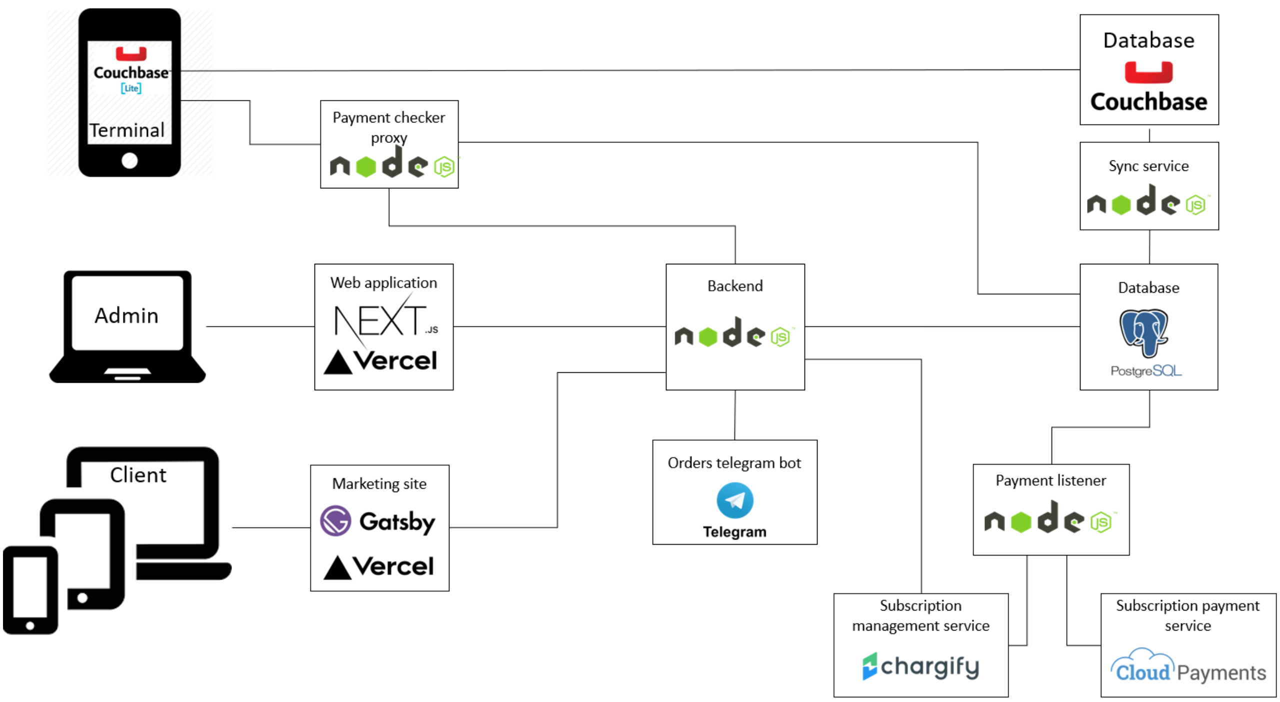


Figure 10.Technology usage schema

Consider the purpose of each of the technologies:

Table 8 - Technologies description

Technology	Purpose
Node JS	An open source, cross-platform JavaScript runtime that runs JavaScript outside of the browser.
Next JS	Open JavaScript framework built on top of React js, which allows to render a page on the server side, due to this, the page processing speed increases, as the load on the user's device is reduced.
Gatsby	A static site generator that makes fast web pages. Gatsby is written in React and is part of the framework of this library. That is, in fact, Gatsby is a React framework with which can quickly build and publish a React application.
Vercel	Web hosting that supports geo-load distribution and automatically conducts load testing, so Restik developers do not need to independently check each new build for load.
Telegram	Messenger with the ability to create chat bots. With the help of such chat bots, the owners of the establishment can receive all the necessary information in the messenger.
Couchbase	NoSQL DBMS with support for master-master replication, together with Couchbase Lite, supports the synchronization of all receipts of the establishment if the Internet connection is lost.
Couchbase Lite	The version of the Couch base that is stored on the client side. This database is part of the Restik client application.
PostgreSQL	Free object-relational database management system. Supports both relational database paradigms and object-oriented approach.
Chargify	Web application designed for automatic subscription management. It independently monitors the subscription status, sends notifications and debits money from the client (however, the function of debiting money does not work on the territory of the Russian Federation)
Cloud payments	A web application designed for online payment on the territory of the Russian Federation, with the help of this application, the owners of the establishments pay for the subscription, and the clients of the establishments can pay for the online order.

Payment listener appeared due to the fact that Chargify does not support payments in Russia. Because of this, the following sequence of actions is used (fig. 11) described in UML notation [43]. At the moment when chargify detects that the user's subscription has expired, information about this is sent to the payment listener, and information is sent to the backend that it is necessary to block the ability to create new checks or an electronic menu for this client. Then the Payment listener sends a message to cloud payments that the user needs to charge the money for the subscription. After confirming the payment, information about the payment is recorded in the database, and a message is sent to chargify that the subscription has been paid.

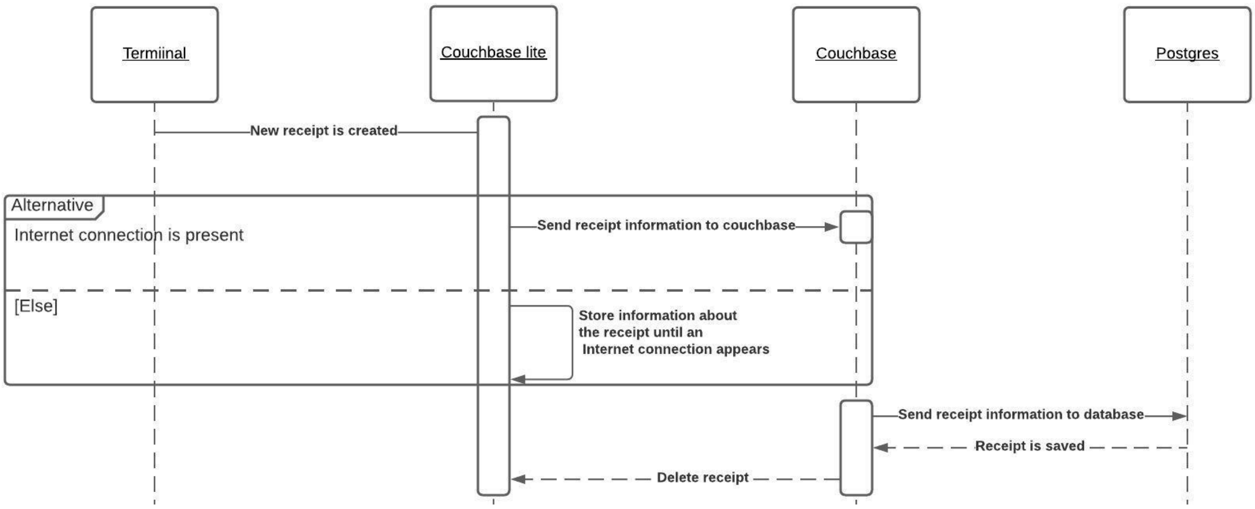


Figure 11. Sequence of subscription payment

Now consider at the purpose of the sync service. Since one of the features of Restik is the ability to create receipts without the Internet, it is necessary to somehow support this feature. This is done with the couch base (Fig 12.).

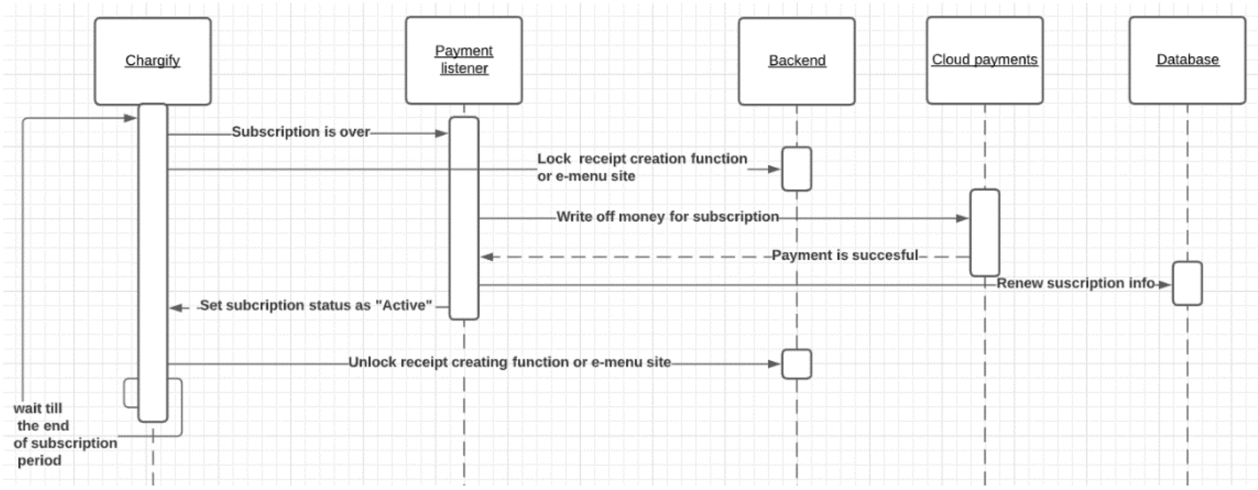


Figure 12.Database synchronization schema

The couch base lite is installed on the client's device, which is a database that is stored on the client's side. At the moment when a new receipt is created in the establishment, information about this is recorded in the couch base lite on the client-side and at the expense of the replication master, the same information is written in the couch base on the service side. When the receipt is closed the couch, base writes the receipt to the database and deletes the receipt information on its side and on the client-side.

Since the information about the receipt is stored in the database, the owner of the establishment can access it from the web application, but this eliminates the need to store information about all receipts on the client's side, only information about the currently open receipts is stored on the client's side. Thus, if the establishment loses its Internet connection, the couch base lite still stores information about all receipts that were opened during this period of time and deletes this information only after it is synchronized with the couch base on the service side and receives a message stating that all information has been successfully saved on the server.

4.2. Formulating requirements for data-driven environment

To describe the requirements for a data-driven environment, it is needed to understand how customers interact with r application so that then find insights and improve the product based on those insights. Customer's interaction with the app could be observed using metrics. In order to obtain these metrics, it is necessary to compile a list of events on the basis of which the metrics will be calculated. To track events, startup management need to understand at what point these events occur within the service and with what technologies these events can be tracked. Thus, the list of questions that need an answer looks like this:

1. What are the key metrics to track?
2. What events do these metrics display?
3. At what point in time do these events take place?
4. How to track these events?

In order to determine the metrics, define what is the main task for a startup. This startup is commercial, which means that the main goal is to make money. A startup makes money with it. that users pay for a subscription, then one of the metrics will be the number of payments by month. Since the number of payments is a metric of growth, it is also necessary to choose a product metric that will indicate how well the product converts the user into money. For a B2B startup, such a metric is the conversion from registration to payment (conversion is the ratio of the number of users who performed the target action to the total number of users), as well as LTV. Thus, the key metrics for a startup will be:

1. Conversion
2. Revenue
3. LTV

In addition to key metrics, it is also necessary to define additional metrics with which it will be possible to judge the behavior of users in the product and, based on this, find new points of growth.

The additional metrics for this startup can be globally divided into 3 categories: finance, decisions, and user behavior. Financial metrics include such indicators as the average income per client per month, the average number of payments from one client (since a client can use several solutions, this number may be more than one) and the conversion from site visit to payment. The solution category includes metrics specific to each solution; for POS systems, this is the number of checks per establishment, the average check of establishments, the total number of checks, and the total revenue of establishments. For a delivery site - the number of delivery orders, the average bill for delivery orders, the number of establishments with delivery orders. Customer behavior includes retention rate, churn rate, conversion from site visit to registration, and the time from registration to payment.

