МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ КОРАБЛЕБУДУВАННЯ
ім. адмірала Макарова
ФАКУЛЬТЕТ ЕКОНОМІКИ МОРЯ
КАФЕДРА ЕКОНОМІЧНОЇ ПОЛІТИКИ ТА БЕЗПЕКИ

«Допустити до захисту»
зав. кафедри
д.е.н., професор НУК Єфімова Г.В.
(наукова ступінь, вчене звання, П.І.Б. завкафедрою)

_____ «___» _____ 2022 р.
(підпис зав. кафедрою)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття ступеню вищої освіти **бакалавр**

на тему: **Великі Дані в Міжнародному Бізнесі: Застосування Великих Даних для стартапів на ранніх стадіях розвитку**

Виконала: студентка IV курсу, групи 4451
Спеціальності 051 Економіка
ОП Міжнародна економіка
(шифр і назва спеціальності)
Казимир Д.В.
(прізвище та ініціали)
Керівник д.е.н., проф. Єфімова Г.В.
(прізвище та ініціали)

The qualification work is developed in the framework of ERASMUS+ CBHE project "Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan" / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.
Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.

Миколаїв – 2022

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
ADMIRAL MAKAROV NATIONAL UNIVERSITY OF SHIPBUILDING
FACULTY OF ECONOMICS OF SEA
DEPARTMENT OF ECONOMIC POLICY AND SAFETY

«Допустити до захисту»
зав. кафедри
д.е.н., професор НУК Єфімова Г.В.
(наукова ступінь, вчене звання, П.І.Б. завкафедрою)

_____ «___» _____ 2022 р.
(підпис зав. кафедрою)

BACHELOR'S THESIS PAPER

on the topic

Big Data in International Business:

Application of Big Data for Early-Stage Startups

Made by 4th-year student, group 4451
Study Program 051 Economics
International Economics
Kazymyr Daria
Supervisor Efimova Hanna, PhD

The qualification work is developed in the framework of ERASMUS+ CBHE project "Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan" / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.
Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.

Mykolayiv – 2022

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ КОРАБЛЕБУДУВАННЯ
ІМЕНІ АДМІРАЛА МАКАРОВА

Факультет економіки моря
Кафедра економічної політики та безпеки
Освітній рівень – перший (бакалаврський)
Галузь знань 05 Соціальні та поведінкові науки
Спеціальність 051 Економіка
Освітня програма – Міжнародна економіка

ЗАТВЕРДЖУЮ
Завідувач кафедри економічної політики та безпеки
_____ Г. В. Єфімова
«____» _____ 2022 року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ
на здобуття ступеня вищої освіти «бакалавр»

Студенту Казимир Дар'я Віталіївна
(Прізвище, ім'я, по батькові)

1. Тема роботи: Великі Дані в Міжнародному Бізнесі: Застосування Великих Даних для стартапів на ранніх стадіях розвитку / Big Data in International Business: Application of Big Data for Early-Stage Startups
Керівник роботи д.е.н., проф. Єфімова Г. В.
Затверджені наказом ректора №__уч від «22» квітня 2022 року
2. Термін подання роботи: 16.06.2022 р.
3. Вихідні дані по роботі: навчальні посібники, наукові праці, монографічні дослідження вітчизняних та зарубіжних вчених, публікації міжнародних організацій, внутрішні корпоративні документи підприємства, а також електронні ресурси
4. Перелік питань, що належать до розробки (найменування розділів) Теоретичні основи застосування великих даних для стартапів на ранніх стадіях розвитку, Практичні основи застосування великих даних для стартапів на ранніх стадіях розвитку, Аналітичні основи застосування великих даних для стартапів на ранніх стадіях розвитку.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Визначення наукового керівника роботи	01.09.2021р. – 15.09.2021р.	
2.	Вибір теми роботи та її узгодження з науковим керівником	15.09.2021р. – 30.09.2021р.	
3.	Вивчення друкованих та електронних джерел, економічних реалій, методичних та наукових видань з теми роботи	01.10.2021р. – 31.10.2021р.	
4.	Складання попереднього плану роботи, узгодження його з науковим керівником	1.11.2021р. – 21.11.2021р.	
5.	Збір статистичної інформації на базовому підприємстві (установі, організації)	21.11.2021р. – 31.01.2022р.	
6.	Розробка теоретичного розділу	1.02.2022р. – 28.02.2022р.	
7.	Розробка аналітичного розділу	1.03.2022р. – 31.03.2022р.	
8.	Розробка проектного розділу	1.04.2022р. – 30.04.2022р.	
9.	Розробка вступу, висновків, списку використаної літератури та додатків	05.05.2022р. – 15.05.2022р.	
10.	Редагування рукопису роботи та ознайомлення з ним наукового керівника	16.05.2022р. – 30.05.2022р.	
11.	Усунення зауважень наукового керівника та завершення роботи	31.05.2022р. – 15.06.2022р.	
12.	Подання рукопису кваліфікаційної роботи, демонстраційного матеріалу та доповіді на кафедральний захист	16.06.2022р.	
13.	Усунення зауважень кафедрального захисту	16.06.2022р. – 19.06.2022р.	
14.	Розробка проекту демонстраційного матеріалу та доповіді	20.06.2022р. – 24.06.2022р.	
15.	Подання роботи рецензенту та отримання рецензії	25.06.2022р. – 27.06.2022р.	
16.	Захист роботи перед АК	30.06.2022р.	

Науковий керівник _____ Єфімова Г. В.
(підпис) (прізвище та ініціали)

Студент _____ Казимир Д. В.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Казимир Д.В. Великі Дані в Міжнародному Бізнесі: Застосування Великих Даних для стартапів на ранніх стадіях розвитку – бакалаврська робота на здобуття ступеня вищої освіти «бакалавр». – Національний університет кораблебудування імені адмірала Макарова, м. Миколаїв, 2022.

Ця бакалаврська робота присвячена поглибленому дослідженню Великих Даних – сучасного бізнес-інструменту, що працює в міжнародному масштабі. Поставивши запитання, чи були Великі Дані лише швидкозмінною тенденцією, ця дипломна робота доводить, що застосування Великих Даних здатне принести компанії переваги над її конкурентами, а також може служити основною передумовою для ефективної діяльності компанії. Поглиблене дослідження Великих Даних у контексті міжнародного бізнесу проводилося поряд з аналітичною та практичною роботою з компанією, яка підпадає під визначення стартапу на ранній стадії розвитку.

З метою навігації читачів у сфері Великих Даних, було насамперед детально розглянуто концепцію, а також описано використання Великих Даних в міжнародному бізнесі. Однак, беручи до уваги, що стартапи на ранній стадії мають свою унікальність, загальне застосування Великих Даних було пояснене трьома дискурсами: стартапи на ранній стадії, ідея продукту яких фундаментально базується на роботі з Великими Даними; стартапи, які працюють із Великими Даними для підвищення своєї ефективності; стартапи, чия ідея продукту походить від Великих Даних.

Маючи визначений шлях до ефективного застосування Великих Даних для стартапів та розрізняючи таке застосування для корпорацій, ця бакалаврська робота аналізує застосування стартапом на ранній стадії розвитку FitSit та формулює найбільш релевантні цільові групи для продукту цього стартапу, даючи чітку сегментацію їх потенційних клієнтів.

SUMMARY

Kazymyr D.V. Big Data in International Business: Application of Big Data for Early-Stage Startups if a thesis paper to obtain a Bachelor's degree at the National University of Shipbuilding named after Admiral Makarov, Mykolayiv, 2022.

This thesis paper is dedicated to the in-depth research of such a modern tool for the businesses working on an international scale as Big Data. Whilst one could wonder if Big Data was simply a matter of a fast-changing trend, this thesis paper proves the argument that application of Big Data is able to indeed bring a company leading advantage over its competitors and may as well serve as the main prerequisite for the company's efficient operations. In-depth research of Big Data in the context of International Business was held alongside analytical and practical field work with a company which falls under the definition of an early-stage startup.

With an intention to navigate the readers in the field of Big Data, first of all, its concept was examined in detail, and its usage in international business-making was described. However, bearing in mind that early-stage startups have their unique specificity, the overall framework of application of Big Data was explained within three main discourses: early-stage startups whose product idea is fundamentally based on the work with Big Data itself, early-stage startups that operate with Big Data to increase their efficiency, early-stage startups whose product idea comes from Data Science.

Having the path to effective application of Big Data for early-stage startups defined and the culture of application of Big Data for early-stage startups and for established on the market corporations differentiated, this thesis paper analyzes the ways how Big Data is being used within a case study of an early-stage startup FitSit. It, thus, formulates and analyzes the most relevant target groups for the startup's product and gives definite predications for the customer segmentation.

ЗМІСТ

TABLE OF CONTENTS

INTRODUCTION

There is no doubt, the world keeps evolving, if not milestone by milestone, then for the least part – step by step. The great minds, moreover, keep contributing to the process of the development of the world around, the life within, and alongside all the major changes happening around the globe by itself. Just as the telescope has given humanity an ability to understand the Universe and the microscope has given mankind insight into microbes, new methods of collecting and analyzing vast amounts of data help us understand the world around us in many ways we are only starting to appreciate. As Andreas Weigend in his prominent work Data for the People says, the real revolution, however, is not be about the computers that compute the data, but about the data itself and how we use it (Weigend, 2015).

Big Data marks the beginning of profound changes. Profound challenges never come alone, though. As an example, with the need to upgrade the equipment for a company or a manufacturer also comes the need to teach the employees to use this equipment. With an opportunity, but also a challenge, to use Big Data, comes a challenge of obtaining and fulfilling the full potential of it. The ability to collect and analyze Big Data has become one of the major breakthroughs of the last decade, and mainly not for the personal or individual use but the commercial and manufacturing ones.

This is where the *relevancy* of the chosen topic steps to the forefront. While there exists a lot of information within the scholarly works that is able to prove the benefits of Big Data as a whole and in general, there still is an open gap for the examination of the importance of application of Big Data for the entrepreneurial purpose, specifically in order to lead International Businesses. What is more, the preliminary research for the topic has shown that even the most modern sources lack an in-depth examination of the importance of application of Big Data for the startups, and even more precisely – early-stage ones. This fact can be supported by the lack of numeric data on the number of various existing startups that are actively using Big Data concepts. Moreover, there is an explanation for the phenomena. It is being believed that early-

stage startups are a lot more likely to be saving their primary resources – both financial and human capital – and not actively using Big Data principles for potentially bettering their products or operations within their businesses.

This thesis paper, therefore, explains why such an approach is rather harmful and inefficient for early-stage startups, and why this can be considered as a mistake which might cost the young entrepreneurs more in comparison to the finances invested into working with Big Data. Large corporations started asking this question of the benefits of application of Big Data years ago but at the time only few saw the real prospects of Big Data analytics. In 2015, only 17% of companies around the world used Big Data in their work. To show the contrast, all large companies are now using Big Data analytics: in the US, more than 55% of companies in a variety of fields are working with this technology. In Europe and Asia, the demand for big data is slightly lower - about 53%. It turns out that over the last five years, businesses started using Big Data three times more. The research question is thus relevant to explore if it was for valid reasons that such attention was given to Big Data. IT, banking, and telecommunications companies were the pioneering industries where they were implementing Big Data. However, these findings are not surprising because it is these sectors that accumulate the largest volume of data whatsoever – banks accumulate huge volume of data through transactions, telecommunication companies accumulate it through geodata, and search engines accumulate it through query histories. This thesis paper, therefore, was curated in a way to address the common misconception that Big Data is only used within the mentioned fields, and it answers the given task by taking a close look at an early-stage startup from the sector of digital health with a high-tech product which is still at the Minimum Viable Stage of production.

The *topic* Big Data in International Business: Application of Big Data for Early-Stage Startups clearly sets the *goals* for this thesis paper. It is of our first and foremost importance to elaborate on the usage of Big Data for companies which are innovative and show great potential, and yet have undefined future in terms of their business paths and own financial stability. Such companies are defined as early-stage startups. What is more, the second goal of this paper is to take a close look at a company which would

represent the group of early-stage startups, and work with it in order to give practical and analytical overview of this company's business behavior with Big Data, as well as to suggest ways which could improve its business-making.

The *object* for the research within this thesis paper is finding ways for better analytical approach to business-making for early-stage companies.

The *subject* for the research within this thesis paper is Big Data itself, as the phenomena that has unlimited potential to contribute to the improvement of short and long-term operations for early-stage companies.

The *methodology* behind the research within this thesis paper has been chosen as the usage of Big Data open to the publicity from scholarly resources, mainly those being research papers published in academic journals, as well as the surveys which were conducted among the whole population of such a country as the United States. To compute the gathered weighted data, the variance estimation method was then used.

This Bachelor's thesis paper will summarize the main information about the concept of Big Data, which, with the right approach, can be largely used as a ground base for not only new products and services, but also high involvement in the overall innovation culture. Some might argue that the topic of Big Data itself is rather an overexplained one – both in the academic sources and popular literature. This paper therefore decided to not only take the very well-known sources of information into consideration for this work, such as Data for the People by Weigend Andreas and Big Data: A Revolution That Will Transform How We Live, Work, and Think by Mayer-Schönberger and Cukier. We're also looking at the sources which go beyond the most popular findings and those that tended to focus on the new approaches to the usage of Big Data and the application of this concept. For this reason, scholarly articles from academic journals as well as a proven track of internet sources with the most up-to-date information were used, and largely contributed to the development of the discourse of the topic Big Data in International Business: Application of Big Data for Early-Stage Startups.

The paper consists of 81 pages, 7 pictures, 3 attachments, 3 tables and 49 resources.

CHAPTER 1.
THEORETICAL FRAMEWORK OF APPLICATION OF BIG DATA FOR EARLY-STAGE STARTUPS

1.1. The Big Data concept and its implementation in international business-making

"Big data is the most valuable commodity of the 21st century, it is the new oil and gas."
Andreas Weigend, former Chief Data Scientist at Amazon.com and the author of Data for the People, in his 2011 talk for the UN Global Pulse programme (Weigend, 2012).

"Big Data will spell the death of customer segmentation and force the marketer to understand each customer as an individual within 18 months or risk being left in the dust."
Ginni Rometty, the ex-CEO of IBM back in 2013 at the 'CMO+CIO Leadership Symposium' in Sydney (Rometty, 2013).

"*Data is the new science. Big Data holds the answers.*"
Pat Gelsinger, an American business executive and engineer currently serving as CEO of Intel (Gelsinger, 2012).

While the most prominent C-level professionals of the world's biggest and most influential technology companies, as well as non-tech related businesses, have been publicly talking about the vital importance of the Big Data over the course of the previous 10 years, there still exist several misconceptions about the whole concept of what the Big Data actually is. The term Big Data itself appeared in 2008 in the journal Nature in an article authored by Clifford Lynch (Lynch, 2008). Although the term was initially introduced to academia and primarily referred to the growth and diversity of scientific data starting from the year 2009, the naming of Big Data was actually widely

disseminated in the business press within the next few years, one of the most widespread examples being the year 2010 when the first products and solutions referring exclusively to the problem of Big Data processing were introduced. By 2011, most of the major information technology vendors and organizations have already been using the concept of Big Data in their business strategies and operational models, including, as can be understood from the quotes previously, IBM, Intel, Amazon.com, but also Microsoft, Hewlett-Packard, EMC, Oracle, and many more. Back in the 2010s, the major information technology market analysts dedicated specified research to the conceptualization of Big Data, and yet, there still is some in-between definition of Big Data itself, as well as its framework.