All the metrics presented above take into account only users who have registered in the product, however, for a complete understanding of how the product works, they are not enough, metrics are also needed that will track exactly how the user comes to registration. To do this, we need to know the marketing channel from which the user came and the conversion from this marketing channel. This list of metrics can also be supplemented depending on the received insights, or when adding new functionality to existing solutions.

4.3. Defining events

After determining the required metrics, determine which events within the application will be used as these metrics. To do this, draw up a use case [40] diagram of the 3 main actors of the product: the manager of the establishment, the employees of the establishment and the clients of the establishment. Start with the establishment manager (fig. 13):

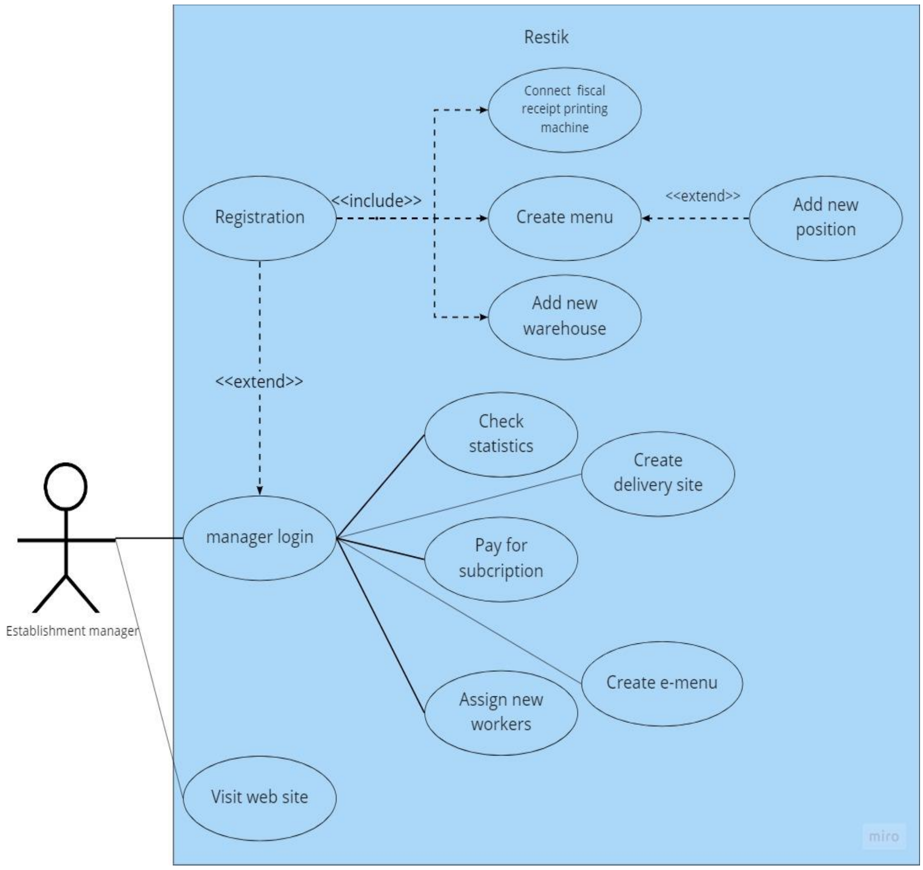


Figure 13 Entablement manager use cases

The manager (owner) can register an establishment, add a menu, new menu items and a warehouse with components. In addition, using the web application, the manager can check the statistics of the institution, create new employees and pay for a subscription. The manager can also create an electronic menu or a delivery site; for each solution, payment is made separately. Using these actions, we can calculate the conversion (number of unique payments / number of registrations), LTV (total amount of all payments / number of unique users) and other metrics. The final list of all metrics with the calculation method will be indicated below.

Next consider the use case of an employee of the establishment (fig 14):

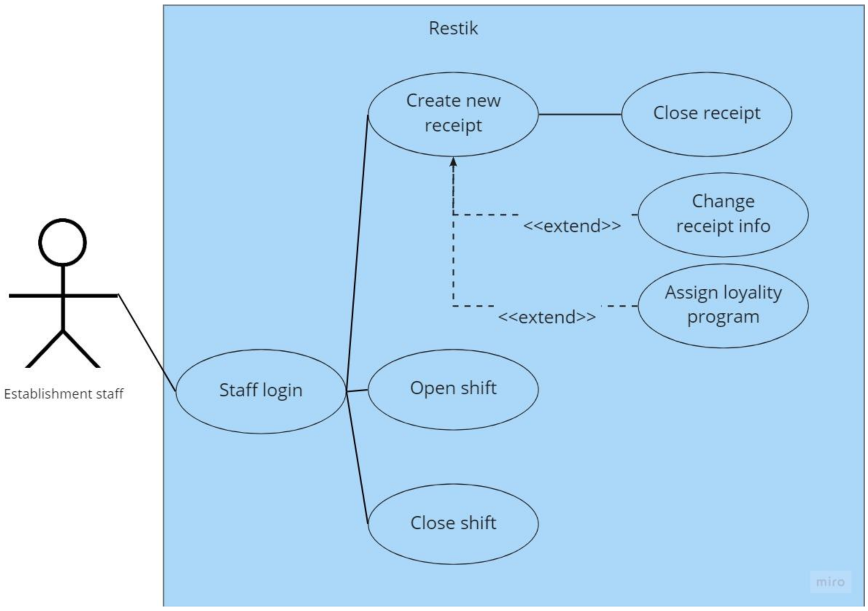


Figure 14 Establishment staff use cases

Of the interested events, one can single out the closing of a receipt and the addition of loyalty programs. Thus, it will be possible to find out how many checks were for the establishment, the average amount of these checks, and also whether the loyalty program was applied.

The last remaining actor is a client of the establishment, the number of actions that he can perform is relatively small (fig 15):

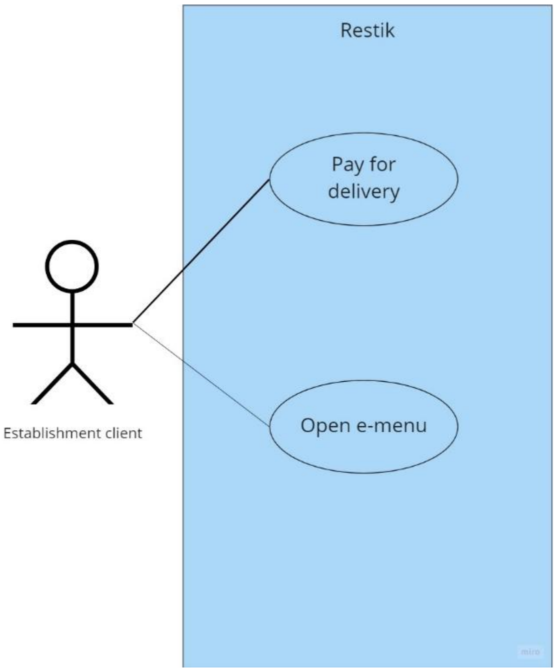


Figure 15 Establishment client use cases

However, knowing this information, the average bill for delivery orders, the number of establishments with orders, and the average number of orders for delivery could be measured.

After analyzing all the events that occur within the application, draw up a final table with a list of metrics and a method for calculating these metrics.In the calculation methods, aggregating functions will be used, which will be further implemented by technical means, these are functions:

1. SUM () - calculates the general characteristics of a specific event (for example, the total sum of all receipts)
2. COUNT () - counts the total number of certain events (for example, the number of checks)
3. UNIQ () - Counts a unique number of specific events.

Thus, knowing all the events and having the ability to apply various aggregated functions to them, get the table 9 containing list of metrics with the calculation method:

Table 9 - Final list of metrics

Metric	Counting method
Registration -> subscription payment conversion	COUNT(Pay for subscription) / COUNT(Registration)
LTV	SUM(Pay for subscription) / UNIQ(Pay for subscription)
Revenue	SUM(Pay for subscription)
Churn rate month N	UNIQ(Pay for suscription month N) / SUM(Pay for subscription mont N-1)
Month revenue per user	SUM(Pay for subscription month N) / UNIQ(Pay for subscription month N)
Average amount of payments per user	COUNT(Pay for subscription) / UNIQ(Pay for subscription)
Web site -> registartion conversion	UNIQ(Visit web site) / COUNT(Registration)
General amount of receipts	COUNT(Close receiprt)
Average amount of receipts per establishment	COUNT(Close receipt) / COUNT(Registrations)
Average receipt price per establishment	SUM(Close receipt) / COUNT(Registrations)
General establismnets receipts revenue	SUM(Close receipt)
General amount of delivery orders	COUNT(Pay for delivery)
Amount of delivery orders per establishment	COUNT(Pay for delivery) / COUNT(Registrations)
Average delivery order price per establisment	SUM(Pay for delivery) / COUNT(Registrations)
Amount of establishments with delivery orders	UNIQ(Pay for delivery)

After determining the final list of metrics and methods for calculating them, it is necessary to choose the technologies with which this will be implemented.

4.4. Defining technology stack

After defining the key metrics, it is necessary to choose the technologies with the help of which it will be possible to collect and visualize data, on the basis of which the analytics will be conducted. After studying the market, the following software products were selected.

4.4.1. Amplitude

Product analytic system that allows to track the behavior of a registered user within the product, the system perceives any user interaction with the product as an "event", this event can have many different characteristics, and the user himself has many different characteristics. As a result, using this analytics system, it is possible to track the number of users who have performed a particular action within the application, and what unites these users.

To integrate Amplitude into the product, there is an SDK that is embedded in the product code. Before starting work, list of events with characteristics, as well as a list of user characteristics was created. Based on the documentation, the following list of events and their characteristics was compiled:

Table 10 - Amplitude event taxonomy

Event Name	Property			
	Name	**Value type**	**Name**	**Value type**
Subscription Payment	Subcription price	integer	Type	"POS", "Delivery site", "E-menu"
Receipt_closed	Receipt sum	integer	Payment type	"Card", "Cash", "Bonus", "Mixed"
Registration				
Delivery_order	Delivery sum	integer	Delivery position amount	integer
Online showcase created				
Electronic menu created				
Card Authorization				

Inside the application, individual events, unique events, and also aggregate properties of events could be counted. This application is used to regularly monitor metrics.

4.4.2. Google Analytics

Analytics system that allows to track the interaction of any user (not only registered) with the application's marketing website. Using this system, it becomes possible to trace the path from the site visitor to registration in the application and paying for the subscription. Analyzing this data, it is possible to understand at what stage users are eliminated, what prevents the user from becoming a client, and which versions of the site are more successful. Within this system, it was decided to perform a cohort analysis to calculate the marketing budget. Cohort analysis involves dividing all users into a group according to some principle and then analyzing each individual group. In this case, users were divided into groups based on the month in which the user first visited the website. Because of this, it will be possible to correctly calculate the marketing budget.

Give an example of how this happens - inside google analytics, it is possible to determine the source from which the user came to the site, as well as track the registration and payment of each individual user. Further, all users are divided into cohorts. Within each cohort, we can count the number of users who paid for the subscription and the total payment revenue. Next, consider the marketing budget for each marketing channel and see how many users this channel brought. For example, 10,000 rubles were spent on marketing on one of the channels, this channel brought 6 paying users, while on average each user brings 1,000 rubles per month, which means that the advertising budget for this channel will be recouped in 2 months (6 * 1000 * 2 = 12000). Based on this data, startup management can direct the budget to the channels that are most effective in terms of the rate of return on investment.