In order to address this question, we have to rely on the academic published papers that address Big Data from the perspective of Economics and Business, as well as the general sources that look at the question of Big Data with a more popularized approach. One of such is the work by Viktor Mayer-Schönberger and Kenneth Cukier 'Big Data. A Revolution That Will Transform How We Live, Work, and Think'. According to it, Big Data is characterized by the large volume of structured and unstructured data generated every minute in the digital environment (Mayer-Schönberger, 2013). In a broad sense, however, Big Data is spoken of as a socio-economic phenomenon associated with the emergence of technological capabilities to analyze huge amounts of basic and complex data or even the entire world volume of data. Moreover, Big Data is also the transformational consequences that can be spotlighted as a result of the computational and statistical power of its interpretation. Indeed, the perception of Big Data within academic papers can be summarized as the one that Big Data refers to structured and unstructured data of huge volumes and significant diversity, effectively handled by horizontally scalable software tools that emerged in the late 2000s and alternatives to traditional database management systems and Business Intelligence solutions (Preimesberger, 2011).

Nowadays, users generate data at many points in a day throughout various when they download or open an application, search for information on the Internet, shop in online stores, cross state borders, or travel to a location with a smartphone in their

pocket. The result is a vast amount of valuable information that companies collect, analyze, and visualize. And it should come as no surprise that our actions online do indeed accurately reflect regular traits of our personalities - people remain themselves, not just when they use websites incognito, and even when they carefully edit their social media profiles. Therefore, algorithms can so easily assess a person's basic character traits (Bachrach, 2012).

There are three criteria to clearly differentiate Big Data from just a compiled list of regular data (Preimesberger, 2011):

- Data must be digital. Books in a national library or stacks of documents in a company's archive are data, and often a lot of it. But the term big data only means digital data that is stored on servers.
- Data must come in objectively large volumes and accumulate quickly. It’s never separated but always compiled with the help of a bigger number of sources.
- Data must be heterogeneous and loosely structured. Orders in an online store are streamlined, and additional statistical parameters are easily extracted from them, such as the average receipt or the most popular products. Therefore, this data does not qualify as Big Data. This criterion is not always mandatory. Sometimes large amounts of structured data that are constantly being replenished are considered big data, especially if they are used for Machine Learning or identifying non-obvious patterns. That is, if Big Data analysis techniques are applied to structured data, we can say that this is Big Data.

So, big data is hard-to-analyze digital information that accumulates over time and comes to the final destination in solid portions, either artificially or being evolved naturally into the making process.

To sum the previous paragraphs up, Big Data is a complex and extensive collection of diverse information. However, this information is presented raw and requires preprocessing to extract valuable pieces that can potentially benefit businesses and organizations, which is a topic this thesis paper will cover in the next chapters.

The second work that this thesis paper will keep circling back to is Andreas Weigend's Data for the people: how to make our post-privacy economy work for you. Weigend finds the analogy between Big Data and oil as the main commodity human society depends on. For more than a century, the economy and socio-political life have been largely determined by oil and the development of technologies to extract, store and process it into products consumed by every inhabitant of the planet. Today, the possibility of processing personal data into products and services is bringing changes to human life comparable to the impact of the Industrial Revolution (Weigend, 2015). Fortunately, information as a resource is radically different from oil. The planet's supply of oil is finite and definite, and as this natural resource becomes more depleted, it automatically becomes more expensive to extract it. In contrast, the amount of information is growing exponentially and the cost of technology to transmit and process it is steadily decreasing (Weigend, 2015). Meaning that there is no point, nor that there is a foreseeable future for this point, when the new oil of the 21st century, as discussed earlier, data and information, will ever be finite.

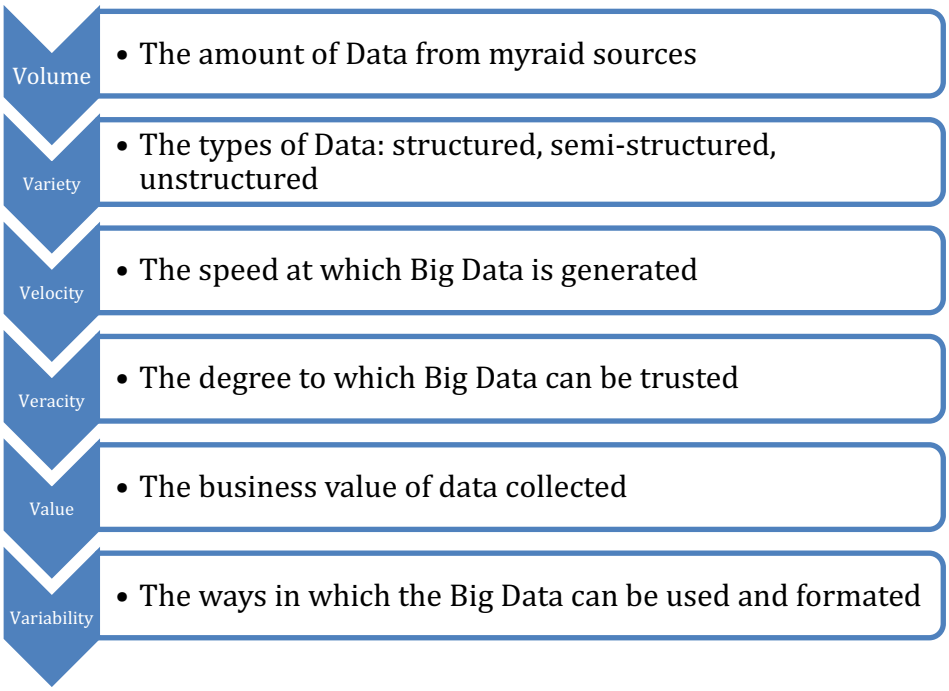
The Information Revolution is another term used in contrast to the Industrial Revolution. Being increasingly rapid in the speed of growth, the Information Revolution reflects the revolutionary impact of information technology on all spheres of society in the last quarter of the 20th century. This phenomenon integrates the effects of previous revolutionary inventions in the field of information (book printing, telephony, radio communications, the personal computer) because it creates a technological basis for overcoming any distance in the transmission of information, which contributes to the unification of the intellectual abilities and spiritual forces of mankind.

As early as 2013, the trend was already demonstrated that 90% of all data in the world was generated in the two years prior to that, so to say that 90% of all data in the world in 2013 was generated in 2011 and 2012 (Science Daily, 2013). Moreover, the number that shows the staggering statistics of data growth has only been steadily increasing in our age. Taking further look at some examples, we could observe the trends of the year 2020 when each person generated 1.7 MB of data every second

(DOMO, 2020). With the increase of the influence and the popularity of social networks, the previously mentioned trend of enormous piling of content and data is also very vivid and visible. For instance, it is being said by some of the most influential digital agencies that every day approximately 350 million photos are to be uploaded on Facebook (Omni Core Agency, 2020), and more than 6 million tweets are to be created on Twitter, 100 million photos and videos are to be shared on Instagram and 350 billion emails are to be sent (Internet Live Stats, 2021).

However, sharing images and text information is not enough to fully understand the scale of Big Data and Big Data as a concept itself. The earlier mentioned Facebook generates 4 petabytes of data every day (Kinsta, 2020). Google processes more than 6.5 billion search queries every day (Internet Live Stats, 2021). Perhaps, that is also the reason why the phrase "Let me check that on the internet" sounds stranger to us than if we say, "Let me Google that". Once global data started growing exponentially a decade ago, it has shown no signs of slowing down. Big Data is aggregated mainly through the internet, including social networks, web search requests, text messages, piles of visuals like pictures and videos, and media files.

Information, unlike oil, will never run out, – summarizes Weigend (Weigend, 2015). From a scientific and statistical perspective, this can be demonstrated by the ever-increasing number of Internet users doing everything online and using various connected devices and embedded systems. For example, by 2025, people will generate 463 more exabytes of data every day (Raconteur, 2020). Therefore, data may no longer be considered static or obsolete quantities that become useless once a certain goal is reached. Rather, they have become the raw material of the enterprise, an essential economic asset used to generate new economic benefits. It turns out that, with the right approach, they can be cleverly reused as a source of innovation and new services. Data can reveal secrets to those with the humility and willingness to listen, as well as the necessary tools (Mayer-Schönberger, 2013).



Picture 1.1. The six Vs of Big Data

Last but not least, when conceptualizing Big Data, we have to remember its characteristics. As shown on the Picture 1.1., a set of 3V's (such as volume, velocity, and variety) attributes was originally developed by Meta Group in 2001 outside the context of the idea of Big Data as a certain series of information and technological methods and tools. It, due to the growing popularity of the concept of a central data warehouse for organizations, noted the equal importance of data management problems in all three aspects (Preimesberger, 2011). Further interpretations appeared with the four V's (adding veracity, used in IBM's marketing materials), five and six V's (adding viability and value), and even seven and eight V's (adding variability and visualization). In all these cases of characteristics, these attributes emphasize that the defining feature of Big Data is not only its physical volume but also many other categories that turn out to be essential to represent the complexity of the task of Big Data processing itself, as well as its further analysis (Weigend, 2015).

There seems to be a mutual unspoken understanding of the need to use Big Data. One can explain it as the way to lead business more efficiently, claiming that each new technology makes us progress in business operations. However, this is a very generic explanation that most of the current resources offer, however much we could argue that

in reality everything is broken down to that very basic need of leading more efficiently and at lower costs to be met. It's not just the data itself that we have to take into consideration, but also the principles of working with it because they also can serve as the ground for further progress for us as business owners or management (Preimesberger, 2011). Otherwise, we would argue that so-called Small Data, for example, the one that you have from your own customer base as a business owner, would be self-sufficient enough to build business models based on homogeneous information or a very small number base. As discussed in the last chapter, Big Data is hard-to-analyze because of its scale of digital information that is being accumulated over time and that comes to you in solid portions. That is precisely why we have to take a further look at the ways of the implementation of Big Data in business-making but also at the principles of working with it within the functionality of the given business, specifically an international one.

First of all, Big Data as an end-result technology and as a principle of work come most often in hand when a deeper process analysis is required for a business. Without a doubt, collecting and analyzing information when following preset metrics allows simple and straightforward adjustments to be made in a timely manner. Such improvements almost always yield tangible results almost immediately. However, as the world keeps evolving, as the engineers keep making progress in the technological field, and as the Internet of Things keeps scaling and increasing in volume and power, we get new ways to operate a business – both in terms of setting new KPIs and metrics and collecting information to be able to evaluate the data we get later on. After all, this is the very basic business operations reality: once the product or service reaches the point of stagnation, the business owner or the responsible manager has to find new ways to develop their operations inside the company to indirectly influence the quality or the quantity of their product or service. That is why analysts, both internal and external, collect and analyze new data which is not directly related to the main metrics of projects, and are very likely to find not the regular small amount of data, but rather Big Data (Malviya, 2018). For example, in e-commerce businesses collect data on cursor movements of the users across the screen. Another example is social media

algorithms which are taught to analyze words, pictures, and emojis in customers' comments or posts on social networks in order to take steps to build or increase customer loyalty.

Even though such data is not directly related to basic operational tasks and the immediate scope of work that are common to meet the regular business metrics, they are very, if not the most, powerful sources of information for business making, especially when leading a company on the international arena. Each country has its businesses operating differently, and even more than that – its customer base, customer demographics, preferences, and decision-making journey are very different from country to country, if not city to city. However, having the modern-day analysis which is supported by the usage of modern principles and technologies, such as Big Data, is very likely to optimize the project to a very high scale. As Andreas Weigend argues, working with such Big Data is like searching for oil – you have to try different places, and apply different strategies to find and extract the hidden resources hidden in that data. Not all attempts will be successful, but in the end, the findings can bring a lot of benefits (Weigend, 2015).

The first problem of international businesses that Big Data is solving is, as mainly described earlier, to analyze the current state of affairs inside the company and optimize its business processes. Big Data can be used to understand which products customers tend to like and prefer more, whether all the technologies and machines of the manufacturer work optimally and efficiently in production, or whether there are any problems with the supply chain for the company’s goods. The most common way is to look at the big pile of Big Data to observe if there are any patterns in that data, to build graphs and charts based on the given data, and later on generate sufficient whole reports. For instance, Big Data helped Intel to find out that they were doing a lot of unnecessary tests when manufacturing their processors. By analyzing Big Data of the company’s then present and history from their dozens of data-intensive projects on the manufacturing facilities they eliminated which tests, in reality, were unnecessary, and this process helped Intel to save about $30 billion in 2013-2014 (Malviya, 2018). As it is well-aware, one could never talk about the framework of the application of Big Data

for early-stage startups only addressing the question theory-wise and having no elaboration on the real-world examples from the market. That's why this chapter will also explore the actual examples of startups who have based their production on the application of Big Data: both in terms of the actual underlying concept for the idea behind the product, and as for the way to scale their business development processes internally and externally for the company.

The second problem of international businesses that Big Data is solving is making predictions. Big Data about past events, such as sales, customer-product interaction, marketing campaign execution rates, price ranges, etc. helps businesses to draw conclusions, as well as predict the future. The concepts of forecasting and analytics are often understood as being synonymous. However, this is incorrect and should be diversified when making predictions about future events. Predictive analytics uses Big Data and is based upon many techniques such as Data Mining, Statistics, Modeling, Machine Learning, and Artificial Intelligence (Malviya, 2018). The key point here is that there are a number of possible methods that can be used to make predictions about the future, but they all are interconnected to Big Data, if not simply impossible to be carried without it. Simply put, the solution to the problem of incorrect prediction making is predictive analytics which is the detection of patterns in Big Data using AI algorithms to predict demand for certain products, segmenting contacts, pricing, and much more (Weigend, 2015). It can be used in various fields, both to make decisions about the volume of production or storage of a particular product in stock and to automatically personalize offers based on the information available about the user. In retail, to even broaden the topic, predictive analytics is used to solve two main tasks: predictive segmentation and conversion probability prediction (Weigend, 2015). To set an example, a logistics company launched the Transportation Management Center using Big Data. They ended up predicting warehouse utilization, meaning, predicting when warehouses would be full and when they would be empty. This helped to plan transport routes and avoid downtime. For example, Pachama (pachama.com) is a technology company with a mission is to bring nature back to life to fight climate change. They do so with a vision to address climate change by reducing the number of

greenhouse gases in the atmosphere, for which they protect and replant forests. The usage of Big Data is interesting in Pachama because they automate the process of counting and observing trees, using lidar, satellite imagery and artificial intelligence to accurately measure changes in forest cover around the world. The platform monitors trees lost to illegal deforestation or wildfires, and as new trees grow, the artificial intelligence system estimates total biomass to understand the full carbon benefit at any given time (MasterBorn, 2022). The company has also built a model to track annual trends over time, particularly in the Amazon rainforest (MasterBorn, 2022).

Thirdly, Big Data helps international businesses solve the need of building accurate models. Based on this kind of data, it is now possible to build a computerized digital model of a physical store, equipment, or oil well. Afterward, this technology also lets the businesses experiment with their artificial models: they can change something in them to better, improve and perfect the outcome, they can track different indicators, or even speed up or slow down various processes in order to analyze them. building accurate models also brings the businesses to the solution of the usage of statistical models. Models make it possible to identify relationships between variables and to understand how variables, working on their own and together, influence an overall system, they also allow one to make predictions and assess their uncertainty (National Academies of Sciences, Engineering, and Medicine, 2013). The major benefit of Big Data when it comes to building statistical models for the benefit of businesses is that while sampling error decreases with increasing sample size, bias does not—big data does not help overcome bad bias (National Academies of Sciences, Engineering, and Medicine, 2013). Having stated that, we should automatically assume that Big Data is the only probable solution to the question of statistics when operating businesses. However, as the earlier quoted research paper states, with massive data sets, it may be easy to find many “statistically significant” results and effects. Whether these findings have substantive relevance, however, becomes an important question, and statistically significant correlations and effects must be evaluated with subject-matter knowledge and experience (National Academies of Sciences, Engineering, and Medicine, 2013). Computational efficiency, therefore, can and should be strived to be

achieved with a consideration that computational complexity is an important factor in Big Data analysis. An example for this problem that this thesis paper wants to reveal is a startup GeoSure (geosureglobal.com). The company has set out to answer the question of safety, specifically a problem of measuring safety around you, by creating a trusted, detailed, real-time security measurement platform (MasterBorn, 2022). They say: Travelers know where the water is safe. But what about the streets? Now, using a cell phone or a smartwatch, the user of GeoSure can measure his or her security at any time around him, and the location could be worldwide, meaning it can be done anywhere in the world. GeoSure now provides safety rankings for at least 65,000 cities and towns around the world. They took such important factors into account as basic freedoms, health and medical care, level of robbery and theft, women's safety, and LGBTQ+ rights protection and safety. The ratings are based on Big Data, which is then verified using advanced statistical algorithms to give the most detailed and accurate location security ratings. Using advanced Artificial Intelligence algorithms, the platform can assess the current measures of security in the place, and it also can predict upcoming threats from the upcoming data to the application. As well as another example being Suplari (suplari.com), a startup that utilizes Big Data for the predictive analytics. Its selling point lies within the innovative accountability and financial performance analysis. These two offerings help the business executives, as well as individual customers to better understand cost effectiveness and evaluate responsiveness, so the product becomes very attracting on the market. Suplari says that they combine Data, Insights, Actions, and then it all equals to Spend Agility. They also say that enterprise Procure-to-Pay (P2P) suites are great, but the users have to wait long times for spend intelligence. The approach of Suplari is different, as they say themselves: it is spent agility, when by using clean data, automated insights, and predictive actions they can offer their clients to continuously manage costs and cash flow.

The fourth problem there exists that can be solved with Big Data when operating businesses on an international scale is automating the routine. There are several algorithms and automotive programs that are powered by Artificial Intelligence and

facilitated with Machine Learning that are taught to do certain tasks such as sorting documents or coming up with the most efficient answers for the customers in the quick response surveys. These algorithms and automotive programs are, first of all, very common specifically for startups, and not necessarily the startups that are on the high-tech level but even the early-stage startups, and, secondly, they are good examples for this paper because they are created with nothing else, but Big Data as a leading technology. Primitive algorithms from voice assistants to neural networks powered by Artificial Intelligence are good examples for the mentioned problem. As an example, to give, many companies have automated there routine in terms of the HR processes and made their recruitment faster and more self-sufficient by introducing a robotic recruiter into their workflow. This robot is able to do simple recruiting work by recognizing the voice, sorting resumes, asking simple questions and accepting answers. This way, human recruiters are given more complex and creative tasks, and not the ordinary manual work, for instance, of actual interviewing processes and the final selection of candidates.

One more major problem that the companies are able to solve with Big Data is tiling Behavior Analysis. Even though this question is not raised in many scientific works, this thesis paper considers this topic of highest important to separately take a look at due to its prominence, but also due to apparent inability of every other source to solve this need for the businesses. As far as traditional statistics are concerned, an event that has a very low probability rate, e.g., only 0.1 percent probability of occurring, will be most likely regarded by managers within the companies and by scientists when doing research as a rare event, and therefore it can be easily left out and ignored. However, as the earlier quoted research paper states, in the realm of massive data, these so-called “rare events” can occur sufficiently frequently that they deserve special attention. (National Academies of Sciences, Engineering, and Medicine, 2013). This means that in general the analysis of tail behavior becomes a key challenge, and it is the Big Data that can position itself on the upfront and solve the issue of Behavior Analysis more efficiently and in a shorter time with all due respect to its complexity and sizing.

Lastly, this thesis paper will mention the problem of large-scale optimization because this is the topic of the upmost importance for absolutely every business, no matter the scale and the field of work, no matter if it is a startup or a big corporate. The direction of the large-scale optimization which this thesis concerns the most is online optimization algorithms. As commonly aware, the current state of a set of parameters of the underlying statistical model should be continuously updates and stored with following the best regulations and guidelines. Moreover, even though there are some existing dynamic modeling tools to handle non-stationary data, they concern themselves mostly with modest-sized data streams (National Academies of Sciences, Engineering, and Medicine, 2013). This is where online model-building algorithms come in very handy and become vital under such conditions. Online methods try to address the time and space complexity simultaneously (National Academies of Sciences, Engineering, and Medicine, 2013). The latter is achieved by not requiring all data to be in memory simultaneously, which alleviates data-storage requirement (National Academies of Sciences, Engineering, and Medicine, 2013). The conclusion which we can make is that for the businesses that choose to operate Big Data to contribute to large-scale optimization it is necessary to study requirements in streaming applications by studying models that can be efficiently maintained/updated in the digital setting where the Big Data is mostly used and stored.

1.2. Overall framework of application of Big Data

Before moving onto the specificity of the Big Data implementation for the early-stage startups, it is vital to reach an agreement on the correct framing of the definition and to explain the specificity behind the term early-stage. In the last few years, such a term as a startup has become quite popular and frequently used, even more than discusses and taught. However, there not always is the mutual understanding of the essence of this concept.

To begin with, a startup is a company created to find a business model that is replicable and scalable. Reproducibility is an ability to sell the resulting solution many

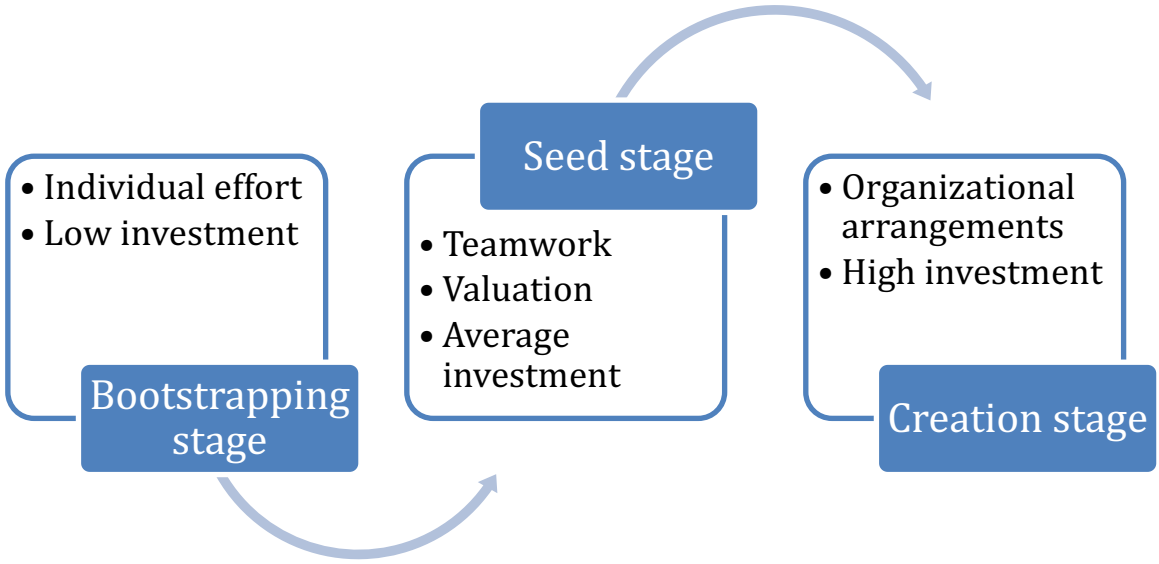
times over the initial starting point. Scalability is an ability to grow the project significantly within big milestones. In addition to this, an important distinguishing feature of startups is that they are technology-based, and most of them are based not on some commonly understandable solution, but a unique technological know-how.

Startups are founded in a way that they are designed to solve problems of their target audience that are possible to be solved exclusively through the usage of technological progress, in this matter of the 21st century. Today's high-tech projects are thought to serve a purpose of making it easier for users to do anything in their daily lives. It is not, however, the job of a startup to drive technological progress forward. They are businesses in the end of the day, so their main objective would always be the financial rewarding at the end of the day. Indeed, when talking about startups, it is unlikely to talk about a promising new company developing its revolutionary solution without a mention of the scale of its investment or the plans for the next series of investment.

The idea behind every startup is the unlimited potential of any businesses. Basically, they say that you would only need the entrepreneurial mindset in order to find a problem and to come up with a solution for it, and later on you will be able to raise money to scale your idea into a working business. It is therefore quite possible within one small team to create a unique software product that provides its users with innovative services, and which may prove to be extremely popular and in demand in the immediate future. Similarly, it is not a problem for a startup to work exclusively as a software-provider or a hardware-provider, nor is it difficult for them to create a combination of hardware along with the software package. Accordingly, only the creation of the product itself, same as its Minimal Viable Product or a Minimal Sellable Product can be a startup: a software product or a physical product put on together, as an example. This is also why retail in its pure form cannot be a startup because of the difficulties arising when scaling the production. This is about the correct timing for a startup to still carry the early-stage title.

The cycle of the startup creation has many steps and is commonly referred to the roadmap of each individual startup. The picture 1.2. of this thesis paper shows the

startup life cycle in the form of a graph. The similarity that there is, is that first of all, after the idea is created and validate, there has to be made a prototype. After the startup's early stage is passed, this prototype would need to be turned into a full-fledged product, transformed, and developed, and scaled up many times. Throughout this process, the startup would need to attract rounds of investments several times, and from the perspective of the processes inside the company, the startup team would grow, and the complexity of their internal operations product would increase. Eventually, the goal of creating each startup is for it to be sold to a large corporation or for then founders to take their shares to the stock exchange and continue operating as a separate company. After all, having passed the early stage of its development, and having gone through many various stages along the way, a startup would turn into a full-fledged business.



Picture 1.2. Early-Stage Startup Life Cycle

It is important to start this subchapter of the thesis paper by mentioning that it will be divided into 3 parts. This is the best approach considered when writing the paper because of the need to differentiate how the early-stage startups can use Big Data. There are 3 main frameworks defined:

- the startups whose product idea is fundamentally based on the work with Big Data itself;
- the startups that operate with Big Data in order to increase its efficiency;
- the startups whose product idea comes from not the Big Data implementation itself, but the related filed of Data Science.

Now, the thesis paper will explore the overall framework of application of Big Data for early-stage startups whose product idea is fundamentally based on the work with Big Data itself. The idea behind this type of startup product idea is to ground their core business model to the usage of Big Data itself. Many young companies all over the world are embracing data-driven business models specifically due to the growing potential of this technology not as a helpful additional resource, but as a unique self-efficient business model. The digital age calls for new challenges, as well as new solutions. That is why in spite of the fact that earlier on the chapters of this thesis paper talked about the implementation of Big Data internally and externally for international businesses, we couldn't miss out on the explanation of data-driven business models, which are due to obvious reasons of agility and flexibility are easier and more feasible to be implemented specifically for the startups, and specifically for their early stage of development. The emergence of new companies based on such new business models – the Big Data ones – opens enormous possibilities for transformation, including those very important changes in the benefits for the consumers. The result gets way more innovative, and thus – product-enhancing.

Moreover, the academic article Data-Driven Business Models: Powering Startups in the Digital Age suggests that such products created within the new Big Data-based business models become available at lower cost, and he businesses end up with better analytics for managing large social challenges at the micro and macro levels, including such drastic points of human society concerns as climate change and public health (Filippov, 2014). As well as that, they immense private-sector success for companies quick enough to seize the initiative and provide the new goods and services to customers (Filippov, 2014).

To give some examples of the early-stage startups implementing the Big Data-driven business models, first of all, we could look at businesses that conduct research on different markets and areas of the markets. This type of early-stage startups will be conducting a variety of research upon its own initiative or upon request with the main task of such business being to collect and analyze data in order to obtain comprehensive information on a wide range of business questions. Key areas would then include market trends, marketing campaign effectiveness, competition and consumer behavior, consumer and public attitudes toward brands, products, customer perceptions, biases, etc. Such early-stage startups then operate as B2C or B2B businesses and charge their customers for the research done within different areas, or they are also able to sell their professional ready-made research in different areas from business to science, medicine, etc. The Big Data collected by such early-stage startups needs to be structured after being collected, and analyzed, so that it can be attractive for other businesses or individual customers to purchase and get use of. For instance, we could support this paragraph by further elaborating on the solutions which use Big Data to better up their marketing and targeting on the areas which are predominately known as not the ones that are very much using such principals. An example would be Finn AI (www.finn.ai) which is a solution from the fintech area, with a specified specialty on Internet banking. Finn AI offers their users an Artificial Intelligence-based chatbot that helps banks and credit unions develop convenient self-service for customers. As you understand, even though it is a fintech solution, it operates differently by choosing a B2B business model when a bank or a credit union is buying the service of Finn AI in order to increase its customer satisfaction rate. Finn AI promises to automate almost 93% of requests which results in 80% decreased waiting time. Finn AI need Big Data in order to create such a unique selling point for the banks and credit unions and also utilizes AI to make the service even more transparent and efficient.

The second example could be an early-stage startup that is based on collaborative consumption using Big Data. Within this business model the most important point of the collaborative consumption would be data sharing and data variety, which both are the business trends of today. It is very common for young startup entrepreneurs to

simply create a platform that connects customers and those who can provide them with the needed services, by which the startup serves as the bridge in-between of a customer and a service-provider and is responsible for nothing more but establishing soothing experience for both the customer and the service-provider. Such startups would then be operating on a collaborative consumption model, which is indeed fundamentally based on Big Data technology. A variety of industries are already highly invested in such a collaborative consumption model startup basis, but there is no limit for the other services of this kind to be offered to target audience across the world.

The third example of the early-stage startups implementing the Big Data-driven business models is connected to the digital marketing services. Online or digital marketing, meaning the social media, online marketing channels, paid advertisement, is solely a ground base for the promising sales streams of the today's businesses. If an entrepreneur doesn't understand how digital marketing operates, he or she is missing out on a lot of business opportunities and potential clients. This is a perfect window of opportunities which becomes open for the young startups of the early stage, because the secret behind great online marketing performance is an effective use of Big Data for marketing purposes. That being said, solid knowledge of resources and technologies of the data analytics when it comes to Big Data, can successfully provide a young startup with a scalable business idea when he or she offers digital marketing services to bigger companies and individual users, this way executing the B2B or B2C business models. As an example, a startup Lemonade (lemonade.com) with the help of Big Data in the form of Artificial Intelligence, as well as by using the behavioral economics, is working in the insurtech field. It minimizes paperwork and executes the key business processes of the insurer by using Big Data to give instant support to the customers. They offer the help of an artificial intelligence bot who crafts the perfect insurance for the users, so there is no more manual work from the side of the employees involved. Big Data helps this startup to reverses the traditional insurance model, and make the life of their customers much easier, as well as that they pioneer in the insurtech solutions.

The fourth idea might be considered somewhat more challenging, but at the same time more rewarding as well because it involves prior knowledge in the field of textual work and textual analysis. This idea covers the textual data analysis or data interpretation services specifically for a reason that most of all such work is already atomized with the Big Data technology. Most Big Data there is now processed using algorithms, including the textual information, which can also be statistically interpreted, even though earlier on many scientists would only consider a qualitative approach to words, and very rarely qualitative. In reality, the ground base of every qualitative work would still be a quantitate repetitive set of actions. Text data analysis would then mean the extraction of numerical metrics from text data. These numerical metrics are then easily analyzed, drawing important conclusions and answers from them. Textual Big Data shouldn't only be thought as articles or google searches, but also emails, social media posts and comments, etc. Therefore, the field is likely to be underexplored, as there is always more and more data exponentially coming to us and not leaving the unlimited space of the internet. If a young startup is adept at textual analysis or data interpretation, he or she could create a company that provides these critical services that the businesses and corporations need, as well as individual users, this way executing the B2B or B2C business models.