Also, this application allows estimating the number of users who came to the site and their geography. Together with amplitude, this allows to create end-to-end analytics and understand the entire path of the user from visiting the site to paying for a subscription, analyze what difficulties the user faces at each stage and take measures to reduce these difficulties.

4.4.3. Python

In addition to ready-made software products, the Python programming language was used for data processing. Python was chosen since there are many libraries for data processing and visualization. The following modules were used in this startup:

1. Pandas [44] is a library for working with data, which allows representing of data in the form of tables (DataFrame), and convenient ways to interact with these tables (clear data, perform additional calculations, filter data by a certain criterion, etc.)
2. NumPy [45] – a library for mathematical data processing. From time to time it was necessary to calculate certain statistics. For mathematical processing, the NumPy library was used, with the help of which it was possible to quickly calculate the statistical reliability of the data obtained, build a theoretical and empirical distribution of a certain

value and compare the difference, and other mathematical operations necessary to obtain high-quality data on the basis of which conclusions can be made.

3. Dask [46] – a library similar to Pandas, but able to parallelize data processing on a processor, as a result, for large amounts of data, a significant acceleration of processing is achieved, compared to Pandas.
4. Matplotlib [47] - library for data visualization. In the course of the work, it became necessary to create custom visualizations for which there are no templates in other data visualization systems. For custom visualizations, it was decided to use Matoplotlib.

After defining the technology stack, measures were taken to introduce the necessary SDK inside the application, as well as reorganized the business decision-making processes and introducing new functionality into the application, then 2 cases will be considered, the realization of which became possible thanks to a new approach to working with data.

4.5. Overview of the implemented solution

After the implementation and configuration of all technological solutions, the diagram of the startup components began to look as shown in Figure 16.

At the backend level, scripts in the python language were added, with the help of which the necessary information is processed and visualized. Google analytics was connected to the marketing site, and the ability to send the necessary events to amplitude was added using the SDK. After setting the amplitude, it became possible to create dashboards, for example (fig 17.), the created list of events fully covers the necessary metrics defined in the defining key metrics section. For each solution (POS system, delivery showcase, e-menu), separate dashboards were created on which the use of the functionality was monitored. In addition, a financial dashboard was created, with information about payments, conversions of interest and the calculation of historical LTV. Also, thanks to the functions built into Amplitude, it became possible to calculate the retention rate and churn rate of users. Ampliude is used as a product analytics system, with its help, startup management can determine exactly how the product works, what value users find in it, and draw conclusions based on this.

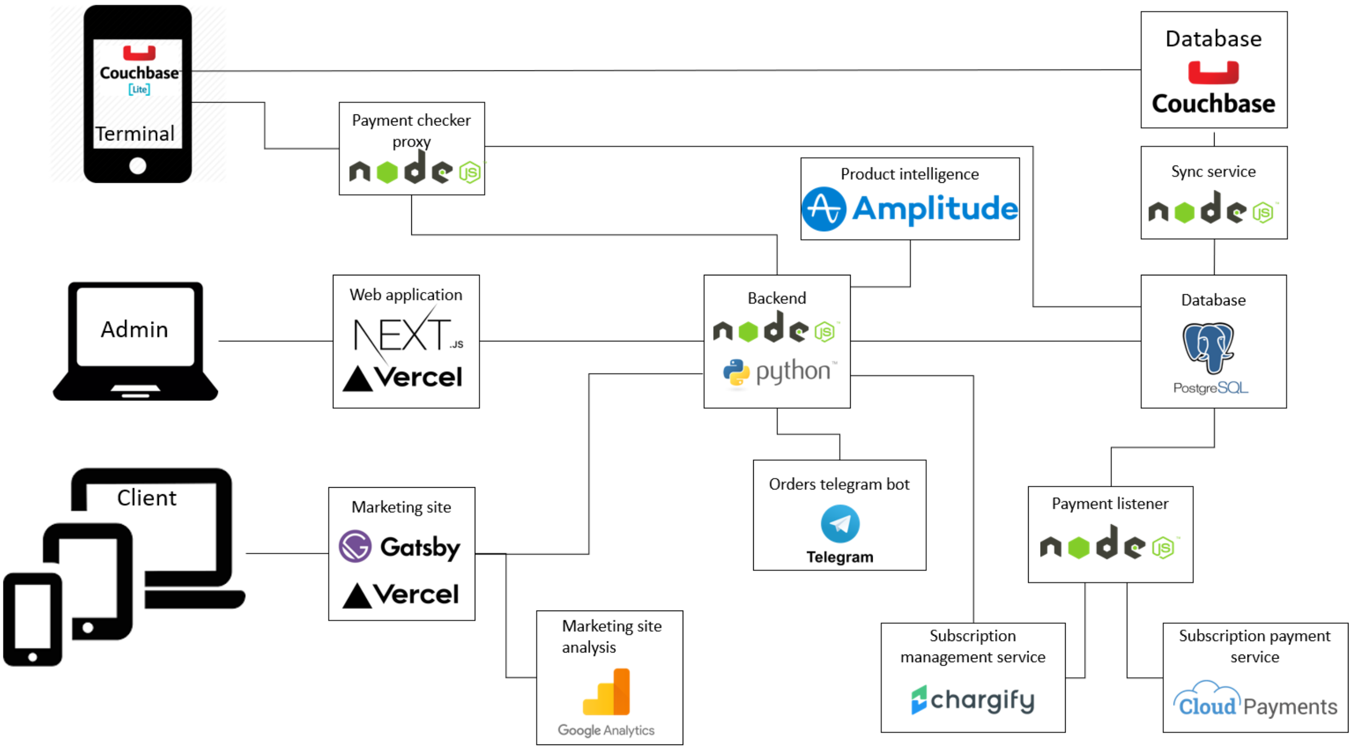


Figure 16. Final component schema

For marketing analytics, google analytics was used. In addition to displaying information about the use of the marketing site using google analytics and google data studio, we managed to create end-to-end analytics.

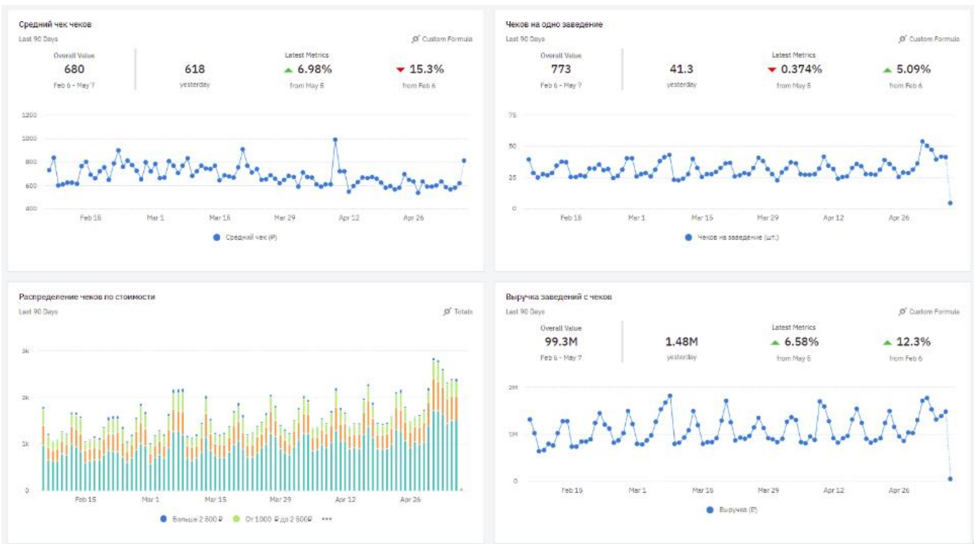
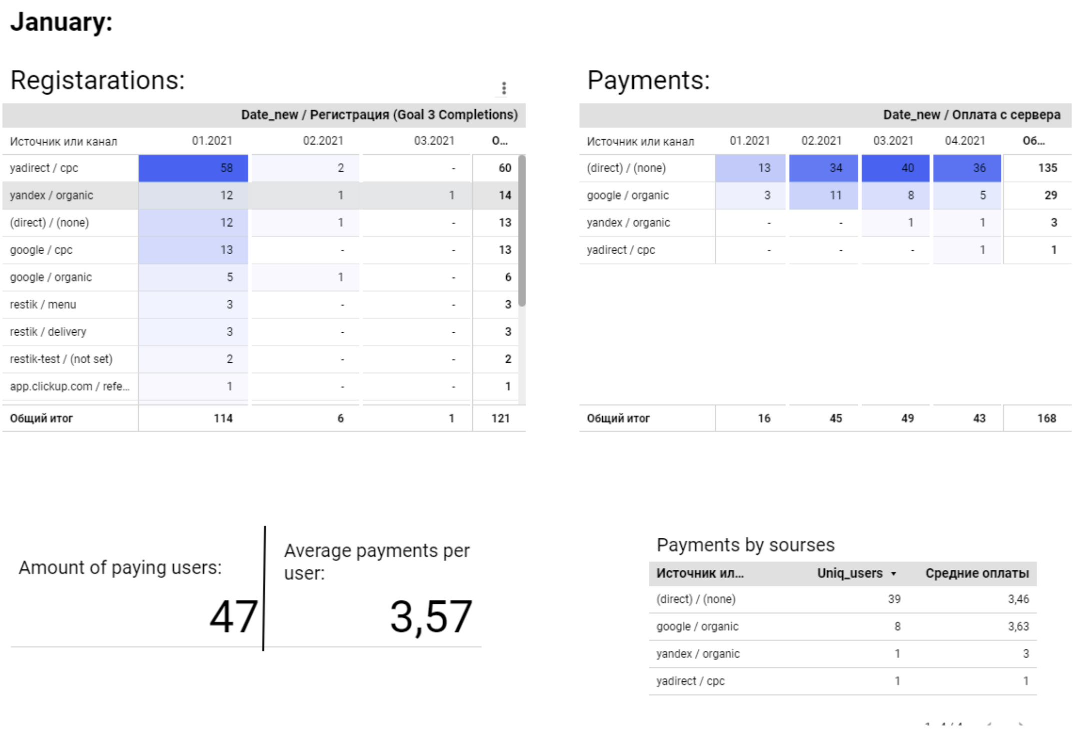


Figure 17 Amplitude dashboard example

In google analytics, all users were divided into monthly cohorts. For each cohort, the display of the traffic source, the number of registrations and the number of payments was configured (fig 18).



January:

Registarations:

	Date_new / Регистрация (Goal 3 Completions)			
Источник или канал	01.2021	02.2021	03.2021	О...
yadirect / cpc	58	2	-	60
yandex / organic	12	1	1	14
(direct) / (none)	12	1	-	13
google / cpc	13	-	-	13
google / organic	5	1	-	6
restik / menu	3	-	-	3
restik / delivery	3	-	-	3
restik-test / (not set)	2	-	-	2
app.clickup.com / refe...	1	-	-	1
Общий итог	114	6	1	121

Payments:

	Date_new / Оплата с сервера				
Источник или канал	01.2021	02.2021	03.2021	04.2021	Об...
(direct) / (none)	13	34	40	36	135
google / organic	3	11	8	5	29
yandex / organic	-	-	1	1	3
yadirect / cpc	-	-	-	1	1
Общий итог	16	45	49	43	168

Amount of paying users: 47

Average payments per user: 3,57

Payments by sourses

Источник ил...	Uniq_users ▾	Средние оплаты
(direct) / (none)	39	3,46
google / organic	8	3,63
yandex / organic	1	3
yadirect / cpc	1	1

Figure 18 Google analytics cohort example

Thus, an environment for data-driven management was created within the startup, examples of decisions made using this environment will be discussed in the next chapter.