Another, fifth idea is basing online commerce on Big Data and creating an early-stage startup with such a business model. There is a large amount of data required to create and run online platforms that facilitate online trading. Moreover, e-commerce is known to be the biggest player when it comes to global business. It is therefore apparent that online commerce business will only keep scaling and growing, with new platforms and services appearing along the way. E-commerce is characterized by the constant collection or receipt of large amounts of data, or Big Data. In this field, a young startup could choose the B2C business model with a strong focus on retail, so that the interpretation of Big Data would become that one key factor which will guarantee profits for his e-commerce early-stage startup. As an example, Tableau (tableau.com) decided to upgrade the Big Data question by creating a service which helps with data visualization, and now their product is the most proffered one in the arena of working

with data. The startup has had 10 years of experience, so it has passed its early stage some time before. However, Tableau has positioned itself as a leader, so its technology is still one of the most fastest growing, making Tableau one of the best examples of Big Data and Data Science companies in the world. The co-founded of Tableau Software, Inc., Mr. Christian Chabot, has been its Chairman and CEO since 2003, and it was him who has led the company to nine consecutive years of record sales and customer growth (Business Analytics, 2022). Teradata (teradata.com) taken as another example for this paragraph, is a startup that uses Big Data specifically in the field of establishing a company based on of Data Science. Now, it is a corporation, but it has grown from an early-stage startup back in 1976, when there was no even framework what a startup is. Teradata has evolved over the years with changing technologies and delivering the best solutions to their customers. The company now provides analytic data solutions, including integrated data warehousing, big data analytics and business applications (Business Analytics, 2022). As of 2014, this company registered a market cap of $7.7 billion.

Last, sixth example would be content creation using Big Data. We all live in the world where we are either creating or consuming content. While some studies show that it is rather a problem and not a blessing of the newest generations it is nevertheless crucial to take advantage of the given opportunities. The one of the simplest and easiest way to do so is to set up an early-stage startup with a focus on Big Data when it comes to creating content. This practice is surely available to almost everyone now, and more and more people successfully apply this concept when they create a business of their own, or when they decide to scale up and improve the content conversion rates. There are several dozen ways to find a your own place in the world of content and create a variety of interesting content pieces: it can be blogging, video channels, social media accounts, podcasts, scientific articles, and so on. For the Big Data implementation in this field, it is obvious that the main idea behind the created content is Big Data itself.

The startup’s focus will be Big Data and Big Data related topics. People are getting interested in this topic due to the reasons described earlier in this paper: not only the business owners need the findings which Big Data can provide, it is also a

matter of a person's own interest in the question of the whole Internet of Things world. The early-stage startup idea then would be to drive traffic to whatever platform it will use to ultimately gain a huge, loyal, and ever-growing stream of followers and potential consumers, and, therefore, future customers. Then the more common streams of revenue appear on the way – in the process, there will be a possibility to find advertisers who would like to advertise their products, possibly also advertising their Big Data related solutions, by which they ultimately give you even more traffic. The advertiser's goal is to gain access to a huge number of your followers. You can simply advertise that you offer ad space on your platform and earn money from it. It will take some time to grow the number of subscribers, so that actions need you to publish engaging quality content on the topic of Big Data on a regular basis. A very well-known company for the students Coursera (coursera.org) could be a perfect example of this idea of usage of Big Data. To explain the choice of Coursera, it supports the idea of startups making their business model based on the Data Science, which is a very well-known education platform. The company's strive to make learning easy, interesting, affordable and accessible to millions of learners all over the globe is absolutely clear (Business Analytics, 2022). However, not many people understand that not only does it sell the courses on Data Science, but it also utilizes Big Data to make the best offerings for their customers. The combination of the use of technology, excellent instruction and flexible schedule made Coursera the largest and most engaging learning platform. Coursera was founded by Andrew Ng with Daphne Koller.

Next, this thesis paper will talk about the overall framework of application of Big Data for early-stage startups that operate with Big Data in order to increase its efficiency. It's not just the big players like big corporates such as Google, Amazon or Facebook that can not only own the Big Data but also take use of it and benefit from its seemingly unimaginable potential. Small and medium sized businesses can also harness the power of their data, as well as publicly open data or buy data from the other providers. The application of Big Data shouldn't be restricted to a particular company size because it is indeed a very powerful instrument. In practice, however,

representatives of such small and medium sized businesses do not take advantage of Big Data. This is not the case for startups, though.

One could assume that it is interconnected to the overall framework of work that the startups follow that make their workflow more productive, as well as lets them be more flexible to implement the usage of such innovative solutions, as, for example, Big Data. It is not always clear how to optimally use such an asset as Big Data to increase the profitability of companies – many options could work well, such as creating competitive advantages, optimizing marketing and business processes. Relevant cases in the publicity are still rare, business media and literature are limited in conveying project specifics, and many actually working models and approaches are protected by Non-Disclosure Agreements. Startups, however, differ in a way that their methodology of work with Big Data, which is at the same time the key characteristic of all the digital industries and just startups, involves interacting with customers via Internet, and it is actively shifting to traditional offline consumer markets, where a significantly greater economic potential is concentrated.

In this regard, all modern companies are seriously concerned about the problem of optimal use of accumulated data. And here is where the flexibility and agility of startups comes to the forefront once again. Automated database processing tools allow the such modern and quick to adapt businesses to use Big Data for better statistical analysis, building more accurate predictive models and making greater data-driven decisions. This thesis paper will explore several approaches of how early-stage startups are able to use the Big Data concept for their benefit if not to base their product idea on the Big Data itself, but to use the techniques within the internal and external operations.

The first very important point would be to make better analysis of opinions and sentiments. Such an analysis of opinions and sentiments can be performed thanks to active communication with the users on the internet, most evidently, on social networks. There the users discuss, like, criticize products, as well as give their honest opinion about the brands and services. By conducting a quick analysis of opinions, the startup can quickly understand assess the current state of affairs of his business and

find out which products or companies the users like or dislike, and why. It is indeed very important to analyze the reasons for user sentiment, not just their behavior. For monitoring the state of things, it is important to pay attention to users' opinions, feelings and attitudes towards a product or brand. If the startup uses the right analytical tools, it can also identify the most influential people within their customer base, the, so-called, opinions leaders who influence the community on social media the most. Afterall, the further analysis of reviews will help the business to improve the product or better the service. Early-stage startups have to be extremely active on social media and on the internet in the browsers because it is their main source of feedback from the potential customers even before the first sales have actually been conducted. Not only is it a great chance to build a good brand awareness, it is also needed to understand your audience, after it has been segmented. There surely will be some changes in the moods and opinions of your users throughout the long journey of your startup product, so that is highly crucial to follow the changes on a big scale – and for the big scale, or the so-called helicopter view Big Data becomes incredibly handy and useful.

The second big point that the startup could take a use of Big Data is the policy of knowing their clients better. It is related to the previous paragraph in many ways, however, here it is possible to talk more about the possible predictions of the future customer's behavior. Customer experience, customer decisions and actions the client from the past can tell a lot about his or her future actions. A person's habits, such as shopping, going on vacation, demographics and so on, form a complete full picture of his lifestyle. This is what is in need for the young startup. When you as a business owner have a good command of your clients' profiles, you can formulate the most individualized offers for them. Big Data will help not only to understand how a customer actually feels about your product, but also to predict if they will decide to move away from you to a competitor.

The third point of upmost usage of Big Data when building or scaling a startup is better segmentation. Businesses always segment their customers, it's not a new concept. What's new is the idea of analyzing Big Data, so that startups can explore completely new segments. This is also an approach to group customers in a different

way, so that the businesses can more accurately understand which features of their products are most important to which specifically customers. This approach to using Big Data can significantly increase usage of a product or a service by offering customers personalized loyalty programs or cashback, knowing what they would prefer from the Big Data we have, or we receive. Another example would be using Big Data to find out the pricing range that your customers of different segments are willing to pay for the product. This would be a serious change in the startups' pricing policy, and a big helper, because it is the early-stage startups that experience the difficulties with the pricing in the beginning of their journey the most.

The fourth point is surely the option to make the startup's next best offer. By predicting what kind of purchase the startup's customer will make in the near future, it is possible to set up timely offers on a related product or service, and so this way it is another stream of revenue that increases sales. Of course, it is rather an art of figuring out the preference of the user and which service or product are the best option for them at the given moment, but Big Data is there to assist the potential concerns and give answers to the questions which arise in the young startup's mind.

The fifth point this thesis paper would mention is the multichannelness that Big Data offers to the users. This is likely to be the most critical factor because of the amount of data there is and the need to structure it in a coherent way. Of course, the options for the client to communicate with the product or service is indeed unlimited: a client will need a financial solution, and the first thing which comes to his mind is not to search for a fintech startups that offers the solution for his problem but to go to the nearest back, and he could reach them at a branch, by phone, online, responding to TV ads, responding to blogs and social media ads, and so on. So, the question that the early-stage startup has would be how to make sure that they get all the possible clients who are in need of their product. Multi-channel and Big Data analysis here give options to the startups to better understand what kind of communication this user is likely to search for solutions and through which channels he would expect the product to come from a business. After all, it could well-described as finding a common language with the customer, the one the business and the client both speak.

To sum it up, proper work with Big Data for early-stage startups helps to see the real sales funnel through all the channels and identify key players. Thanks to the qualitative analysis of Big Data in all the channels, the startups will be able to answer its questions of which channel is responsible for the purchase and what are the trends and patterns leading to a purchase.

Next thing to be opened and covered by this thesis paper is the overall framework of application of Big Data for early-stage startups whose product idea comes from Data Science. “The road to recovery is paved with data,” said Kate Smaje, senior partner, McKinsey & Company. Data is providing the fuel to power better and faster decisions. High-performing organizations are three times more likely than others to say their data and analytics initiatives have contributed at least 20 percent to EBIT (McKinsey & Company, 2020).

Data Science is a type of data analysis formed at the intersection of statistics and mathematics, on the one side, programming, on the other side, and knowledge in any business area, on the third side. As was said earlier, each of us, using the phone, produces a huge amount of data every second: location, preferences, photos and likes. Computer technologies for processing huge arrays of such data in real time make it possible to draw certain conclusions at the request of a particular business and monetize this process. Data Science allows the businesses to accurately and from the right angle predict the outcomes of today's events that can be very useful for every business, and especially an early-stage startup.

Data Science is in demand for the analysis of sales, markets, customer segmentation and research of their behavior. Most likely, in this case, the work of an analyst to a large extent comes down to data aggregation, building dashboards and setting up marketing campaigns. The last task is an optimization task, so that the business figures out within which models it should build its sales funnels and overall operations to predict the client's reaction to advertising, as well as all kinds of promotions and discounts rates. Risk management is actually a separate industry, and quantitative analytics is a tool for controlling and managing risks which is extremely useful for startups that are just starting along their journey. The most common interest

point here is credit risk management, and advanced methods for calculating regulatory and economic capital and provisions for expected losses from investment defaults.

Data Science, on the other hand, can be also a source for the early-stage startups whose product idea comes from Data Science itself. Here it is important to take a look at R&D, or Research and Development. As part of the work of R&D laboratories, there will be some fundamental research often carried out which would ultimately result in the development of new algorithms, either connected to Data Science or Big Data, or to other technologies like Artificial Intelligence or Machine Learning architectures. The specialists required in such areas are much more specialized, and they tend to dig deeper than classic data scientists.

This way Data Science can be a product to be developed of trainable agents in computer games, control of robotic equipment, unmanned aerial vehicles, autonomous driving. It is more likely for early-stage startups to sell their products or services based on Data Science within the B2B business model, because it is less likely for individual customers to carry out needed processes to invest finances into Data Science on their own.

As another interesting example, there is a way for such big industry players as technological giants (Google, Yandex), online trading platforms (Amazon, Alibaba), or social networks (Facebook, Instagram, WeChat) to often buy early-stage startups in order to turn them into their own internal structural units. And it is a rewarding experience for startups financially. A steady trend in recent years is associated with the transition of everything and everyone to cloud platforms. And while there is a debate if such big tech companies are monopolies, they keep getting bigger market shares by buying out contradicting to their vision startups. In this connection, entire ecosystems of partner services are then built based on platforms such as Azure, AWS, or Google Cloud. In particular, these services offer customized access to Machine Learning and Big Data capabilities.

1.3. The effective application of Big Data for early-stage startups

It is of crucial importance for human being to be able to quickly adapt to fast-changing environments and worldwide matters. As well as it is even more crucial for the businesses to be adaptable when it comes to problem-solving decisions within the fast pace of their work. Startups, for their biggest advantage, are there to always quickly adapt new approaches and transform their previous ways of company operations. Even before the global health crisis, 92% of business owners surveyed by McKinsey & Company believed that their business models were unsustainable at the pace of digitalization of the era they were operating in (McKinsey & Company, 2020).

This strong need to follow the trends of the digital age and propose new solutions to the old problems is very-well met with application of the modern-day technologies, one of which is Big Data. This subchapter will explore those benefits that application of Big Data brings to mostly, but not only, early-stage startups.

With a variety of data sources, such as, for example, CRM systems, advertising and social platforms, loyalty programs, IoT and databases, there are increasing demands on technology and skills to manage, analyze and extract value from Big Data. Once the integration phase is complete and the entire data architecture in a company is in sync, businesses often have a need to apply machine learning techniques to find new answers. This is how the field of Data Science emerged for professional analytics capable of leveraging existing data sources and generating new knowledge and insights, extracting hidden patterns and actionable insights. Professionals in this industry are required to have business experience, effective communication skills, the ability to accurately interpret results and use statistical methods, the knowledge of programming languages, software packages, libraries, and infrastructure. For this reason, there exist at least four typical approaches to structuring and modeling for calculations with using Big Data:

- LTV (lifetime value, lifetime customer value).
- Churn rate (probability of customer exit).

- ROPO (research online purchase offline, the impact of online channels on offline sales).
- MMO (marketing mix optimization, optimal media exposure).

With the proposed frameworks, the departments will be able to quantitively and qualitively measure the use of implementation of Big Data, as well as speed up the execution of their projects within their newly founded startups.

Visualization of the results of work with Big Data is also extremely important. Data scientists use BI (Business Intelligence) tools to help managers identify trends and patterns by presenting information in a way that is accessible to non-technical experts. Data visualization provides a graphical representation of information that satisfies the need to communicate Big Data analysis results to those outside of the analytics department. To achieve the best possible result, the data visualization specialist strives to achieve a fine balance between form and function. Visualization types for dashboards include charts, tables, graphs, maps, infographics and more. Thus, the best BI platforms visually display key business metrics and their dynamics, allowing stakeholders to use this information for visual Big Data analytics.

Most Big Data experts agree that the amount of data generated will continue to grow exponentially in the future, as we have already mentioned earlier. According to some estimates, the global data sphere will reach 175 zettabytes by 2025 (McKinsey & Company, 2020). The main factors contributing to this are the growing number of Internet users who do everything online and the proliferation of connected devices and embedded systems.

With public and enterprise cloud providers such as Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform transforming the way they store and process Big Data, experts also predict that the future of Big Data will remain in the cloud. And hybrid and multi-cloud environments are seen as the future of enterprise deployment of Big Data projects.

So-called fast data, which allows businesses to see and explore real-time processing of streams, is expected to become increasingly important. With streaming technologies enabling organizations to analyze such information in as little as one

millisecond, fast data will become a critical means of generating revenue for businesses. Examples of such projects include real-time user scoring for premium online stores, b2b services, real estate developers' websites, and so on, where losing a customer is extremely costly to the company. The introduction of emerging Machine Learning and Artificial Intelligence technologies into Big Data analytics tools is only expected to help develop this trend further.

Big Data arrays are so large that simple Excel sheet cannot handle working with them, and that is why special software is in need to be used to work with Big Data. Such software would be called horizontally scalable because they are able to distribute tasks among several computers simultaneously processing information. The more computers are being involved in the work, the higher is the productivity of the process.

Such software is based on MapReduce, a parallel computing model. MapReduce is a distributed computing model introduced by Google, used for parallel computing over very large, up to several petabytes, datasets in computer clusters (Dean, 2004). The model works like this:

- First, the data are filtered according to conditions set by the researcher, sorted and distributed among individual computers (nodes).
- Later, the nodes compute their blocks of data in parallel and pass the result of computation to the next iteration.

Therefore, MapReduce is not a specific program, but rather an algorithm that can be used to solve most big data processing problems.

Some examples of software that is based on MapReduce:

- Hadoop, a set of open-source programs for file storage, scheduling, and data collaboration. The system is designed so that if one node fails, the load is immediately redistributed to other nodes without interrupting computation (Apache Hadoop, 2022).
- Apache Spark, a set of libraries that allow users to perform calculations in memory and multiple access to the results of calculations. It is used for a wide range of tasks, from simple data processing and filtering to Machine Learning (Apache Spark, 2022).

Big Data professionals use both tools: Hadoop to build data infrastructure and Spark to process real-time streaming information.

The three main professions in Big Data are data engineers, data scientists, and data analysts. Data scientists specialize in Big Data analysis. They look for patterns, build models, and use them to predict future events. To master this profession, one would need an understanding of basic mathematical analysis and knowledge of programming languages such as Python or R, as well as the ability to work with SQL databases. The data analyst uses the same set of tools as the data scientist, but for different purposes. His or her tasks are to do descriptive analysis, interpret and present data in an easy-to-understand form. It processes the data and produces results, making analytical reports, statistics, and forecasts. Big Data is also handled by other specialists who should be on board of the startup to make it work efficiently and bring the best possible results. For some of the specialists, Big Data itself may not be the main field of work, but they still need to have a very good command of Big Data in order to make their work successful. Such professionals within a startup could be:

- Interface designers who analyze behavioral research data to create user interfaces.
- NLP engineers who develop software for chatbots and call center automation by analyzing natural language.
- Marketing analysts who research massive amounts of data to build marketing practicies and campaigns, and personalize advertising.
- Engineers and programmers in startups that process all the kinds of data.

Lastly in this chapter, the thesis paper will explore the difference in the culture of application of Big Data for early-stage startups and for established on the market corporations. There are a lot of individual users, as well as businesses and ventures that believe that Big Data and Data Science is meant only for big organizations having high-class infrastructure. It is generally believed that such a concern stems from a misconception of the data science as a whole. Data science is not created by machines and heavy equipment that can only be found on the big corporations, but rather it is statistics, analytics, programming, presentation, and professional employees who know

how to make the best use of data itself, and how to bring value to their organization. The size or headcount of the company therefore doesn't really matter, even though there is a common understanding that the amount of financial resources does play a big role in this question. However, leaving the finances aside, and assuming that in this regard the opportunities are absolutely similar if not the same, there still will be some differences in the application of Data Science for an early-stage startup and a key industry player as a big venture.

Startups are very dependent on building their product on time, meaning, when there is market need for such a product, so there's no room for extensive research which will be very time-consuming, and there is no space and time for reinventing the chosen solution, they have to be quick and flexible. In corporations, though, there always is a lot of time available to play out their assumptions before actually releasing the decision and implementing the new approach to the operations. Sometimes, however, such processes of making a decision can take many months. However, it is still believed that it's better to use proven models or even proven entire solutions in a strategic and well thought out way in terms of a given business problem (MasterBorn, 2022).

One of the key differences between implementing Big Data and Data Science in a startup project and a corporate project is the amount of risk taken (MasterBorn, 2022). As VentureBeat reports, about 87 percent of Machine Learning models never make it to production (VentureBeat, 2022). While large corporations can afford it and they have such risks already counted and included in the final planned costs of production, it can be a turning point for a small early-stage startup. This is also why investors ask for a thoughtful plan of actions from the startup on top of aa solid idea, quality data, and sometimes offer to hire a good data scientist to the team.

If mentioning a good data scientist, this concept can also vary between big companies and startups and working there looks different from an employee perspective as well (MasterBorn, 2022). Being a data scientist in a corporation, the employee is most likely to work in an big and multi-member team where each person is only responsible for particular assigned tasks, as it always happened in the big companies. In such an environment, it is very likely that any proposition from the

employees' side to change something or introduce an innovation that he or she proposes will have to be firstly many times discussed at meetings with managers and colleagues, and by that the window for new changes has already been closed, or the employee has lost the motivation to do this task, if not his work in general. Working in a small company or a startup, especially early-stage one, the person may be the only one responsible for the entire Data Science and Big Data field or will be a member of a very small and hard-working and highly invested team which would then result in much more room for maneuverability but also much more pressure and responsibility for the project (MasterBorn, 2022).

So, as described earlier, business development opportunities and scaling the operations of a company using Big Data and Data Science are not just available to huge corporations or Artificial Intelligence companies, but also to smaller startups and early-stage projects. However, properly designed solutions have to be supported with the right decision-making process from the side of the startup or corporate management.

CHAPTER 2.
PRACTICAL FRAMEWORK OF APPLICATION OF BIG DATA FOR EARLY-STAGE STARTUPS

2.1. Breaking down an existing early-stage startup within the Business Model Canvas framework

To start this chapter, we firstly need to go back to the topic of application of Big Data for early-stage startups and take a specific example of one startup which we will be looking at in order to create a great example of the application of this concept, as well as to be able to analyze this example and define the specific ways for the analysis to further elaborate on in the next chapter. We have to keep in mind that the chosen startup has to be an early stage one and has to also follow the main patterns of the ideas that an early-stage startup as a working unit is typically known for.

For this, let's firstly define the profile of the startup we will be working with. We're taking a startup FitSit (https://fitsit.io/) as an example for this thesis paper.

Picture 2.1. The logo of the early-stage startup FitSit

So, FitSit is an early-stage startup in the field of digital health innovations. The startup as a company has been created in January 2020, meaning that the team has been working on the project for a bit more than 2 years. That fits the description of a company that is currently on the early stage due to the small number of years that the team co-founding team has been working on the project. Secondly, the startup has not raised any investments, and has not overgone any investment rounds. They have not been funded by any other sources as well, and only had invested their own finances into the project, meaning that the team of the startup has been bootstrapping for the time of the startup production cycle. Thirdly, we could talk about the product stage that

makes this startup an example of an early-stage startup, and it is precisely the fact that this startup doesn't have a ready product yet and is currently on the stage of the early development of the project, or the MVP or a minimum viable product. To further describe this startup, we would have to take a look at the product that they are offering.

FitSit says that they create a web application that helps to create a habit of keeping a healthy posture and, therefore, fights back and neck pain. To give you some further context of this application, FitSit offers short posture practice sessions, when the user's upper body is monitored by a computer's webcam. The user actively sits, meaning that there is nothing else for the user to do but to actually sit with the correct posture, while the app will be giving immediate feedback to the user and will remind the user to keep up to a new habit during the workday. FitSit's team says that it helps users to have a correct posture while they are simply sitting at work, and this one habit brings huge benefits to overall health. FitSit is good to prevent back and neck pain, and relieve this kind of pain when needed, as well as to build an attractive healthy posture.

So, to briefly sum up the business model canvas, we will go over the main points of the canvas which will be summed up in the Table 2.1.

Problem	Back and neck pain connected to long-term sitting in front of the computer at work. Musculoskeletal disorders (MSDs) remain the most common work-related health problem in the European Union. Besides the effects on workers themselves, they lead to high costs for enterprises and society. Working with computers, in particular, outstrips the other risk factors. Roughly 3 out of every 5 workers in the EU report MSD complaints. The most common types of MSDs reported by workers are backache and muscular pains in the upper limbs. The most prevalent risk factors are standing, repetitive hand movements, and working with computers. During the last 5 years, for most physical risk factors prevalence is slightly decreasing, except for working with computers, laptops,

Problem	smartphones. Working with computers, laptops, smartphones has been a rapidly developing trend in Europe. According to the European Working Conditions Survey, the percentage of people working with computers for at least a quarter of their working day increased from 47 % in 2000 to 53 % in 2010 and 58 % in 2015. For sitting (60 % in 2015) it is not possible to determine a trend, but it is likely that this has increased as well (assuming that most of the work done on computers, laptops is done when sitting). The risks from working with computers, laptops, smartphones or from sitting for at least a quarter of the time predominantly occur for clerical support workers, managers, professionals, and technicians, and associate professionals. Regardless of the specific nature of the relationship between sitting, computer work, and MSDs, sedentary behavior at work is hazardous for health (cardiovascular pathologies, cancer, diabetes, etc.) and this occupational risk needs to be prevented, especially in a context in which sitting at the workplace is increasing.
Solution Solution	The algorithms within the Artificial Intelligence and Machine Learning, to be defined later, however, the ground base is a unique Rule Engine that is able to understand wrong and correct human postures. And this exactly unique Rule Engine is what makes the application of Big Data extremely important for this early-stage startup.
Unique Selling Points	• There are no recordings being made, and the webcam only watches the body of the user with his or her permission for a limited amount of time. • The parts of the body will be respectively assigned with virtual coordinates.

	• The given feedback will be exclusively real time, there are no recordings of people made and stored. • The application is EU GDPR policy compliant. • It is built upon the AWS databases (which are also HIPAA compliant).
Unique Selling Proposition	There is medical expertise behind the idea of FitSit. The concept of FitSit was developed by understanding what patients who have back, and neck pain experienced in real life. Having a degree in physiotherapy and working as a physiotherapist in the Czech Republic and abroad, one of the co-founders came up with an idea of how to help his patients, especially those sitting at their desks long hours every day. His idea is able to now help people from all over the world to fight back pain.

Table 2.1. Business Model Canvas for FitSit

Customer's Behavioral Incentive lies within two main grounds. Psychological explanation of FitSit is that changing the habit once and forever may be a very hard thing to do at once, and FitSit is able to provide support for the users during their journey, as well as to keep them highly motivated to keep the process going further on.

Scientific explanation is that FitSit uses the most effective ways to build a habit of correct posture. A medical study from Israel confirms that sustained improvement in reducing musculoskeletal risk among office workers using computers can only be attained with the photo-training method – aka using self-modeling webcam feedback (Applied ergonomics, 2010). Success for the team of FitSit means that the users are able to spend the whole day sitting in the office without pain. They achieve it by changing their positions several times throughout the day: switching from a relaxed passive position to active sitting with a straight back. FitSit suggests so because the

number of hours the person spends sitting doesn't really matter – the way he or she is sitting matters. Multiple studies concluded that sitting duration alone is not associated with low back pain, the relationship is only apparent when the user is sitting in sustained awkward postures like lordosed, kyphosed, slouched (Hartvigsen, 2000; Lis, 2007; Roffey, 2010). Quality, not quantity, matters.

Computer office workers with back pain have worse ergonomics indexes for chair workstations and worse physical risk than workers without pain (Rodrigues, 2017). So, the most basic step for the users would be setting up a proper ergonomic working environment, ensuring healthy behavior in various measures of your life, and excluding any possible pain risk factors (International Journal of Occupational Medicine and Environment Health, 2018).

Last but not least, FitSit's team always insists on the necessity to visit a physical therapist to learn about the user's personal needs and conditions. There is a lot of information on the web, but only a proper medical consultation has real value. One might doubt why he or she needs FitSit then, especially if the employer pays for the sessions with a physiotherapist. The effectiveness of ergonomic instruction over time is minimal if it is not combined with additional support, like constant reminders (Walsh and Schwartz, 1990), which is exactly the purpose FitSit serves. Education increases knowledge levels, but it is not translated into behavioral changes in maintaining a correct posture (Daltroy, 1993).

So, to cap it all, in the practical part of this thesis paper, we will be going over an early-stage startup that is working within the digital health innovation solution and that offers its customers a unique technology which can be scaled in different ways upon the technology which is now already existent and can be sold to the potential clients as an MVP version of the future product. The product itself, what is more, represents a health-tech solution and carries a highly innovative implementation. Especially with the upcoming trend to work remotely, a digital physiotherapist FitSit works with a customer's posture wherever they are to help them fight back and neck pain. FitSit helps them not only to create the habit, but also to maintain it afterwards through our notifications, scheduling options, medical recommendations, advanced

techniques and more. The users then also get more motivated, when seeing their daily correct sitting score graphs and percentages.

2.2. Practical case study of applying Big Data for FitSit's APIs strategy

The first point which couldn't be missed out on when understanding how Big Data could be helpful when talking about effective application of it for an early-stage startup FitSit is exercising the correct Machine Learning and Artificial Learning APIs strategy. API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other (Mulesoft, 2022). API is a description of the ways (a set of classes, procedures, functions, structures, or constants) that describes the ways that one computer program can interact with another program. It is often implemented by a separate program library or operating system service. It is used by programmers when writing all sorts of applications. Existing Machine Learning and Artificial Intelligence APIs can be divided into major groups.

The first are giant services that are designed to solve various tasks. Such services also provide data storage and preprocessing infrastructure, continuously running servers to process incoming data and form answers online, have user-friendly and intuitive interfaces to build models and assess their quality, to visualize data, model structure and obtained results. This group includes solutions from major companies: Amazon Machine Learning, Microsoft Azure Machine Learning and Microsoft Cognitive Services, Google Cloud Prediction API and Google Cloud Machine Learning, IBM Watson Cloud and AlchemyAPI, BigML and others (Mulesoft, 2022). The definite advantages of such services are access to high-performance computing machines, servers, and infrastructure for working with Big Data, high quality documentation and availability of ready-to-use libraries for working with APIs from various programming languages.

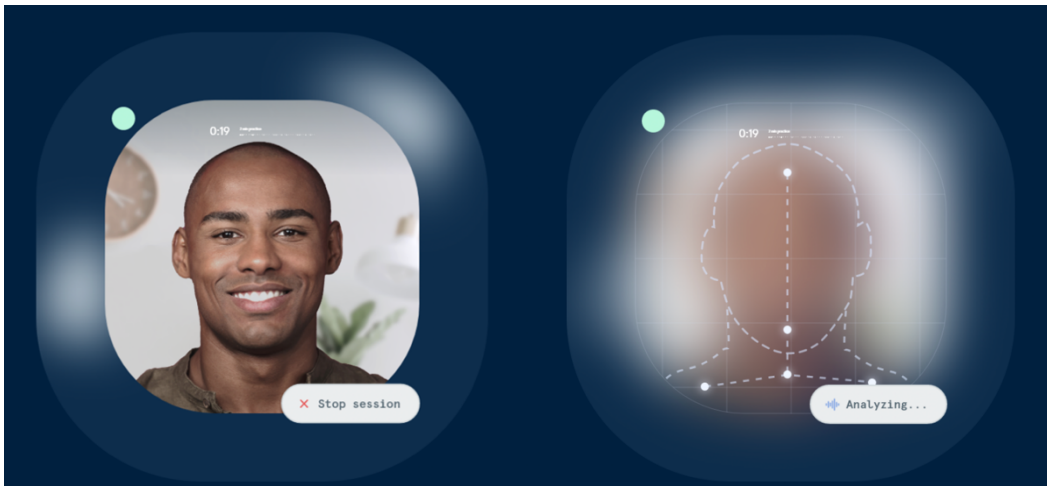
The second of the largest Machine Learning and Artificial Intelligence API packages allow the use of models and algorithms based on deep learning, mostly related to image and natural language processing, which are quite difficult to

implement on their own. Deep learning tasks are often quite complex, requiring advanced theoretical knowledge and significant computing power to learn and use the models. In this case, the use of APIs turns out to be particularly appropriate even for those who are familiar with Machine Learning. The range of capabilities of such API services is very wide. For example, the video and image analysis capabilities of Microsoft Cognitive Services include image description (who or what is depicted on it and what is happening), categorization, type identification (photo or picture), identification of adult content, identification of dominant colors, search in photos for specific people such as celebrities, logos and objects, near real-time video analysis, text recognition in images, thumbnail generation, motion detection and tracking. Face detection, sex, age, pose, glasses, beards, and other features for people in photos and videos, photo verification (comparing a photo from the database and a photo from the camera), grouping photos of the same people, fuzzy video stabilization, frame thumbnail creation (Mulesoft, 2022).