5. EXAMPLES OF DATA-DRIVEN APPROACH USAGE

This chapter will explore the practical application of a data-driven management approach. After setting up the infrastructure for data collection, Restick's management began to apply this structure. This chapter will look at 4 cases of how a data-driven approach helped make management decisions.

5.1. Case 1: Salesperson mistake that lead to new product positioning

At the time of this case, Restick did not yet have 3 separate solutions, there was only a POS system with the ability to create an electronic menu and a delivery site.

This case began with a cohort analysis of conversions from registration to pay. From the experience of past cohorts, we knew that on average, in the first month after registration, 20 % of registered users pay for a subscription, and then within 3 months this number decreases to about 12 %. However, an oddity was noted in the behavior of the November cohort (Fig 19.):

Cohort	09.2020	10.2020	11.2020	12.2020	01.2021
September	18,6%	16,5%	14,6%	12,5%	12,5%
October		21,3%	18,3%	15,6%	14,0%
November			20,4%	10,6%	5,5%
December				19,7%	15,7%
January					23,1%

Figure 19 Cohorts conversion up to January

In the first month (November), the behavior of the cohort did not differ from the previous cohorts, but then within 2 months the conversion dropped to 5.5 %. At the same time, at this moment there were no significant changes in design and functionality, and the behavior of the December cohort was similar to the behavior of the other cohorts. It was decided to conduct a survey among customers who paid for a subscription in November, but stopped paying for it in the following months, in addition to reasons such as “did not like the product” and “the institution closed”, about 35 % of respondents said that the product “did not implemented the functionality they need ”. Based on the results of the survey, the following story emerged:

In October, during the discussion of the further development of the startup, a proposal was made to introduce new functionality. Further, user research was carried out, which showed that the functionality was really in demand, but there were no announcements that it would be implemented. However, a one of the salespersons who was attracting new customers in November told customers during presentations that this functionality would soon be rolled out. The introduction of the new function turned out to be a rather complicated process (this is why there was initially no announcement that it would be added since no one could say with certainty how long the implementation would take).

As a result, the function was added in mid-January (for this reason, the December cohort remained at the level of average), but part of the customers who came in November left the application without waiting for the required function. As a result, explanatory work was carried out with the salesperson, and all customers who came in November were sent notifications that the function was implemented, and a gift trial period for 2 weeks. As a result, we got following conversions (fig.20):

Cohort	09.2020	10.2020	11.2020	12.2020	01.2021	02.2021	03.2021	04.2021
September	18,6%	16,5%	14,6%	12,4%	12,4%	12,4%	11,0%	11,0%
October		21,3%	18,3%	15,6%	14,0%	11,2%	11,2%	11,2%
November			20,4%	10,6%	5,5%	9,4%	11,4%	11,4%
December				19,7%	15,7%	12,7%	12,7%	12,7%
January					23,5%	15,4%	14,0%	14,0%
February						24,1%	21,0%	19,5%
March							24,6%	21,3%
April								24,9%

Figure 20 Cohorts conversion up to April

Also, as a result of this event, it was decided to periodically attend sales meetings with clients. At that time, the startup had 2 salespersons. When analyzing their negotiations with clients, an interesting feature was revealed, the salesperson A when advertising a product, mainly focused on functions common to all automation systems that would be easier to perform with the help of a Restik (keeping checks, warehouse accounting, analytics), and salesperson B mainly advertised specific functions that are poorly represented in competitors creation of an electronic menu and a delivery site).

Based on this, the idea came up to conduct an experiment in order to understand which type of positioning is better. For the experiment, it was decided to create A/ B test. In this test, a control group of users is compared to a test group (or set of test groups) in which one or more conditions are changed in order to find out how these changes affect the target. In this case, the test and control groups differed in the presentation that the salespeople showed them. The experiment design is provided in Table 11:

Table 11 - Experement 1 design.

Hypothesis	Restic's positioning from specific functionality point of view in direct sales attracts customers better
What will be changed in product	Test group: mainly specific functions will be described Control group: mainly a general overview of the product will be described
Test group	New users, that attracted by salesperson
Key metrics	Conversion from registration to payment in first and second month
Effect	Users from test group should have higher conversion than control group
Action plan	If test group conversion is higher start positioning from specific functions point of view. If conversion is the same, keep everything as it is (that means that positioning is doesn’t matter for users) If control group conversion is higher start positioning from general product point of view.

Next, it was necessary to make the result comparable. To reduce the influence of other factors, 2 presentation scripts were created. One focused on specific features, while the other focused on the overall ease of use of the product. Further, each salesperson, when communicating with a client, randomly selected one of the scripts and made a presentation on it. In total, according to the results of the month, salespersons held 44 meetings, while the group for which the presentation was conducted from the point of view of specific functionality included 24 clients (let's call it group A), and the group for which the presentation was conducted from the general point of view of the product included 20 clients (call it group B). From group A in the product 9 users registered and paid in the first month, and from group B 3 users.

Accordingly, the conversion of group A was 9/24 = 0.375 or 37.5 %, and the conversion of group B was 3/20 = 0.15 or 15 %. Now it is necessary to compare these 2 conversions to understand whether we can, based on the collected data, assert that the conversion of group A is higher than the conversion of group B, in other words, whether this difference is statistically significant. In this context, "statistically significant" means whether the difference between the conversions of the 2 groups can be explained by chance, if not, then we believe that the difference is related to the changes we made in the presentation. It should be noted that from the point of view of strict statistics, further calculations are not entirely correct (for example, for a sample of less than 30, it is necessary to use not the normal distribution, but the t-distribution [26]) and can be used only in the case of comparing the conversions of 2 groups, however, for this task we can use the assumptions.

For further calculations, it is necessary to explain the difference between *real conversion* and *conversion rate [48]*. The real conversion shows what percentage of the total population of all establishments after registration will pay for the first month of subscription. In this case, in order to find out the real conversion, it is necessary that all possible establishments on the Earth (the general population) are registered in the product. Unfortunately, this is not possible, so we cannot know the actual conversion, but we can estimate this conversion based on a specific sample. After assessing the conversion, we need to understand how well that rating describes the actual conversion. To do this, we can use the confidence interval. In the context of a conversion, a confidence interval with a confidence level P is an interval that, with a probability P, contains the actual value of the conversion. To calculate the confidence interval, we will use formula (9):

$$\bar{x} \pm Z_{\frac{a}{2}} * \frac{\sigma}{\sqrt{n}} \tag{9}$$

Where:

$\bar{x}$ - mean of data values

σ- standard deviation

n - sample size

Z – trust coefficient

So, find all these values. In order to find Z coefficient we need set trust level (in this case 90% or 0,9) divide it by 2 and look through Z-score table [28]. In our case this coefficient will be equal to 1,65.

Nest it is needed to calculate mean and standard deviation. We know the size of each of the samples (24 and 20). The standard deviation is calculated by the formula (10) [28]:

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \qquad (10)$$

Where

σ- standard deviation

x - data values

$\bar{x}$ - mean of data values

n - sample size

The average of the data values is calculated using the formula [28]:

$$\overline{x} = \frac{\sum x}{n} \qquad (11)$$

Where:

$\bar{x}$ - mean of data values

x - data values

n - sample size

To calculate the average, it is needed to understand what is the average in the sample. For this case, take 0 if the user did not pay for the subscription after registration, and 1 if the user paid for the subscription after registration. So, the average is essentially equal to the conversion (9/24 = 0,375 in the case of group A and 3/20=0,15 in the case of group B). calculate standard deviation. Put everything in the formula (10) we will get following result:

$$\sigma_a = \sqrt{\frac{\sum_{k=1}^{k=24}(X_k - 0{,}375)\text{^}2}{24}} \qquad \sigma_b = \sqrt{\frac{\sum_{n=1}^{n=20}(X_n - 0{,}15)\text{^}2}{20}}$$

Depending if customer pay for subscription or not X_k and X_n equal to 1 or 0. Standard deviation of each of the groups is 0.047842 in the case of group A and 0.02851 in the case of group B. Now we know all necessary values, so put them in the formula 9, for group A, standart deviation will be calculated like:

$$0,375 \pm 1,65 * \frac{0.047842}{\sqrt{24}}$$

And for group B:

$$0,15 \pm 1,65 * \frac{0.02851}{\sqrt{20}}$$

Obtained results is shown in table 12:

Table 12 - Confidence interval for 2 groups of users.

Group	Confidence interval
Group A	[0,359;0,391]
Group B	[0,139;0,16]

These values are necessary in order to understand how the conversion rate relates to the actual conversion. Consider Group A, for example, management needs to find out if the conversion in Group A was above 35 %. Since 35 % (0.35) is less than the lower limit of the confidence interval of group A, we can argue that with a 90 % probability, the actual conversion of this group is higher than 35 %. But, if it is necessary to find out whether the conversion rate in this group was higher than 37 %, it will not be possible to give an unambiguous answer based on the available sample, since 37 % (0.37) lies in the confidence interval. Thus, we have solved the problem of comparing the conversion of one group with a constant, which is solved in the following way:

1. Calculate conversion rate based on a sample of users
2. Calculate the confidence interval that describes the actual conversion
3. If the interval contains a constant, then based on the collected data, the actual conversion and the constant are indistinguishable. If the interval does not contain a constant, then the real conversion significantly differs from the constant.

The initial task was to compare the conversions of the 2 groups and see if the difference between them is statistically significant. In order to solve this problem, reduce it to the problem of comparison with a constant. Comparing 2 conversions with each other is tantamount to comparing the difference of these conversions with 0. It should be noted that in this case the standard deviations will be added to each other, and the conversions will be subtracted. Using this logic, we will get following formula (12):

$$C_1 - C_2 \pm Z * \left(\sqrt{\frac{\sum(x_1 - \overline{x_1})^2}{n_1} + \frac{\sum(x_2 - \overline{x_2})^2}{n_2}} \right) \tag{12}$$

If the obtained confidence interval contains 0, then we can say that the difference in these conversions is statistically insignificant, that is, it can be explained by chance, if the interval does not contain 0, then we can say that the difference is statistically significant. In this case, at the 90 % confidence level, the confidence interval of the difference in conversions is:

$$0{,}375 \ - 0{,}15 \pm 1{,}65 * \left(\sqrt{\frac{\sum_{k=1}^{k=24}(X_k - 0{,}375)\hat{}2}{24} + \frac{\sum_{n=1}^{n=20}(X_n - 0{,}15)\hat{}2}{20}} \right)$$

Which is equal to [-0.44; -0.096]. This interval does not contain 0, that is, the difference between 2 conversions is statistically significant. However, it should be noted that at 95 % confidence level this interval is equal to [0.472; -0.024] and contains 0, but since 90 % confidence level was initially chosen, we can argue that the difference in conversion between the 2 groups is not due to chance, but a change in the approach to product presentation.

Later in this chapter, many problems of comparing conversions of 2 user groups will be solved. To simplify the presentation, for each comparison, only the final result of calculating will be displayed, the confidence interval calculated according to the algorithm indicated above.

So, after all the calculations, we found out that in direct sales, focusing on specific functionality increases the conversion from registration to payment. Thus, the original hypothesis turned into a concrete fact. At the same time, at the end of the experiment, the startup's management was just drawing up a new marketing strategy in order to increase the CTR of advertising banners and the conversion from click to pay. A proposal was made to use this fact when drawing up a new marketing strategy. Thus, a new hypothesis appeared - "Using advertising banners with specific functionality will attract users better than conventional banners."