Examples of successful applications of Machine Learning and Artificial Intelligence APIs in business, and specifically startups, as well as product or app development are numerous. With a small monthly number of requests (a few thousand) most of the reviewed API services can be used for free, so startups have the opportunity to try Big Data providing services in practice and test when developing. Further on, the payment to those services will depend only on the actual capacity a startup uses (meaning, they pay only for what they use), and the fees are small for using the services, especially at the initial stages when the startup's product application has few users (Mulesoft, 2022). The services themselves chose such a business strategy that for their customers no large initial investment will be required, and the costs can pay off quickly.

Now, let's take a closer look at FitSit's example of the usage of Big Data when creating the product. By this, we will explore how an early-stage startup can use Big Data on practice in order to excel at their product and production operations. To start the explanation with we have to address the product of FitSit once again. It is a web application that helps the users to create a habit of a healthy posture. FitSit offers short practices when the upper body of a user is monitored by a computer's web camera (as

if they are on a video call). However, it is more important to look closely and in-depth for the actual solution behind the product that is presented by the team of FitSit to the broad panel of the audience.



Picture 2.2. Visually representative explanation of the FitSit application

Picture 2.2. depicts the way FitSit operates by showcasing a person having a session during which his face is monitored by a web camera. The process behind the technology is being done within two stages, both of which are driven by Artificial Intelligence, meaning that they use Big Data as the main agent within the whole duration of their processes. First of all, for the web camera to be able to track the posture and at the same time not to disobey to the general rules presented by the GDPR policies and not to obtain any personal information from the users, the camera doesn't actually take any recordings of the users. Moreover, it doesn't take screenshots or pictures when working with the picture of the users. What is actually does is that it intake images in real time. Meaning, that there is no recording made, and the algorithms build in this product are able to track the pictures at once.

The second stage is described as an automotive processing of the images which were received by the camera earlier on. There is no need for the technology to store data from the users because the computational analysis of the posture is being done at the same time as new visuals are being received from the new images of the users. What's more, this allows the product to not even use the images in order to assess the

upper body of the users. The idea behind this approach is that the camera sees not the body parts, but various coordinates of the points of the human body. So, we could say, that the qualitative image is transformed into quantitative data.

Moreover, this quantitative data is being piled up altogether. Every other second the application of FitSit receives a new data set with an upcoming every other image, and what is more, one data set also consists of a huge amount of data per one data set. Having stated all that and knowing that the users use the application for at least 20 minutes per day, each minute having 60 seconds in it, it leads us to 1.200 seconds of work of the application, or 1.200 different images received and transferred into a numeric view. Knowing that the most typical picture resolution would hold from 8.0 to 14.7 pixels, it makes us only sure that the number of coordinates is also big enough to consider these data sets as a very representative example of Big Data. Therefore, we are taking this example of application of Big Data as for the Artificial Intelligence and Machine Learning algorithms. Posture of the users are then tracked without the actual tracking as proven earlier with the help of a specifically designed Rule Engine. Feedback from this Rule Engine is what is for the least part provided to the user in time which is near to the real time. The analytics are then done actually in real time and with least latency. This all makes us believe that the data sets used are great examples of the Big Data implementation, and that they also make it possible for the startup not to store any data of the users. Basically, taking this scheme described earlier as a separate example of the application of Big Data, we are able to easily say that the concept is being highly innovative and the whole product idea is built upon the usage of Big Data.

To further deepen the discussion in this chapter, we will elaborate on the term Data-as-a-Service which is a data distribution model or data management strategy in which users do not handle the data collection, storage, integration, processing, and analysis processes themselves, but instead outsource these tasks to specialized cloud providers to get the result which they were aiming for initially. This approach ensures that users get the data they need for their corporate or personal purposes, but without the cost of infrastructure and additional staff. This is exactly what FitSit does – by

providing the product pf this web application they also offer a service of Data-as-a-Service.

The Data-as-a-service model operates on two basics: volume-based, where payment is based on data volume, or pay-per-call, where payment is taken for each API call to the data provider platform. The second plane is type-based, which is pre-structured by providers by type or attribute, such as geographic, financial, and historical data. According to the DaaS paradigm, data is provided to the user on demand, regardless of where the provider and consumer of the information are located geographically. And it can certainly provide a savings in data management, which for large organizations can add up to a very decent amount.

So, the Data-as-a-Service offered by FitSit is that the company takes their unique product such as a Rule Engine able to give real-time feedback for the posture and combines it with the service which is an example of Data-as-a-Service, such as of collecting Big Data sets which are the coordinates of the upper human body. Knowing that, we could therefore imply, that the service has gone one step further, which is not just the DaaS model, but the Big Data-as-a-Service, or BDaaS model. BDaaS is a combination of Big Data analytics technologies and cloud computing platforms that can help companies, alongside many other different key improvements for the product provided or for the operations within the company, to reduce the costs and time to deploy Big Data projects. In addition, BDaaS enables enterprises to manage Big Data in the cloud and provides easy access to data for all departments at all times, as well as for the users.

Of course, as with any cloud solution, the main risks of the DaaS paradigm are data security and privacy. Data leakage risks, ensuring information security, compliance requirements for sensitive data, such as the requirement to host personal data on servers that are located in the state - these are all important issues that a DaaS provider must address.

Now, let’s take a further look at the innovativeness of the idea behind the FitSit’s product that would make them as an early-stage startup an example of a high-tech solution providers which are pioneering in terms of the innovation with their solution.

The framework which the team of FitSit is now still developing, and which is already being sed in their first product the web application, will be able to teach cameras to read human body language from real-time video. The engine is a low-latency human motion capture system which would be compatible with web cameras, enabling the product to uncover human potential from pixels. This can become an individual product or a core technology for various solutions.

Let’s now sum up what would the implementation of Big Data mean for this innovation within FitSit. There is no doubt for the most common usage of Big Data being that Big Data analytics tools help extract valuable information that can be translated into business strategies and solutions that will become key for innovations. Changing strategy, developing new products and services to solve specific customer problems, improving marketing methods, customer service, or employee productivity are just the tip of the iceberg when it comes to examples of the implementation of Big Data. Big Data can help the companies to find new ways to innovate and extend their brand reach.

What FitSit does is that it uncovers human potential from pixels. Earlier on, we knew the examples of machines being able to uncover human potential from pixels by, for example, the technology called Face Recognition. By selecting a single Machine Learning algorithm, which is unconditionally based on the Big Data itself, and entering the accurate data, we’re able to obtain the result which lets us to the technology of Face Recognition.

This technology is being done in a way that, first of all, it is necessary to examine the image and find all the faces on it. Second, the machine needs to focus on each face and determine that, despite an unnatural face rotation or unimportant lighting, it is the same person. Third, it would need to identify unique characteristics of the face that one can use to distinguish this face from the faces of other people, such as eye size, facial elongation, etc. Finally, the machine must compare these unique facial characteristics with those of other people they know in order to identify the person's name from the database it has been given.

Computers are incapable of doing all the steps simultaneously, as human brain does, so the data scientists and programmers always teach the machines step by step of this process. In other words, they connect several Machine Learning algorithms in one chain. Let's look closely how FitSit does so step by step and how it uses Big Data for this process with the help of Table 2.2.

Step #	Process description	Final outcome
Step #1	This pipeline starts with Face Detection. Obviously, it is necessary to select all the faces and upper body parts which are located on a visual that a web camera captures before trying to recognize and differentiate them. For each individual pixel, it considers its immediate surroundings as well. The goal is to highlight how dark the current pixel is compared to pixels directly adjacent to it.	Having done so and knowing that FitSit's Rule engine is able to tell where the hand, the chin, or the shoulder is, the machine will then consider each individual pixel in the image in sequence.
Step #2 Step #2	Then the Rule Engine draws an arrow showing the direction in which the image becomes darker. To detect the body parts in this HOG image, all the machine has to do is to find the part of the image that most closely resembles the known HOG structure derived from the group of images of similar kind used for training. So, specifically for this the team of FitSit used the database of pictures of people sitting with correct and wrong postures in front of their computer desks. Then, they trained the machines on all of	The original image is converted into a HOG representation that shows the main characteristics of the image regardless of its brightness.

	these pictures provided to recognize the body parts correctly in all of the cases.	
Step #3	The next step involved was to make the same body part, like the face or the shoulder, viewed from different directions and angels accurately, and that is where the data scientists from FitSit used an algorithm called Anthropometric point estimation. One of the ways to do this for Face Recognition specifically is an approach proposed in 2014 by Wahid Kazemi and Josephine Sullivan. The basic idea is that 68 specific points, or so-called marks, present on each face are being highlighted, such as the protruding part of the chin, the outer edge of each eye, the inner edge of each eyebrow, etc.	The ML algorithm is set up to search for 68 specific points on each face. Similarly, the other body parts are also being assigned coordinates which are then turned into numeric representation of the visuals captured by the app and assessed with Rule Engine.
Step #4	In this step it comes to the very root of the implementation of Big Data in FitSit's product and the main problem of the very distinction of correct and wrong positions of the body parts and the deviations of the body parts for the normal, or physiotherapeutically speaking correct, position.	The Rule Engine of FitSit learned to extract some basic characteristics from each part of the body.
	In this part, the technology made it possible to take characteristics from an unknown part of the body and compare them to the characteristics of known ones. For example, it was able to measure each ear, determine the distance between the eyes, the length of the nose, or the distance between the chin and the	After repeating this step millions of times for millions of images of thousands of different people, the neural network can reliably create 128

Step #5	chest, between the right and left shoulder, etc. The algorithm of the FitSit's Rule Engine then looks at the characteristics that it was generating for each of the body parts for the normal circumstances. It slightly adjusts the neural network so that the probable small deviations are still accurate.	characteristics for each body part of each new person. Specialists in ML at FitSit call these 128 characteristics of each person that is a user of FitSit a set of personal traits.

Table 2.2. Step by step description of the FitSit algorithm's pipeline

The idea of reducing complex input data, such as an image, to a list of computer-generated numbers has proven to be extremely promising in Machine Learning. All the team of this early-stage startup had to do was to run many thousands of images of correct and incorrect postures of people captured by the webcam through their pre-trained network and obtain 128 features for each person to be able to then use them for the upper body parts recognition and evaluation. To cap it all, we proved what is innovative and scalable for an early-stage startup FitSit whose product is built upon the technology of Big Data.

CHAPTER 3.
ANALYTICAL FRAMEWORK OF APPLICATION OF BIG DATA FOR EARLY-STAGE STARTUPS

3.1. Application of Big Data for the customer segmentation

Modern businesses, especially ones that are leading business internationally and work within various geographies with customers from different countries and continents, need to know their customers by sight, be very good at understanding their needs, pains, and even sometimes be able to predict or guess the customers' wants. This is when the question of the implementation of Big Data to create customer portraits of the ideal cases come in handy and serve as a great perspective to the ability to segment people according to their interests.

Customer segmentation is an analysis of customer characteristics, behavior, supply, demand, and more. This all pile of information certainly makes the basis as a huge number of data which can be combined into big datasets and be referred to as Big Data. The companies tend to always be looking for customers and searching for some common characteristics among them, or, if possible, some divisions and classifications of the groups of customers, thereby they are in a way building different collections of their users and clients. Customer segmentation does not have a single standard, though. It is rather based on the business perspective, and highly dependent on the type of the company, if it is early or late stage, and the headcount of the departments. Combined with the actual or potential application scenario to the target object, product or service, different businesses within different industries can be fulfilling their own customer segmentation strategy.

This thesis paper, therefore, would focus on the customer segmentation strategy for the early-stage startups – to further explain what it would mean, we have to say that the customer base in the cases of either early-stage startups or recently founded companies normally don't have a solid understanding of their customers, and definitely not the final defined portray of their users. Even though they are rather focusing on the

customer base itself and not the categories of the clients with similar qualities, it is still early timing for such enterprises to be directly centralized around only one idea of the target audience and close themselves up for potential other groups of customers, or even international markets. In the case of early stage startups segmentation is highly needed to recognize users interested in a product and willing to buy it – such users are sometimes also called early adopters, meaning, they are the ones who are most likely spend their finances on the product or service, even if it hasn't yet reached the final stage of production, but they are willing to pay because they find some certain benefits in their product and the minimum stage of the product is already enough. In the case of FitSit, it is necessary to understand who would be most willing to buy the product in a way that it is offered now, in order not only to boost the production by motivating the founding team that they are on the right track, but also to actually get some cashflow in the daily operations of the company.

Early-stage startups also use customer segmentation based on Big Data to define priorities for the efficient use of marketing resources. FitSit, for example, hasn't raised any investments, so it would not make sense to hire a professional marketing agency to promote their product, and also even less sense to be trying out the marketing strategy themselves step by step, and waste the scare financial resources in order to play out their assumptions for the marketing. Instead, FitSit should rather use open Big Data, or that Big Data at their free reach on the internet in order to create that one right image for their product for specific categories of potential clients and choose the most optimal strategy for promotion of their application to those early adopters. Later, there comes the phase of the accurate assessment of the market with growth potential, in order to boost sales when the startup stops exercising the most basic principles of the early stage.

Therefore, this thesis paper will analyze the Big Data in order to find the target customers of the target value for the startup FitSit, according to the results of customer segmentation for early-stage startups. We will focus on the open resources in order to increase customer viscosity and achieve product significance.

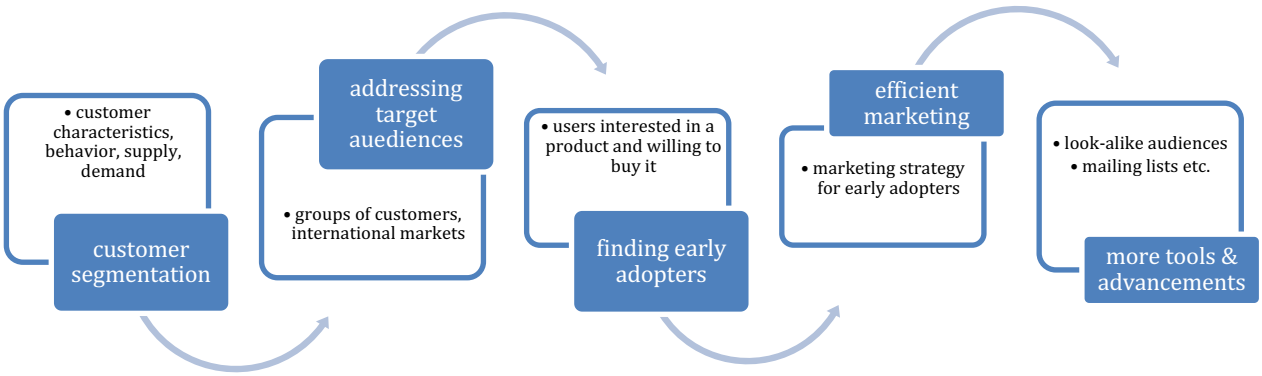
This step will then help to create the strategy of the targeted marketing for different target audiences, assembled according to the results of customer segmentation. Accurate marketing strategy would help to meet the needs of different clients and improve customer satisfaction. Even more than that, it can also serve its purpose by providing important reference value for customer management decisions, as well as design and product development based on customer distribution.

Another potential usage of the analysis of this thesis paper for the startup FitSit is creation of look-alike audiences. This tool would help to search for new audiences for the same product of the same stage of production based on the creation of the subscriber base that is as similar to existing customers as possible. It would allow FitSit to spend less money on advertising campaigns, increase their effectiveness and attract customers that are interested in the product or service most.