Based on this hypothesis, an experiment displayed in table 13 was made. For the purity of the experiment, 7000 ad impressions were purchased from one supplier. When loading an advertisement, users saw one of 2 variants of an advertising banner (a banner advertising specific functions - functions, a banner advertising the system in general - general). For each type of banner, the CTR and conversion from a click on the banner to registration were calculated, the result is shown in Figure 21, for clarity, the step with displaying the banner has been removed from the graph, since the CTR is in the region of 2-3 percent, and due to economies of scale, the results poorly visible.

Table 13 - Experement 2 design.

Hypothesis	Using “specific function” advertisement banners attract users better
What will be changed in product	Test group: Will see a banner advertising specific function Control group: Will see banner advertising general functions
Test group	Users who have been shown ads
Key metrics	CTR, conversion from click to registration
Effect	Users from test group should have higher key metrics than control group
Action plan	If test group key metrics is higher start buying only “specific function” advertisement banners. If conversion is the same, keep everything as it is (that means that advertisement banner is doesn’t matter for users). If control group conversion is higher start buying only “specific function” advertisement banners.

The results of the experiment showed that the conversion of both impressions per click and from click to registration for the “functions” version of banner is 2.57 % and 0.54 %, respectively, and for the “general” version of banner 2.32 % and 0.41 %. Since to calculate the statistical significance, it is necessary to compare conversions separately, the conversion from click to registration was calculated, for the “functions” version it was 21 %, and for the “general version” 17.6 %.

Despite the fact that the “functions” version of banners performed better than the “general” version, the confidence interval of these two groups for the impression-to-click conversion was [-0.0097; 0.00466], and for the click-to-click conversion registration [-0.0822; 0.15163]. Both of these intervals contain 0, that is, the results are statistically insignificant, and with a 90 % probability are due to chance, and not to real changes.

The results were disappointing for the management of the startup, the hypothesis was not confirmed, but this result still clarified the product model - in advertising banners, the emphasis on specific functionality or general functionality does not matter.

At the same time, there was still the fact that in direct selling, such positioning affects the conversion. A new hypothesis was put forward - "Focusing on specific functionality on a marketing site, will increase conversion from site visit to registration."

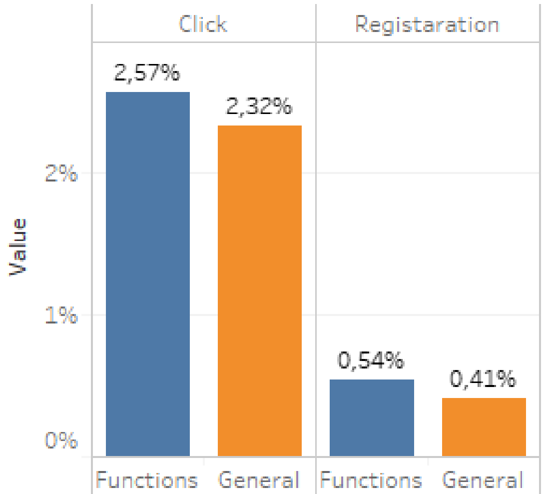


Figure 21 Result of banners A/B testing

However, the question arose of how to test this hypothesis. The marketing site already had information about the functions of creating an electronic menu and an online delivery site. For verification, a risky step was taken to conduct a worsening A / B test. In this A / B test, the conditions of the test group are deliberately deteriorated, in order to test how this deterioration will affect the targets, as a result, an experiment was made, shown in Table 14:

Table 14 - Experement 3 design.

Hypothesis	Focusing on "specific functionality" on the marketing site will increase conversion from site visit to registration.

What will be changed in product	Test group: Will see site version without description. Control group: Will see with description.
Test group	New web site visitors.
Key metrics	Conversion from visit to registration.
Effect	Users from test group should have lower key metrics than control group.
Action plan	If test group key metrics is higher conduct further investigation If conversion is the same, return previous version of website to every user If control group conversion is higher count hypothesis as correct.

It is important to understand that worsening A / B tests have limits of applicability, and the results of such tests must be approached with extreme caution, however, in some cases, such tests help to quickly test the hypothesis of interest.

For the experiment, 2 versions of the site were created, in the control group, under the description of the POS system, the specific functions of Restick were indicated point by point (the same as usual). The other was just a link to a POS system without specifying functions. When users visited the site, they were redirected to one of the pages in a 50:50 ratio. In order to avoid repeated visits to the site, users were tracked through cookies. The experimental sample included only users who did not have a cookie on the site. The test results are shown in Figure 22. The conversion of the control group was 0.52 %, and the conversion of the test group was 0.21 %. At the same time, the confidence interval was equal to [-0.0056; -0.00043] at a confidence level of 90 %. The interval does not contain 0, which means that the difference in conversions is statistically significant. As a result, after all the experiments, the following facts were checked and confirmed:

1. In direct selling, focusing on specific functionality increases registration-to-pay conversions.
2. Banner ad content does not affect the conversion from click to registration in any way.
3. The lack of a short description of specific functions lowers conversion.

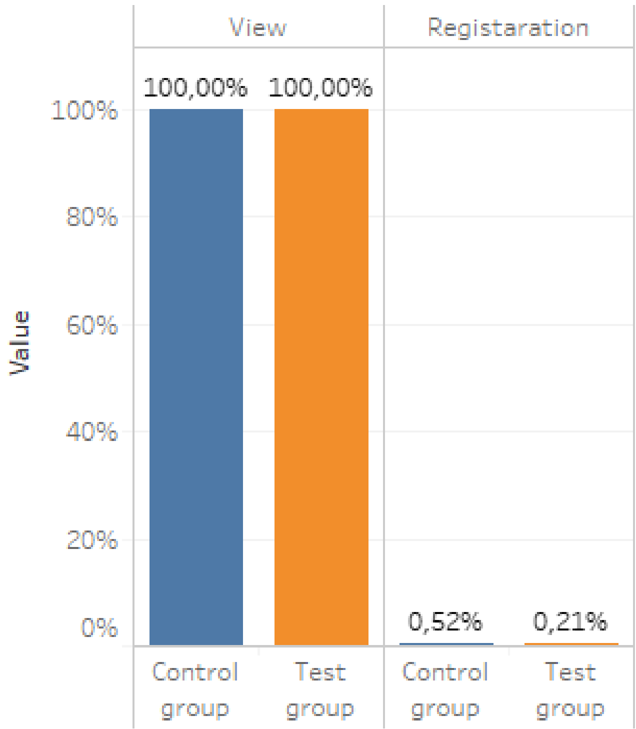


Figure 22 Results of degrading A/B test

By themselves, these facts could not give much, however, on their basis, it was decided to conduct a survey among the owners of establishments and ask them to tell why they chose / did not choose Restik. As a result of the interview, the following interesting fact emerged. Most of the respondents chose the system according to the following principle - they either saw the banner, or followed the link to the site, then scrolled the site in search of the necessary functions of the automation system. If they could not find a description of a given function, they continued to search other automation systems. This explains why the banner conversion checker did not produce any results. At the sight of a banner, the owners of establishments simply clicked on it, not really reading into the banner's content, the main search took place precisely on the site itself.

Based on the data received, the management decided to change the positioning of the product. If earlier Restick was positioned as a simple and fast automation system for the restaurant business, now the startup has begun to position itself as cloud solutions for various tasks of the restaurant business. Further, the entire product was divided into 3 separate solutions.

And the marketing site started showing each solution separately. Thus, due to the salesperson's mistake, Restik eventually came to a new positioning in the market.

5.2. Case 2: New monetization approach

After splitting the product into 3 separate solutions, all solutions were still available for the price of 1 subscription. That is, by paying for a subscription, the user received all 3 solutions and himself chose which one to use.

At this point, the initial investment began to come to an end, so management began to look for new ways to monetize. One of the proposals was to take payment for each decision separately. If earlier users using all 3 solutions paid 1000 rubles a month, now they would have to pay 3000 rubles. At the same time, it was experimentally impossible to check how this would affect the product, since this would require assigning different prices for different users. Technically, such an experiment is possible, for example, for some new users to give a 100 % discount on a solution, and not to give another one, but this would be dishonest to users, and the results of such an experiment would be highly questionable. In any case, the main reason for the impossibility of such an experiment was the decision of the startup management, which categorically refused to experiment with subscription prices.

In this case, the effect can still be estimated, for this it is necessary to build a financial model based on the data that is already available in the startup and understand how such a decision can affect users.

To begin with, it was necessary to understand how many users use more than one solution, for this a diagram was built in Figure 23. After the analysis, it turned out that 60 % of users use only one of 3 solutions, that is, the planned change will not affect them in any way (provided that that they would continue to use only one solution), another 36 % of users used 2 solutions and only 4 % of users used all 3 solutions.

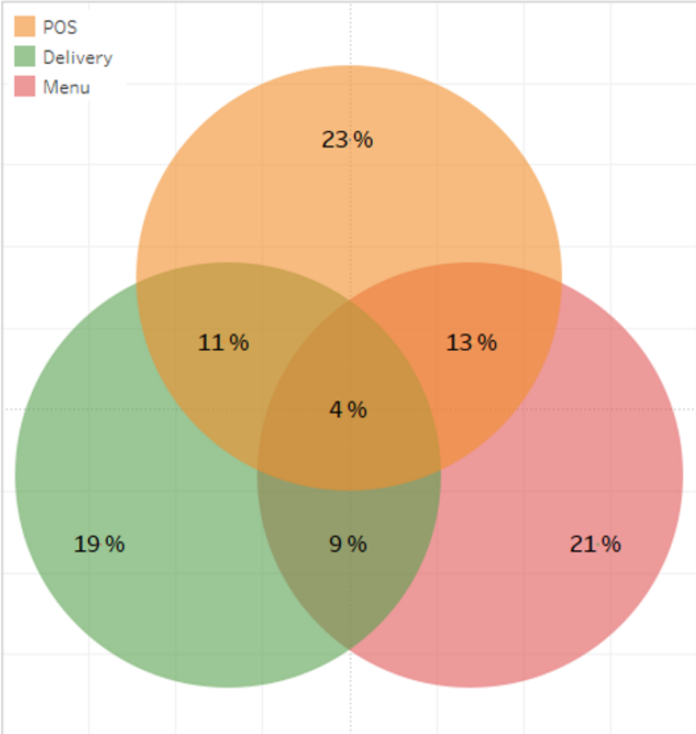


Figure 23 User groups intersections before new monetization approach

Based on the obtained distribution of the use of subscriptions, we will construct a financial model [50]. To do this, assume that the app uses 100 users. In this case, the total monthly revenue will be 100,000 rubles (100 users multiplied by the cost of a subscription of 1,000 rubles). Next, simulate a situation in which each subscription is paid separately, based on the data from Figure 23, we get that for a month the revenue will be: 60 users with one subscription * 1000 rubles + 36 users with 2 subscriptions * 2000 rubles + 4 users with 3 subscriptions * 300 rubles = 144,000 rubles.

The expected effect of withdrawing money for each subscription is that some users will stop using the product. Let's calculate how many users must leave the product so that the revenue in the case of a separate payment for each subscription is lower than the revenue from payments for all subscriptions.

For users who only use one solution, essentially nothing will change, so let's say they stay anyway. Let's change the number of subscribers using solutions 2 and 3, so that the total amount becomes less than 100,000. For convenience, let's call users who use 1 solution - group 1, 2 solutions - group 2, 3 solutions - group 3, as a result we get a table 15:

Table 15 - Subscription payment financial model

Group 1, unit	Group 1 payment sum, rub	Group 2, unit	Group 2 payment sum, rub	Group 3, unit	Group 3 payment sum, rub	Total income
60	1000	13	2000	4	3000	98000
60	1000	15	2000	3	3000	99000
60	1000	16	2000	2	3000	98000
60	1000	18	2000	1	3000	99000
60	1000	19	2000	0	3000	98000

Now, in table 16, expressed in relative units what percentage of users from group 2 and group 3 must leave the product so that the revenue becomes less than the revenue from a single payment for all subscriptions.

Table 16 - Modeling the percentage of departed users

	Group 2	Group 3
Case 1	64 %	0 %
Case 2	58 %	25 %
Case 3	56 %	50 %
Case 4	50 %	75 %
Case 5	47 %	100 %

Thus, based on the current situation in the distribution of users, we can say that the decision to introduce a separate payment for each subscription will be unsuccessful if after that more than 64 % of users who pay for 2 subscriptions leave the product, or more than 58 % of users who pay for 2 subscriptions and more than 25 % of users who pay for 3 subscriptions, etc.

Next, we will consider the revenue of users of each of the groups, since thanks to the analytics system, we know the average number of checks and the average price of a check for each of the groups, the result is shown in Table 17:

Table 17 - Group income comparison

Group	Average amount of receipts daily, unit	Aberage receipt sum, rub	Total income daily, rub
Group 1	95,8	468,6	44892
Group 2	149,1	492,3	73402
Group 3	223,4	513,4	114694

This table includes receipts from both the delivery site and the automation system (representatives of groups 2 and 3 definitely have some of this). Now calculate what percentage of the daily revenue is the subscription payment for each of the groups in the case when the price is charged for all solutions (case 1) and for each solution separately (case 2). The calculation results are presented in table 18:

Table 18 - Subscription cost as a percentage of daily income

	Case 1	Case 2
Group 1	2,23 %	2,23 %
Group 2	1,36 %	2,72 %
Group 3	0,87 %	2,62 %

It should be borne in mind that this percentage is calculated from the daily revenue, and the subscription is paid for a month. That is, in the case of 1, users from group 1 need to pay 2.23 % of the daily revenue to use Restik for one month, while users from group 3 in the first case need to pay 0.87 % of the daily revenue once.

Based on the constructed models, the following conclusions can be drawn:

1. Unsubscribing for each technology solution will not pay off if more than 50 % of users using 2 or 3 solutions leave after that.
2. When paying for a subscription for each product separately, the one-time payment from the daily revenue for all users is leveled.

Based on these facts, the following decisions were made:

1. Warn all users in advance that subscription fees will change, in order to find out the reaction, and to understand the potential churn if there were many emails from customers with the desire to leave, to say that after the meeting, the management decided not to introduce subscription fees.

2. Provide all customers with calculations of how much, on average, a one-time daily revenue will be spent on paying for a subscription.

In this case, it was experimentally impossible to test how the product would change after the introduction of the new payment policy, but anyway, using the available data, a financial model was created that helped the management understand what results would mean that the decision was unsuccessful. The data-driven approach may not always give a clear answer to what positively affects a startup and what is negative, but in any case, using this approach, management can learn new information that can help in decision-making. These changes were introduced in January 2021. The monthly churn of users is shown in the graph in Figure 24:

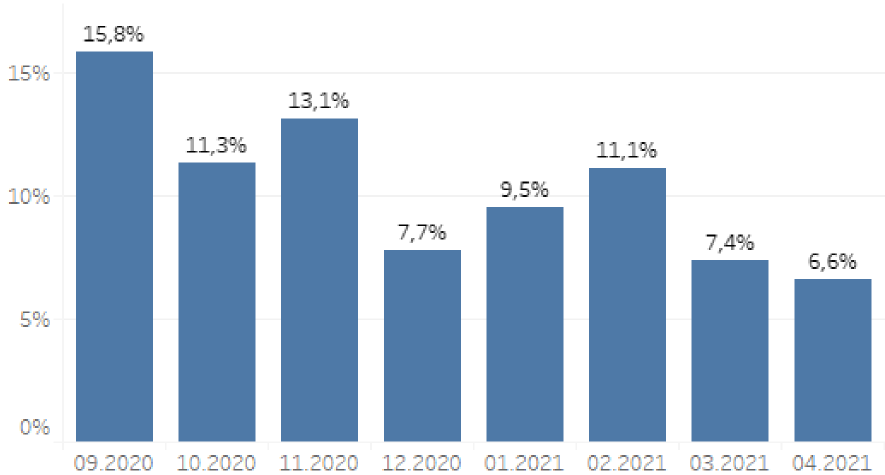


Figure 24 Dormant users per month

This graph should be read as follows, for example 7.7 % in January means that 7.7 % of users who paid in the previous month (December) canceled their subscription in January. As expected, churn rates in February were higher than in the previous 2 months, as in January there was still a rule of one payment for all solutions, and in February, establishment had to pay for each solution separately, but then the churn rate dropped to a historic low. At the same time, total revenue in March increased by 35 % compared to February. The worst predictions regarding this solution did not come true, and a significant part of the users remained in the product. In this case, the data-driven approach was not suggested specific actions but was able to provide additional information about the product, on the basis of which the decision was made to introduce a new method of monetization.

5.3. Case 3: New marketing strategy

The next case considers choosing the optimal marketing channels in terms of return on funds and creating a new marketing strategy.

After defining a new marketing strategy, the marketing costs in the startup's budget were significantly increased, and the management wanted to evaluate the marketing channel for acquisition. There was a need to understand how the spent funds attract new users. For this, Google analytics, already integrated into the product, was used. An example of calculating the effectiveness of marketing channels will be shown on a cohort that came in March. So, in March, the metrics indicated in table 19 were obtained:

Table 19 - March cohort marketing channels

Sourse	Registarations, units	Payment, units	Uniq payment, units
(direct)/organic	17	6	4
google/organic	21	4	3
yandex/organic	16	2	2
yadirect/cpc	14	5	4
mail.ru/refferal	9	1	1
google/cpc	11	3	3
inst_target/delivery_video	19	5	4
a2is.ru/refferal	5	2	1
inst_target/emenu_video	14	4	4
i.instagram.com/refferal	10	2	1
away.vk.com/refferal	9	1	1

The Uniq payment column shows how many unique payments were made from a particular source, in other words, the uniq payments column shows the number of users who came from the marketing channel, and the payments column shows how many payments were made by users who came from this channel (these numbers may not coincide, as to how one and the same user can buy several solutions, and pay for a subscription for each of them).

At the same time, if the sources say organic, this means organic traffic - users who went to the site from regular search results, and not from paid advertisements [49]. Therefore, in further calculations, organic sources will not be taken into account. Next, we need to calculate how much it costs to attract each user to each of the channels, as well as how to quickly return the funds spent on the channel.

First, combine some of the channels, since, in fact, they are a single channel. For example, advertising via Instagram d in this table has 3 manifestations (inst_target / delivery_video, inst_target / emenu_video, i.instagram.com/refferal), while the payment is made for all 3 sources at once. When purchasing an ad, a few creatives are simply loaded and the desired impression share of those creatives. Since in this task it is necessary to analyze marketing channels, and not specific creatives, this operation can be done.

The task is to determine, based on the available data, an investment in which marketing channel is most effective in terms of the rate of return of marketing costs and draw up a predictive model that can be used to calculate the marketing budget for each of the channels. To do this, for each of the channels, we calculate the following metrics [35]:

1. CAC - client acquisition cost, the cost of attracting one registered client, is calculated by dividing the marketing channel budget by the number of registered users.
2. Conversion from registration to payment - this conversion can be obtained by dividing the number of registered users by the number of unique users who paid for the subscription (in this case, the number of unique users is equal to the number of unique payments).
3. Average payments per user - answers the question how many times, on average, per month users pay for a subscription, calculated by dividing payments by unique payments.
4. Average revenue per paying user (ARPPU) - shows how much revenue one paying user will bring on average (that is, a user who made a conversion from registration to payment)
5. Payback period – in how many months the funds invested in the channel will be returned.

Dwell on the Payback period per user indicator in more detail. It is necessary to understand how, knowing the conversion, the average number of payments, ARPPU, and the cost of user acquisition, determine the return time to the marketing budget. Make calculations by dividing the marketing budget by divided by the cost of acquisition, we will get the number of registrations from the marketing channel. By multiplying registrations by conversion, we get the number of paying users. By multiplying the paying users by ARPPU, we get one month's channel revenue. In order to find out the payback period of the channel, we need to divide the marketing budget by revenue for one month. However, in calculating revenue for one month, the marketing budget was already used, so bringing everything together in a single formula, marketing budget will be reduced and as a result following formula (13) will be obtained:

$$PP = \frac{ARPPU * Conv}{CAC} \quad (13)$$

Where

PP – Payback period

ARPPU - Average revenue per paying user

Conv – Conversion from registration to payment

CAC – Client acquisition cost

Further, all metrics were calculated for each channel, the result is presented in Table 20. The following conclusions were obtained from this table:

1. The yadirect source shows the shortest payback period and the largest conversion from registration to payment.
2. The lowest cost of attracting from the source is mail.ru, but the payback period of this source is the longest.
3. Sources google and inst target are relatively similar to each other, google have a higher conversion, while inst target has a large profit per paying user.
4. a2is.ru has the highest acquisition cost and the second fastest return on investment, but it should be borne in mind that the figures for this source are obtained on the basis of only 1 paying user.

The data obtained is not entirely correct from the point of view of strict statistics, however, based on it, we can make an approximate forecast of which marketing channels are most effective in terms of return on funds.

Table 20 - Model of payback period

	yadirect	mail.ru	google	inst_target	a2is.ru
CAC, rub	714,3	555,6	909,1	930,2	1000,0
Conversion from registration to payment, percent	29 %	11 %	27 %	21 %	20 %
Average payments per user	1,25	1	1	1,22	2
ARPPU, rub	1250	1000	1000	1222,2	2000
Payback period , months	2,0	5,0	3,3	3,6	2,5

The results came as a surprise to the management of the startup, since before that the marketing budget was distributed as follows:

1. Inst_target – 40 %
2. Google-35 %
3. Yadirect – 10 %
4. Mail.ru - 10 %
5. A2is.ru – 5 %

That is, 10 % of the budget was invested in the most efficient channel in terms of return on funds, and 75 % of the budget was invested in channels with similar characteristics. Based on the model obtained, the management revised the distribution of the budget for marketing channels, and it was also decided to evaluate the channels on a monthly basis and invest the budget in the most effective ones.