Next, FitSit would be able to use this information for their targeting of the mailing lists. This is a tool that allows startups and enterprises to filter user groups by geolocation, gender, age, and interests and send them advertising messages. Therefore, when creating promotional materials FitSit will be able to be more precise, and later send specific messages to very well-defined audiences. Specific information will be received only by those customers who are most interested in it.

Lastly, when leading the startup internationally, as FitSit is planning to do, they would be able to create heat maps. This tool would then allow FitSit to analyze the target audiences presented by this thesis paper based on the location, and to highlight the most interested in company's services or products user specifically in a certain territory. Heatmap itself is just a nice tool of the visualization of the Big Data analysis results, in which statistical and numerical values are being displayed using color.

To cap it all, the Picture 3.1. would add up to this chapter by explaining how the implementation of an accurate customer segmentation strategy should be done.



Picture 3.1.1. The scheme for implementation of an accurate customer segmentation strategy

As this thesis paper focuses on the implementation of the Big Data framework, we will also take close looks only at the big datasets when finding the target customers for the startup FitSit and leading the customer segmentation process. To do so, this paper uses the 2010 National Health Interview Survey (NHIS). The core and additional occupational health questions from the 2010 National Health Interview Survey provided the data sufficient to examine the main potential audience and formulate the probable customer segmentation strategy.

However, it is highly important to note that this survey NHIS is a yearly cross-sectional survey of the civilian and non-institutionalized population in the United States exclusively, so we're simultaneously setting the demographics for this analysis within the United States. One could assume that the probability of the result being very similar if not the same for the countries in Europe is very high. However, it is important to note the geographical locations can highly influence the behavior of the inhabitants, so FitSit would have to keep in mind that this analysis should only be used to elaborate

further on the demographics of the customer segmentation in the United States. The NHIS core questionnaire is the same every year for the U.S., but the supplementary questions change every year, collecting additional data on specific health concerns. This thesis paper will also use the second source of the Big Data for the analysis, and it being the PubMed. PubMed is an English-language text database of medical and biological publications created by the US National Center for Biotechnology Information (NCBI) based on the biotechnology section of the US National Library of Medicine (NLM). This information altogether will be sufficient for our research and analysis presented by this chapter.

In order to get the most out of the Big Data concept implementation for this research, this thesis paper will focus on the studies that was carried out to see the frequency of musculoskeletal problems in frequent computer and internet users in order to collect big datasets of information on the users. This will help us to define the demographics and personal characteristics of the users, by which, simultaneously, we will also understand the factors which rule in the customers' decision-making processes. Data from such research can be therefore taken into consideration for understanding of the prevalence of low back pain, as well as related workplace psychosocial risk factors. This all contributes to mapping the correct profile of the potential user of the FitSit's application, and even more importantly – it provides enough base to define the startup's early adopters.

The data that this paragraph will be going over is presented in the attachments as Table 1. According to the key findings, it was observed that out of 150 participants 47% experienced the symptoms of musculoskeletal problems. Which proves that the assumptions that the startup had initially about the pain points of their customers was correct.

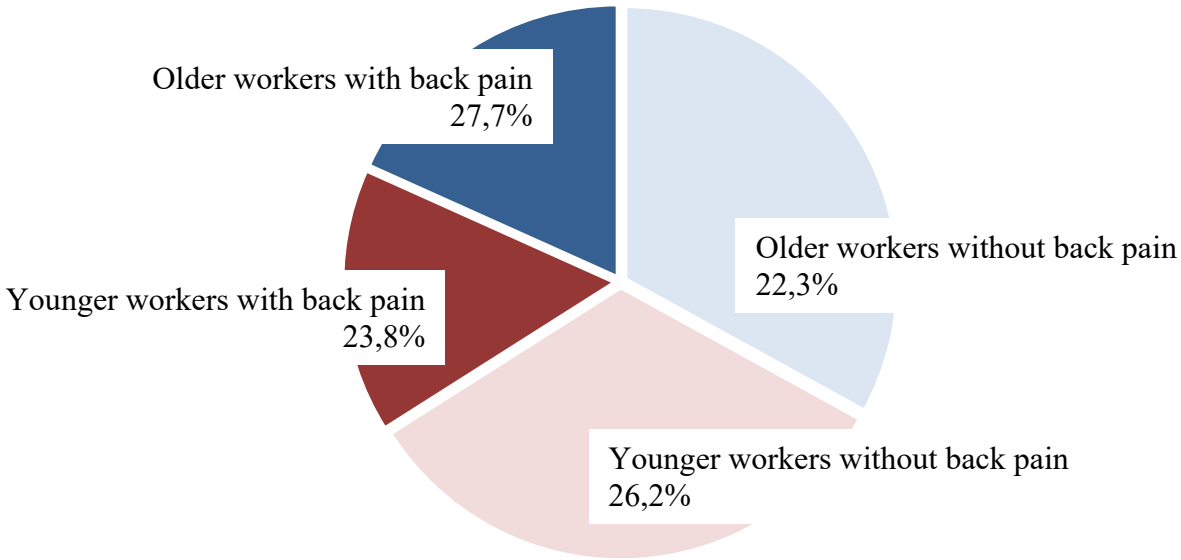
The prevalence of low back pain was 25.7% for all workers, 24.5% for males, 27.1% for females, 23.8% for younger workers and 27.7% for older workers (Yang, Haldeman, 2016). Table 1 also proposes the numbers for the prevalence of low back pain within the gender and age-group specificity, with 22.5% for males in the

younger age group and 28.8% for females in the older age group (Yang, Haldeman, 2016). Partially, this data is depicted in the Table 3.1 below.

All workers		Male Workers		Female Workers		Younger Workers (18-40)		Older Workers (41-64)	
% in Pop	% back pain	% in Pop	% back pain	% in Pop	% back pain	% in Pop	% back pain	% in Pop	% back pain
	25,7	53,0	24,5	47,0	27,13	50,0	23,8	50,0	27,7

Table 3.1. Key findings about the prevalence of low back pain

While demographic and socioeconomic characteristics of workers in the US are presented in Table 1, it also shows the relationships between low back pain and demographic and socioeconomic factors analyzed in the logistic analyses (Yang, Haldeman, 2016). The findings display the information that compared to the 18–25 age group, workers in 26–40, 41–55, and 60–64, had an increased risk for low back pain, controlling for other risk factors (Yang, Haldeman, 2016). The main trend can be seen in the Picture 3.1.2. which depicts a pie chart showing the prevalence of low back pain by older and younger age groups of workers.



Picture 3.1.2. The pie chart showing the prevalence of low back pain by older and younger age groups of workers

What is more, female workers also had a limited increased risk for low back pain, compared with male workers (Yang, Haldeman, 2016). Another conclusion which can be made and that would serve as a very good incentive for the choice of a profile of an early adopter for our customer segmentation is what Model A shows. Indeed, it was found that workers who had master or above level education have a lower risk for low back pain, compared with the workers with high school education.

Switching to Table 2 and revealing the findings from this spart of the analysis, this table shows us associations between the emerging psychosocial/organizational factors and low back pain. Interestingly enough, workers who reported that their work-family imbalance was being exposed and challenged, or that were themselves exposed to a hostile work environment or job insecurity had increased prevalence for low back pain, compared with those were not exposed to these risk factors (Yang, Haldeman, 2016). Female workers who were exposed to hostile work environment had the highest prevalence of low back pain (37.9%), compared with female workers exposed to other work-related psychosocial factors (Yang, Haldeman, 2016). If the gender question has always been an important topic to be discussed within many discourses, our analysis also revealed that some patterns observed are rather similar was for male workers as well. As an example, younger workers who reported job insecurity had the highest prevalence of low back pain (36.2%) compared with workers in the same age group who were exposed to other work-related psychosocial factors (Yang, Haldeman, 2016). Male workers (21.5%) and younger workers (22.5%) who worked alternative shifts had the lowest prevalence of low back pain (Yang, Haldeman, 2016).

Next, this thesis paper will go over the Table 3 in the attachments and it will be this part that will present the logistic regression analyses of the associations between low back pain and the psychosocial and work organizational factors, as was done in the 2010 National Health Interview Survey (NHIS). The findings show that while controlling for demographic characteristics, socioeconomic status, other health and health behavior-related factors, and other work-related variables, all workers who experienced work-family imbalance, were exposed hostile work, or had job were more likely to have low back pain (Yang, Haldeman, 2016). The sex and age stratified

logistic analyses for the psychosocial and work organizational factors indicated similar patterns (Yang, Haldeman, 2016). However, the one the major findings also show that the associations between low back pain and those of the work organization factors were not the same in different demographic groups of workers (Yang, Haldeman, 2016).

The analysis of the findings that this thesis paper has is mostly centered around the idea that musculoskeletal symptoms are associated with prolonged use of computer, which in many cases correlated with prolonged internet use, and that this problem is often left unreported and unrelated by those who suffer. Moreover, it will be various groups of country's population that are influenced by this problem, and we could argue that all the people, no matter the race, nationality, gender, or other demographic considerations are potential under big exposure to the problem. The key findings themselves might be coming from several different ideas. First, the responsibility could lie within the lack of awareness among students starting their career and office workers with long track of computer work about poor ergonomics, as well as their inability to report the problem to the right authorities. Secondly, it is that the time which the employees sit in front of their computers sustaining one type of the posture actually matters more than their incorrect posture. Those who had the symptoms of musculoskeletal problems in were the ones who used computers for more than 6 hours without often movements.

3.2. Formulating and analyzing the profiles of the most relevant target groups for the startup FitSit

So, based on this information this thesis paper is giving a customer segmentation of those from the target audience of FitSit application to better understand the users in general, as well as to make a detailed customer profile of an early adopter. Picture 3.2 summarizes the customer groups in one schematic overview, while the detailed explanation is given later in the chapter.

Picture 3.2. Proposed customer segmentation strategy for FitSit

Customer Group #1 will be based on the predications that the FitSit application should be used as a medical product, or an e-health application, as it is. Therefore, the first group in our customer segmentation would be people who are currently dealing with back and neck pain themselves. The age categorization on this case is not highly important because the main feature which differentiates those people is their need to solve the problem of inefficient treatment or poor results of the treatment. It is confirmed that 70 to 80% of people will suffer from work-related back disorders at some point in their life by the European Agency for Safety and Health at work.

This customer group could be reached out to through cooperation with physiotherapists, orthopedists, medical and rehabilitation centers that have a solid base of customers of their services. It is very likely that the business model to reach this category of customers would be B2B2C – in this case FitSit will find not the customers themselves but the cooperation with medical practitioners in the field and try to sell their product through them. The main distinguishing characteristic of the target group in this audience is that they have to be experiencing back or neck pain problems on a regular basis, as well as that they have to reach out for the medical services and seek a solution to fight back pain.

Their persistence of already being in search of the solution, or the market's inability to meet the needs of this customer group, would give FitSit a great chance of being introduced into the customer's routine as an innovative solution. By focusing on the patients with diagnosis like back pain, neck pain, scoliosis, tinnitus, migraine etc. FitSit will be able to reach a very solid target audience with keen interest in their new treatment offering. Moreover, this doesn't have to be the local market because the demographics of people with back pain go beyond one country. FitSit application can become if not the substitute product for the initial treatment with the physiotherapist, then a part of the whole treatment, so that the startup doesn't take the customers from the working medical practitioners. Lastly, this customer segment also opens a new way to sell the product, and the receiving end to be the insurance companies. Insurers are very likely to include FitSit into their insurance plan because it would prevent people with higher chances for these diseases from putting themselves in danger of back pain, and, on the contradictory, would treat this issue before it arises.

Now, this thesis paper is on to the Customer Group #2. This group of users will be represented by the office workers and employees of big companies with strong company culture. The prerequisite for the users to be involved into this customer segment is that they have to be working for an employer who highly values the health of the employees and is already implementing various solutions for their workers. As an example, some companies, like SAP in all their offices, have a physiotherapist on board whom the employee can contact and get professional medical advice on how to

maintain a healthy posture or what to do to get rid of acute pain. Another example and the fact that proves that companies invest in their employee's well-being is that they buy ergonomic chairs, set up standing desks, and, for example, build gym stations inside their office buildings. It is being done so since, as for the primary market segment research that this thesis paper has conducted, in 2020 59% of Germans spend at least a half of their workday at the offices (IBA Studie, 2020). Working towards helping the office workers who are employed in SMEs and corporates, including insurance companies as well, the estimation for the Total Addressable Market in Germany – the country that is known to be very proficient when it comes to setting up the culture of ergonomics and treat its citizens with high social responsibility – to roughly be 20 million people. However, as the first step it makes more sense to focus exclusively on those companies that have already implemented some well-being programs for their employees within their monthly agenda.

Therefore, through different sales channels FitSit should be able to attract attention of the SMEs, corporations, insurance, and healthcare companies to work with bettering the health of their employees. The strongest relations can be established by working with the HR departments of those companies, as it is precisely the HR managers who are thought to create a solid base of benefits to satisfy their company's employees. As the first step to establish such meaningful cooperation with the companies and thus sell the FitSit application as a B2B service, the startup has to agree on a testing period with the companies based on the Letter of Intent conditions, and only afterwards to sign the contract adjusted to every other business case accordingly. This way, the companies would appreciate FitSit being the early-stage startup and willing to be flexible, however, they will be obligated to use the product nevertheless. It is known for the big companies to take very long time when doing budgeting for an upcoming year or signing a new contract, and the LoI cooperation mitigates those risks for an early-stage startup.

The next in this chapter will be the definition of the Customer Group #3. For this we will take into consideration the main assumption that it is not only important for the customer to have a strong interest in the solving of the product, but they also have to

have potential to use the application. What we mean by that is the fact that offering an application which is rather innovative in the field of digital health would be much easier to a young working person who is very used to working with computers than to a senior who has been having back pain for a very long period of time. The interest coming from young people is thought to be more validated since they are much more involved into different subscriptions on the internet in the field of fitness and healthy lifestyle.

The last mention is also a great selling point – the topic of the healthy lifestyle is a very popular nowadays, so it would not cause any misconceptions for the customers who are already using many healthy lifestyle services, and in a way are persuaded that there should not be any immediate result, but rather a very long and continuous journey to keep.

What is more, it is now known that 21% of Germans switched to the home office in 2020 without proper ergonomics, and the trend keeps increasing more and more, as companies adapt to the remote work culture on a regular basis (t3n.de Digital Pioneers, 2020). Another explanation as to why the profile of such a potential user of FitSit application is more likely to become an early adopter of the application is that these people are more likely to face the real root of the problem. By working as freelancers or being employed for companies but keeping some manner of remote work, they are facing sitting for 8-10 hours in front of their laptops or personal computers without proper ergonomics. If previously there was no need for them to search for a solution to the problem of back pain, now they see the scale of this problem and that it is a real health-threatening disease.

This all depicts the need for FitSit to establish a B2C business model in which the customers will directly buy the application on the subscription basis. Using promotional channels on social media and various digital marketing tools, FitSit should be ablet to reach out to the potential individual users and offer them some differences in plans of the paying subscription, so that there is room for choice and less commitment for the beginning of the customer journey. FitSit, therefore, can have a selling point for individual customers that is it easy to purchase, easy to use, and easy to see the result product.

As a conclusion for this chapter, we have to say that while there are no paying customers on board of FitSit yet, the Customer Group #3 is thought to be the easiest one to test and their customer base should be found as the most approachable one, even though it will be harder to attain users later on, and the user acquisition rate for this segment will be lower. However, if we're in a situation of the early-stage startup, then our primary goal is to say if there is any product-market fit for the product in general which is a question that could be answered by directly and indirectly talking to the customers from the segment of the Customer Group #3.

CONCLUSIONS

All in all, this Bachelor's thesis paper has now theorized, analyzed and provided the practical ground for its topic of the Big Data in International Business: Application of Big Data for Early-Stage Startups. One might argue that this topic is yet underexplored within many discourses, however, we strongly believe that there is always room for discussion when it comes to such life-changing digital assets and modern-day tools as Big Data.