However, these are not all the insights that were obtained from the analysis of the effectiveness of marketing channels. It was also unexpected that more than a third of users came from organic channels. Based on this fact, it was decided to analyze what percentage of registrations and payments fell on organic channels, and what percentage were paid monthly. Since google analytics was only connected for 3 months at that time, the data was also available for only 3 months. For each month, the conversion from registration to payment was calculated, taking into account the division of marketing traffic into organic and paid, the result is shown in Figure 25:

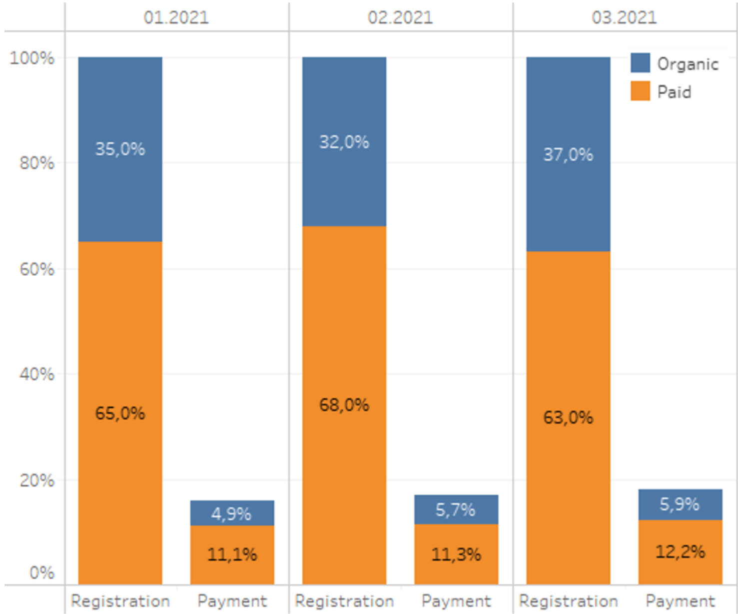


Figure 25 Conversion with marketing channel segmentation

That is, on average, more than one third of the total number of users came not from paid traffic, but from organic traffic. This fact puzzled the management, since it was assumed that no more than 10 % of organic traffic comes to the product, and all budgets and the economy were calculated based on this hypothesis.

Based on the results of this analytics, an interview with users was conducted, where the question was asked about how exactly users came to the decision to choose our product. It turned out that such a product as an automation system was chosen not just by an advertising banner. Before choosing a system, all the owners of establishments, one way or another, conducted a study and compared various systems with each other. To do this, they searched websites and thematic forums for information about various automation systems. At the same time, Restik rarely got into various review articles, however, due to the search history, the owner fell into the target group for displaying an advertising banner Restik.

Here it is worth clarifying the peculiarity of the work of google analytics, when a user clicks on a banner his cookie was remembered and on the basis of this a google id was assigned, and then all site openings, registrations and payments from this google id were counted by the system as coming from a paid channel. However, in reality, between the transition from the paid channel and registration, there were many more site openings, and searches for other automation systems.

The usual user journey from the first contact with Restik to payment was as follows:

1. The user was looking for information about various payment automation systems.
2. Because of this, the user was shown targeted ads from Restik.
3. The user clicked on the ad, and from that point on, google analytics assumed the user came from a paid channel.
4. The user remembered that there is such a system as Restik, but after clicking on the advertisement, he continued to search for other systems.
5. After their own analysis, some users chose Restik and registered in it, while if the first contact with a user occurred through an advertising banner, such a user was counted as coming from a paid channel.

At the same time, the management believed that users choose Restik, just like any other product, that is, they see ads, go to the site, get acquainted with the product and download it. Based on the data obtained, a new marketing strategy was proposed, one of the points of which was to increase the number of mentions of Restik in various review articles.

Thus, the analysis of marketing channels not only helped to select the most effective channels in terms of return on funds, but also to better understand the end user and change the marketing strategy in accordance with this.

5.4. Case 4: Consequences of e-menu implementation.

The last case will describe how incorrect data analysis and unverified hypothesis ultimately led to the creation of a solution that was spent on a lot of effort and money, but the result was not at all what was expected.

It will be about introducing a function to create an electronic menu. At that time, Restik did not yet have a data-driven management environment. And all the analytics was carried out using direct queries to the database. Customer behavior was tracked by 2 queries, one counted the total number of receipts from all customers, and the other the number of registrations. And on the basis of this data, the results of the introduction of new features were assessed.

In April 2020, one of the investors proposed to make the function of creating an electronic menu in the product (at that time, Restik had only a POS system and the ability to create a delivery site). According to the investor, since the lockdown in Russia has ended, but the trend towards contactlessness has remained, such an offer will definitely be in demand on the market. From the point of view of a data-driven management approach, this is a hypothesis that needs to be tested, but at that moment everyone simply believed the investor and part of the team of programmers was transferred to the creation of an electronic menu.

The functionality was introduced at the end of May 2020. And the team was very pleased with the results as the number of registrations and the number of checks in establishments began to grow rapidly, on average every month the number of checks increased by 20 percentage points, and the number of registrations by 15 percentage points. The growth of checks was associated with the introduction of the electronic menu function (this was also an unconfirmed hypothesis).

More and more efforts of programmers were directed to finalizing and improving the electronic menu, this task had the highest priority in the backlog, but in October-November the number of checks fell sharply, as shown in Figure 26. This drop was associated with the second wave of COVID-19, while itself the e-menu creation function was still considered one of the main drivers for increasing registrations in the product, so its development was still extremely active. It was assumed that after the end of the lockdown, establishments will continue to register in the same volumes as in the summer of 2020.

After connecting analytics systems and creating an environment for data-driven management, this case was also analyzed. Since by this moment the e-menu was already a separate solution, it became possible to find out what share of the revenue each solution brings. The result was very unexpected for the management - 10 % of the total revenue for the month came from the e-menu, and 45 % of the revenue came from the delivery site and POS. This fact did not in any way correlate with the management's confidence that e-menu is the main driver of registrations.

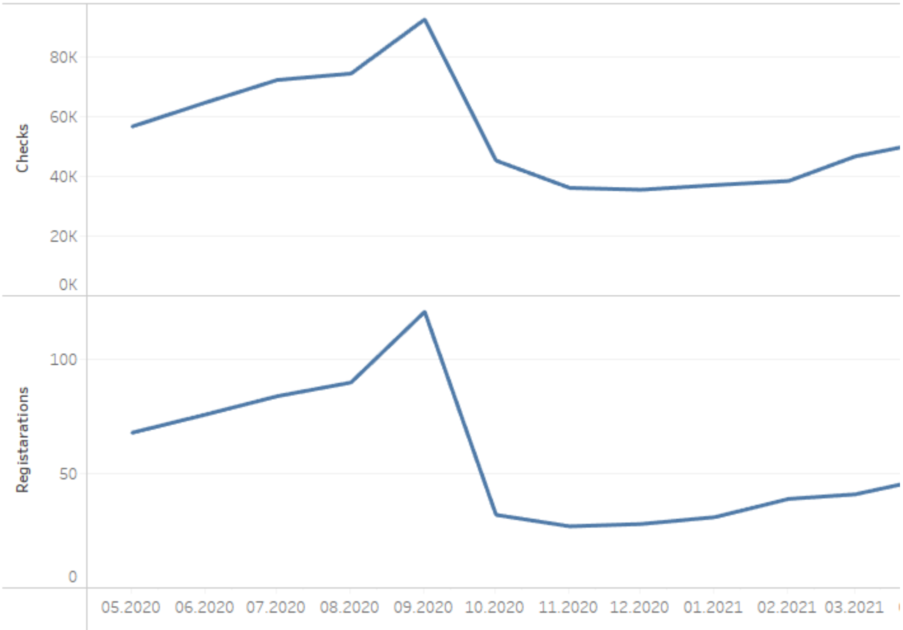


Figure 26 Amount of receipts and registrations per month

A detailed study was carried out on what caused the growth in the summer of 2020, as a result, the following facts became clear:

1. In the summer of 2020, the holiday season began in Crimea, in this regard, many seasonal establishments were opened that registered with Restik (in the summer months, establishments from Crimea accounted for about 60 % of all registrations and payments).
2. Seasonal establishments closed at the end of September 2020, and all customers who came from Crimea left the product.

At the same time, the beginning of the season in Crimea coincided with the release of the e-menu, and the end of the season with the beginning of the second wave of COVID-19. Thus, for management, changes in the product were driven only by internal factors, when in fact external factors influenced it. As a result, 7 months of programmers' work were spent on the function that brings 10 % of the revenue, while everyone thought that this was the right approach, since the hypotheses of the startup management coincided with the data. But this is the danger of hypotheses, even if the hypothesis is consistent with the data, it still needs to be tested, since the real reason may be completely different.

If the data-driven approach had been implemented earlier, the team could have run an experiment back in June to see if growth is really driven by the e-menu, and see that growth is actually driven by another factor. This would save many hours of programming and marketing team work.

However, even this unsuccessful case gave an important lesson to the entire startup team - even if everything goes as planned, it is necessary to confirm all the hypotheses from which the product is built, which is one of the principles of data-driven decision making - to trust facts, not hypotheses.

6. CONCLUSION AND DISCUSSION

The objective of the study was to understand what a data-driven approach is, how this approach can be applied in a B2b startup developing an application, and what is the effect of applying the approach. The research results were achieved by performing the following tasks:

1. Analysis of the current state of the European start-up market and the startup life cycle.
2. Analysis of principles, processes and methods of data-driven decision-making approach.
3. Compiling a list of possible metrics and technologies for a B2B startup.
4. Creation of a data-driven environment in the analyzed startup.
5. Application of the approach based on the example of specific cases.

The main research methods were:

1. Basic general scientific methods.
2. Case analysis.
3. Literature review.
4. Financial modeling.

Research work also has the following limitations:

1. The thesis considers startups that creating apps on B2B market.
2. Data-driven decision-making approach is not a standardized framework, the description of the approach in this thesis is a consequence of the analysis of cases and literature and does not pretend to be an industry standard.
3. This work is not a guide to action, but only describes the approach and in practice shows how this approach can be applied.

At the beginning of this work, an analysis of the European start-up market was carried out, and an approximate life cycle of a startup was described. Next, various cases of applying the data-based decision-making approach were considered and, based on the analysis of these cases, 3 principles of data-driven-decision making were taken away:

1. Trust fact, not hypothesis.
2. Compare the Comparable.
3. You are interacting with the product model, not the product.

Based on these principles, a data-driven decision-making process and methods were developed and described. which can be used to implement this process.

These methods are experiment, data manipulation, statistical approach, and user research. To track the results obtained using the approach, a list of possible metrics for a B2B startup was compiled and classified. Metrics can refer to growth metrics or product metrics:

1. Growth metrics - show the current state of the business
2. Product metrics - show how well a product is performing its task.

Also, metrics can be divided into the following 4 categories, depending on the tasks for which they are used:

1. Behavioral
2. Financial
3. Feature-based
4. Custom

Also, technologies that allow using the Data-driven decision-making approach was considered.

Then a practical example of using the data-driven approach was provided. For this, a diagram of the components of the analyzed startup was described, then, based on the use cases, a list of necessary metrics was compiled, the main metrics for this startup are:

1. LTV
2. Revenue
3. Conversion rate

In addition to these metrics, were also chosen metrics that track user behavior within the startup and marketing effectiveness. Then technologies were selected - Google analytics for tracking marketing channels, Amplitude for tracking user behavior within the product, as well as various scripts written in Python, all the selected technologies were implemented in the startup.

Further, 4 cases of applying the approach were considered:

In the first case, using cohort analysis and testing hypotheses with an experiment, a new way of positioning a product was created, which consisted of dividing the product into 3 separate solutions, and then promoting each of these solutions.

In the second case, a financial model was created, with the help of which a decision was made on a new approach to monetization, which helped to increase revenue, while not requiring additional marketing costs.

In the third case, based on the analysis of marketing sources, a new way of distributing the marketing budget was created, and the general marketing strategy was also adjusted. The startup's management began to better understand users and refined the product model they were interacting with.

In the fourth case, an example of a failed decision was made based on an unverified hypothesis. At the same time, initially everything looked as if all hypotheses were confirmed by the data, although in fact there were completely different events that misled the management, and as a result, a significant amount of time and effort was spent on a decision that brings relatively little revenue

As a result, the key metrics of the product were improved, and the product itself reached a break-even point. Thus, we can say that when applied correctly, the data-driven decision-making approach helps to increase the likelihood that the consequences of the decisions made will be expected. This, in turn, helps to reduce the uncertainty within which the startup operates. And this is probably the main task of the data-driven decision-making approach - to reduce uncertainty and provide information about which development strategy for a startup is most likely to lead it to success.

7. SUMMARY

In this paper, a data-driven approach to startup management was investigated. In order to investigate the approach, various cases were analyzed and the principles of the approach were derived from them. Then, the methods and tools of the approach were highlighted. These methods and tools were then implemented in a real startup using specific software solutions.

In the case of the studied startup, the approach helped the management to identify the positioning weaknesses, find new points of growth and attract users. However, this was achieved because of careful preparatory work. To implement the approach, not only some business processes had to be redesigned, but also new services had to be added to the application architecture.

It is important to remember that the approach does not guarantee that the decision will be successful, it helps to understand the probability that a particular decision will lead to the consequences that are expected from it. In order to implement the approach, the initial product hypothesis of the startup must be correct as well, otherwise no approach will help the startup become successful.

On the one hand, the approach is very flexible and can be applied in many different industries, but on the other hand, because of this flexibility, the approach cannot be standardized. In each individual case, the approach has to be adapted to the specific case. However, if the product hypothesis is correctly defined and all the preparatory work to implement the approach is done, most likely the data-driven decision-making approach will help startup managers to make the right decisions, so that the startup can grow and develop.

8. REFERENCES

1. Krishna, A. Agrawal and A. Choudhary, "Predicting the Outcome of Startups: Less Failure, More Success," *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, Barcelona, Spain, 2016,pp. 798-805, doi: 10.1109/ICDMW.2016.0118.
2. D.Feinleib, “Why Startups Fail”, 2011 Apress, doi: 10.1007/978-1-4302-4141-6
3. S.Kalogiannidis, F.Chatzitheodoridis “Impact of Covid-19 in the European Start-ups Business and the Idea to Re-energise the Economy,” *Sciedu Press Vol.12, No.2, doi:* 10.5430/ijfr.v12n2p55 URL:
4. Billio, Monica and Varotto, Simone, A New World Post COVID-19: Lessons for Business, the Finance Industry and Policy Makers (July 31, 2020). Venezia Edizioni Ca’ Foscari - Digital Publishing, 2020, doi:10.2139/ssrn.3665230
5. F.Giones, A.Brem, J. M. Pollack, T. L. Michaelis, K. Klyver, J. Brinckmann,"Revising entrepreneurial action in response to exogenous shocks: Considering the COVID-19 pandemic", Journal of Business Venturing Insights,Volume 14,2020,ISSN 2352-6734, doi:10.1016/j.jbvi.2020.e00186.
6. Grandhi, B., Patwa, N. and Saleem, K. (2020), "Data-driven marketing for growth and profitability", *EuroMed Journal of Business*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/EMJB-09-2018-0054
7. Eric R., “The Lean Startup.” No 3(29)2011, ISBN:9780307887917
8. Jamie Pride “Unicorn Tears: Why Startups Fail and How To Avoid It” , ISBN: 0730348695
9. “The Complete List Of Unicorn Companies”, [Electronic resource] .- available at: https://www.cbinsights.com/research-unicorn-companies (accessed: 10.05.2021)
10. “EY Start-up-Barometer Europa”, Die Welt (Welt am Sonntag), April 2020
11. D.Tzabbar, J.Margolis,"Beyond the Startup Stage: The Founding Team’s Human Capital, New Venture’s Stage of Life, Founder–CEO Duality, and Breakthrough Innovation". Organization Science 28 (5) 857-872, doi:10.1287/orsc.2017.1152
12. J.Paschen,"Choose wisely: Crowdfunding through the stages of the startup life cycle",Business Horizons,Volume 60, Issue 2,2017,Pages 179-188,ISSN 0007-6813,doi:10.1016/j.bushor.2016.11.003.
13. J. C. Picken,"From startup to scalable enterprise: Laying the foundation",Business

Horizons,Volume 60, Issue 5,2017,Pages 587-595,ISSN 0007-6813,doi:10.1016/j.bushor.2017.05.002.

14. M. H. Ahmed, C. Fei, V. Li, F. C. Lee and Q. Li, "Startup and control of high efficiency 48/1V sigma converter," *2017 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2017, pp. 2010-2016, doi: 10.1109/ECCE.2017.8096403.
15. M.Cagan, "Inspired: How to Create Tech Products Customers Love (Silicon Valley Product Group) 2nd Edition", 2018, ISBN:1119387507
16. Horowitz, «The Hard Thing About Hard Things: Building a Business When There Are No Easy Answers», 2014, ISBN: 9780062273208
17. Matassi M, Boczkowski PJ, Mitchelstein E. “Domesticating WhatsApp: Family, friends, work, and study in everyday communication”, New Media & Society, 2019;21(10):2183-2200. doi:10.1177/1461444819841890
18. Pompermaier L., Chanin R., Sales A., Prikladnicki R. (2019) MVP Development Process for Software Startups. In: Hyrynsalmi S., Suoranta M., Nguyen-Duc A., Tyrväinen P., Abrahamsson P. (eds) Software Business. ICSOB 2019. Lecture Notes in Business Information Processing, vol 370. Springer, Cham. doi: 10.1007/978-3-030-33742-1_33
19. Arundale, Keith & Mason,”Private Equity & Venture Capital: Riding the COVID-19 Crisis.”, Colin, 2020 Doi:10.30687/978-88-6969-442-4/014.
20. K. C. Wong and Yang Wang, "Pattern discovery: a data driven approach to decision support," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 33, no. 1, pp. 114-124, Feb. 2003, doi: 10.1109/TSMCC.2003.809869.
21. H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 343-347, doi: 10.1109/ICIP.2014.7025068.
22. A.Kusiak,"Innovation: A data-driven approach",International Journal of Production Economics,Volume 122, Issue 1,2009,Pages 440-448,ISSN 0925-5273,doi:10.1016/j.ijpe.2009.06.025.
23. S. Moro, P.Cortez, P.Rita,"A data-driven approach to predict the success of bank telemarketing",Decision Support Systems,Volume 62,2014,Pages 22-31,ISSN 0167-9236,https://doi.org/10.1016/j.dss.2014.03.001.

24. Khanna D., Nguyen-Duc A., Wang X. (2018) “From MVPs to Pivots: A Hypothesis-Driven Journey of Two Software Startups”. In: Wnuk K., Brinkkemper S. (eds) Software Business. ICSOB 2018. Lecture Notes in Business Information Processing, vol 336. Springer, Cham. https://doi.org/10.1007/978-3-030-04840-2_12
25. Behl, A., Dutta, P., Lessmann, S. *et al.* A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. *Inf Syst E-Bus Manage* **17,** 285–318 (2019). https://doi.org/10.1007/s10257-019-00452-5
26. Luca, M., & Bazerman, M. (2020). *The power of experiments*. Cambridge, Massachusetts: The MIT Press. ISBN: 0262043874
27. Fan, W., & Geerts, F. (2012). Foundations of Data Quality Management. *Synthesis Lectures On Data Management*, *4*(5), 1-217. doi: 10.2200/s00439ed1v01y201207dtm030
28. Dixon, W. J., & Massey, F. J., Jr. (1951). *Introduction to statistical analysis.* McGraw-Hill.
29. Peter E. Rossi,Greg M. Allenby, Bayesian Statistics and Marketing. Marketing Science 22 (3) 304-328, 2002 https://doi.org/10.1287/mksc.22.3.304.17739
30. S.Marsh “User Research: A Practical Guide to Designing Better Products and Services”, Kogan Page Publishers, 2018, ISBN: 0749481048
31. Grimm, P. (2010). Social Desirability Bias. In Wiley International Encyclopedia of Marketing (eds J. Sheth and N. Malhotra). https://doi.org/10.1002/9781444316568.wiem02057
32. E. Kurilovas,"On data-driven decision-making for quality education",Computers in Human Behavior,Volume 107,2020,105774,ISSN 0747-5632,https://doi.org/10.1016/j.chb.2018.11.003.
33. Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. A Review of Data-Driven Decision-Making Methods for Industry 4.0 Maintenance Applications. *Electronics*. 2021; 10(7):828. https://doi.org/10.3390/electronics10070828
34. P.Seele,"Predictive Sustainability Control: A review assessing the potential to transfer big data driven ‘predictive policing’ to corporate sustainability management",Journal of Cleaner Production,Volume 153,2017,Pages 673-686,ISSN 0959-6526,https://doi.org/10.1016/j.jclepro.2016.10.175.
35. M.Jeffery, “Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should

Know 1st Edition”. 2010, ISBN: 978-0-470-50454-3

36. ”The Retention Analysis chart: an overview” [Electronic resource] .- available at: https://help.amplitude.com/hc/en-us/articles/230543327-The-Retention-Analysis-chart-an-overview (accessed: 07.05.2021)
37. Kemell KK., Wang X., Nguyen-Duc A., Grendus J., Tuunanen T., Abrahamsson P. (2020) Startup Metrics That Tech Entrepreneurs Need to Know. In: Nguyen-Duc A., Münch J., Prikladnicki R., Wang X., Abrahamsson P. (eds) Fundamentals of Software Startups. Springer, Cham. https://doi.org/10.1007/978-3-030-35983-6_7
38. Chou, D.C., Bindu Tripuramallu, H. and Chou, A.Y. (2005), "BI and ERP integration", Information Management & Computer Security, Vol. 13 No. 5, pp. 340-349. https://doi.org/10.1108/09685220510627241
39. Bresciani, Sabrina & Eppler, Martin J.: Gartner's Magic Quadrant and Hype Cycle. ecch reference no: 908-029-1, 2009.
40. S.Oved, “Welcome to AppsFlyer” [Electronic resource] .- available at: https://support.appsflyer.com/hc/en-us/articles/212017846-Welcome-to-AppsFlyer (accessed: 05.05.2021)
41. J.Steele, N.To, “The Android Developer's Cookbook: Building Applications with the Android SDK (Developer's Library) 1st Edition”, Addison-Wesley Professional, 2010, ISBN: 9780132478014
42. ”The State of Developer Ecosystem 2020” [Electronic resource] .- available at: https://www.jetbrains.com/lp/devecosystem-2020/ (accessed: 07.05.2021)
43. J. Rumbaugh, I. Jacobson, and G. Booch. 2004. “Unified Modeling Language Reference Manual, The (2nd Edition)”. Pearson Higher Education., isbn: 0321245628
44. “User guide” [Electronic resource].- available at: https://pandas.pydata.org/docs/user_guide/index .html (accessed: 07.05.2021)
45. “NumPy Documentation” [Electronic resource].- available at: https://numpy.org/doc/ (accessed: 07.05.2021)
46. “Dask” [Electronic resource].- available at: https://docs.dask.org/en/latest/ (accessed: 07.05.2021)
47. “Matplotlib: Visualization with Python” [Electronic resource].- available at: https://matplotlib.org/ (accessed: 07.05.2021)
48. Hubbard, D. *How to measure anything workbook*. Hoboken, New Jersey: Wiley. 2014.

ISBN: 9781118539279

49. Fiorini, P., & Lipsky, L.. Search marketing traffic and performance models. *Computer Standards & Interfaces*, *34*(6), 517-526, 2012 .doi: 10.1016/j.csi.2011.10.008

50. Sawyer, T, *Financial Modeling for Business Owners and Entrepreneurs [recurso electrónico]*. Berkeley, CA. 2015. ISBN: 978-1-4842-0370-5