This thesis paper, therefore, gives more detailed overview to the question of importance of application of Big Data for the entrepreneurial purpose, specifically in order to lead International Businesses, and even more specifically – for the early-stage startups. Moreover, there have been made some key findings which answer the research questions posed in the introductory part of this paper, as well as some insightful findings for the specific example of the early-stage startup FitSit, the team of which has agreed to collaborate further and use the information and implement the results of the research in the development of their strategy. Now, the paper will present the findings which were made:

1. Big Data may not be the primary source that the business owners are thinking about when leading their business, and it may not be directly related to basic operational tasks that are usually in place to meet the regular business metrics. However, Big Data is a very powerful source of, first and foremost, information for businesses. So, especially when leading a company on the international arena where each country has its businesses operating differently and, thus, differentiates in its customer demographics, preferences, and decision-making journey, business owners are ought to be using Big Data. Big Data as such, as well as the range of modern-day resources, principles, and technologies, are giving opportunity to optimize the projects and operation to their maximum potential.

2. Big Data addresses many problems that the businesses are facing on their way, including such as analysis of the current state of affairs inside the company and the need to optimize their business processes, automating the routine in terms of

operations, correct estimations and feasible predictions for sales, customer-product interaction, marketing campaign execution rates, and price ranges. Moreover, Big Data helps international businesses solve the need of building accurate models through computerized digital models. Lastly, there is potential for Behavior Analysis for better targeting, and large-scale optimization through online optimization algorithms.

3. The most basic, and, therefore, most visible changes that Big Data may bring for the companies are an ability to make better analysis of opinions and sentiments, the policy of knowing their clients better, more clear and precise customer segmentation, the predications of the steps to make the startup's next best offer, and the theory of multichannelness as a whole. Big Data for early-stage startups is basically a technology that is giving the vision for the real sales funnels. Thanks to Big Data, the early-stage startups are able to answer their most important question of their entrepreneurial journey: which channel is in reality responsible for the purchase and what are the trends and patterns of the customers leading to a purchase.

4. The Big Data analytics have also introduced such emerging technologies as Machine Learning and Artificial Intelligence. They both individually, and especially in cooperation, offer new business development opportunities, which are open not only to the well-established players on the market but also newcomers, such as early-stage startups. However, one of the key differences between implementing Big Data and Machine Learning or Artificial Intelligence in a startup project and a corporate project is the amount of risk taken.

5. An example of usage of the Machine Learning or Artificial Intelligence fields was being described in relations to the early-stage startup FitSit. Its infrastructure is built on the work with Big Data and its availability within the ready-to-use libraries for working with APIs from various programming languages. FitSit's product is an innovative solution by itself with an even more innovative programming solution inside, as it the framework which the team has been developing is able to teach cameras to read human body language from real-time video. Its Rule Engine is a low-latency, human motion capture system compatible with web cameras, enabling the team to uncover human potential from pixels.

6. As for the analysis and the predictions for the early-stage startup which were made by using Big Data, we have to say that while there are no paying customers on board of FitSit yet, there should be done proper customer segmentation. We propose to do it in 3 groups, with the Customer Group #3 being the easiest one to test and approach. As an early-stage startup, FitSit has tested the idea of its product, and it now has to test the assumptions of its potential customers, and to do so, the team needs to get in touch with the group of their early adopters. This way finding the real product-market fit for the product we be achieved even without any direct sales, but by indirectly communicating with their actual customers from the segment of the Customer Group #3.

RESOURCES

1. Weigend, A. (2017). Data for the people: How to make our post-privacy economy work for you. New York: Basic Books. ISBN 9780465044696.

2. Mayer-Schönberger V. & Cukier K. (2014). Big Data. A Revolution That Will Transform How We Live, Work, and Think. Brno: Computer Press. ISBN 978-80-251-4119-9.

3. Rometty G. (2013). IBM's CEO on data, the death of segmentation and the 18-month deadline | Marketing Mag. Marketing Mag | Marketing Mag [online]. Copyright © 2022 Niche Media. Available from: https://rogerluethy.wordpress.com/2013/02/20/ibms-ceo-on-data-the-death-of-segmentation-and-the-18-month-deadline/

4. Preimesberger, C. (2011). Hadoop, Yahoo, 'Big Data' Brighten BI Future. EWeek.

5. Atal, M., & Mike, M. (2018). Big Data for Managers: Creating Value (1st ed.). Routledge. https://doi.org/10.4324/9780429489679

6. Filippov, S. (2014). Data-Driven Business Models: Powering Startups in the Digital Age. Available from: https://www.researchgate.net/publication/270508502_Data-Driven_Business_Models_Powering_Startups_in_the_Digital_Age

7. Dean J., & Ghemawat S. (2004). MapReduce: Simplified Data Processing on Large Clusters: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA

8. Apache Software Foundation, Licensed. (2018). Apache Hadoop [online]. Available from: https://spark.apache.org/

9. What is an API? (Application Programming Interface). MuleSoft. (2022). Integration Platform for Connecting SaaS and Enterprise Applications [online]. Available from: https://www.mulesoft.com/resources/api/what-is-an-api

10. 6 Astonishing US-based Data Science Startups Working for a Better Tomorrow. (2022). MasterBorn: Your React and Node.js Trusted Partners. MasterBorn: Your

React and Node.js Trusted Partners [online]. Available from: https://www.masterborn.com/blog/US-based-data-science-startups

11. Top Entrepreneurs in Big Data and Data Science / Business Analytics. Analytics Vidhya - Learn Machine learning, artificial intelligence, business analytics, data science, big data, data visualizations tools and techniques. (2022). [online]. Available from: https://www.analyticsvidhya.com/blog/2015/08/top-entrepreneurs-big-data-analytics-data-science/

12. Machine Learning is Fun! Part 2. Using Machine Learning to generate… | by Adam Geitgey. (2022).| Medium. Medium – Where good ideas find you. [online]. Available from: https://medium.com/@ageitgey/machine-learning-is-fun-part-2-a26a10b68df3

13. Yang H., Haldeman S., Lu M.L., & Baker D. (2016). Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey. J Manipulative Physiol Ther. doi: 10.1016/j.jmpt.2016.07.004. Epub 2016 Aug 25. PMID: 27568831; PMCID: PMC5530370.

14. How open API economy accelerates the growth of big data and analytics - KDnuggets. Machine Learning, Data Science, Big Data, Analytics, AI - KDnuggets (2022). [online]. Available from: https://www.kdnuggets.com/2016/06/open-api-economy-growth-big-data-analytics.html

15. Leem J. G. (2016). Big data and pain. The Korean journal of pain, 29(4), 215–216. https://doi.org/10.3344/kjp.2016.29.4.215

16. Borhany, T., Shahid, E., Siddique, W. A., & Ali, H. (2018). Musculoskeletal problems in frequent computer and internet users. Journal of family medicine and primary care, 7(2), 337–339. https://doi.org/10.4103/jfmpc.jfmpc_326_17

17. Yang H, Haldeman S, Lu ML, Baker D. Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey. J Manipulative Physiol Ther. 2016 Sep;39(7):459-472. doi:

10.1016/j.jmpt.2016.07.004. Epub 2016 Aug 25. PMID: 27568831; PMCID: PMC5530370.

18. Hong S, Shin D. Relationship between pain intensity, disability, exercise time and computer usage time and depression in office workers with non-specific chronic low back pain. Med Hypotheses. 2020 Apr;137:109562. doi: 10.1016/j.mehy.2020.109562. Epub 2020 Jan 9. PMID: 31945657.

19. Malińska M., Bugajska J., & Bartuzi P. Occupational and non-occupational risk factors for neck and lower back pain among computer workers: a cross-sectional study. Int J Occup Saf Ergon. 2021 Dec;27(4):1108-1115. doi: 10.1080/10803548.2021.1899650. Epub 2021 May 13. PMID: 33704014.

20. FitSit - Your digital physiotherapist. (2022). [online]. Available from: https://fitsit.io/

21. National Academies of Sciences, Engineering, and Medicine. (2013). Frontiers in Massive Data Analysis. Washington, DC: The National Academies Press. https://doi.org/10.17226/18374.

22. Apache Hadoop. (2006). Apache Hadoop [online]. Available from: https://hadoop.apache.org/

23. Big Data Intro - GlobalPulse - GAbriefing2011 - YouTube. (2021). YouTube [online]. Google LLC. Available from: https://www.youtube.com/watch?v=lbmsDH8RJA4

24. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P. & Stillwell, D., (2012) Personality and Patterns of Facebook Usage, Proceedings of the 4th Annual ACM Conference on Web Sciences, Evanston, IL, June 22–24, 2012 (New York: Association for Computing Machinery, 2012), pp. 24–32, Available from: http://dl.acm.org/citation.cfm?id=2380722

25. Big Data, for better or worse: 90% of world's data generated over last two years – ScienceDaily. (2022). [online]. ScienceDaily: Your source for the latest research news [online]. Available from: https://www.sciencedaily.com/releases/2013/05/130522085217.htm

26. Data Never Sleeps Infographics. (2022). [online]. BI Leverage, At Cloud Scale, In Record Time | Domo. Available from: https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf

27. Facebook by the Numbers. (2020). Stats, Demographics & Fun Facts. Omnicore: Medical & Healthcare Digital Marketing Agency [online]. Available from: https://www.omnicoreagency.com/facebook statistics/

28. Twitter Usage Statistics - Internet Live Stats. (2020). Internet Live Stats - Internet Usage & Social Media Statistics [online]. InternetLiveStats.com. Available from: https://www.internetlivestats.com/twitter-statistics/

29. Wild and Interesting Facebook Statistics and Facts (2021). Kinsta - Managed WordPress Hosting for All, Large or Small [online]. Kinsta Inc. All rights reserved. Available from: https://kinsta.com/blog/facebook-statistics/

30. Patel, P. (2018). Role of big data in us presidential election. Available from: https://www.researchgate.net/publication/325333423_role_of_big_data_in_us_presidential_election

31. Google Search Statistics - Internet Live Stats. (2022). Internet Live Stats - Internet Usage & Social Media Statistics [online]. InternetLiveStats.com. Available from: https://www.internetlivestats.com/google-search-statistics/

32. Big Data: How do your data grow? (2022). ResearchGate | Find and share research [online]. Available from: https://www.researchgate.net/publication/23235946_Big_Data_How_do_your_data_grow

33. A Day in Data - Raconteur. (2021). Raconteur - Content for Business Decision-Makers [online]. Available from: https://www.raconteur.net/infographics/a-day-in-data/

34. Khvoynitskaya, S. (2021). The future of big data: 5 predictions from experts for 2020-2025. Available from: https://www.itransition.com/blog/the-future-of-big-data

35. Lippel, H. (2022) Big Data in the Media and Entertainment Sectors. ResearchGate | Find and share research Available from:

https://www.researchgate.net/publication/299617099_Big_Data_in_the_Media_and_Entertainment_Sectors

36. Reuters Institute for the Study of Journalism | Reuters Institute for the Study of Journalism. (2022). [online]. Available from: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-04/Big%20Data%20For%20Media_0.pdf

37. Kobielus, J. (2013). The Four V's of Big Data, IBM Big Data & Analytics Hub [online]. Available from: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

38. BAIN & COMPANY. (2013). Travis Pearson and Rasmus Wegener, Available from: http://www.bain.com/Images/BAIN_BRIEF_Big_Data_The_organizational_challenge.pdf

39. Big Data in Media and Entertainment. (2020). Qubole. Available from: https://www.qubole.com/big-data-in media-and-entertainment/

40. Subroto, A. & Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning | Journal of Big Data | Full Text. Journal of Big Data | Home page [online]. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1

41. Jose, A. (2020). Big Data for Executives and Market Professionals – eBook. Available from: https://joseantonio.xnewdata.com/publish

42. STATISTA Online. (2022). Number of Netflix paid subscribers worldwide [online]. Available from: https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers worldwide/#:~:text=Netflix%20had%20195.15%20million%20paid,Netflix's%20total%20global%20subscriber %20base.

43. How Netflix uses big data to create content and enhance user experience. Homepage - ClickZ. (2022). [online]. Available from: https://www.clickz.com/how-netflix-uses-big-data content/228201/

44. Van Dijk, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology [online].Available from: Google Scholar.
45. Uthayasankar, S. (2017). Critical analysis of Big Data challenges and analytical methods - ScienceDirect. ScienceDirect.com | Science, health and medical journals, full text articles and books. [online]. Available from: https://www.sciencedirect.com/science/article/pii/S014829631630488X
46. Chen, J., Chen, Y.Du. X. Li, C. Lu. J. Zhao. S. & Zhou. X. (2013). Big data challenge: a data management perspective. Frontiers of Computer Science.
47. Gandomi. A., & Haider. M. (2015). Bevondthe hvpe: Big data concepts, methods, and analytics. International Journal of Information Management.
48. Edwards. R., & Fenwick. T. (2015). Digital analytics in professional work and learning. Studies in Continuing Education.
49. Phillips-Wren, G., & Hoskisson, A., (2015) An analytical journey towards big data Journal of Decision Systems

ATTACHMENTS

Table 1

Description of Study Population and Low Back Pain by Sex and Age Group, Demographic Characteristics, Socioeconomic Status and Other Factors, Weighted Percent

Source: Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey.

	All Workers		Male Workers		Female Workers		Younger Workers (18–40)		Older Workers (41–64)	
Variable	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**
Low back pain		25.7	53.0	24.5	47.0	27.13	50.0	23.8	50.0	27.7
Demographics										
Age										
18 – 25	15.3	21.2	14.9	18.3	15.9	24.2				
26 – 40	34.6	25.0	36.1	24.3	33.0	25.9				
41 – 55	36.7	27.6	35.9	26.3	37.6	28.9				
56 – 64	13.4	28.0	13.2	27.4	13.6	28.7				
Sex										
Male	53.1	24.5	53.0	24.5			54.1	22.5	52.0	26.6
Female	46.9	27.1			47.0	27.1	45.9	25.3	48.0	28.8
Ethnicity and race										
Non-Hispanic White	67.8	27.1	67.6	26.6	68.0	27.8	63.7	25.7	71.9	28.4
Non-Hispanic Black	10.9	23.2	9.4	18.5	12.4	27.2	11.6	20.8	10.1	25.9
Hispanic	14.7	24.5	16.7	22.5	12.4	27.5	17.9	21.8	11.5	28.7
Non-Hispanic Asian	4.9	15.9	4.7	14.1	5.2	17.8	5.0	13.3	4.9	18.5
Non-Hispanic Others	1.8	26.0	1.6	24.9	2.0	27.1	1.9	24.9	1.6	27.4
Education										
Less than high school	12.2	29.2	14.0	26.1	10.2	34.1	13.3	26.8	11.2	32.1
High school	22.0	25.7	22.8	25.1	21.2	26.6	20.6	22.6	23.5	28.5
Some college	31.8	29.2	29.6	27.3	34.3	31.1	34.5	26.9	29.1	32.0
College	22.0	21.6	22.1	21.4	22.0	21.8	22.2	20.5	21.8	22.7
Master and above	12.0	21.1	11.6	20.9	12.4	21.3	9.5	19.4	14.4	22.2
Earning										
< $14,999	19.0	28.8	14.5	26.3	24.1	30.4	24.1	26.6	13.9	32.5
$15,000 – $34,999	30.5	27.1	27.3	26.0	34.2	28.1	34.3	24.8	26.7	30.0
$35,000 – $64,999	30.3	25.7	32.0	26.2	28.3	25.0	28.1	24.6	32.4	26.6
> =$65,000	20.2	24.8	26.2	24.1	13.5	26.4	13.5	21.9	27.0	26.2
Other related factors										
Leisure-time physical activity	51.7	24.4	56.0	23.7	46.9	25.4	55.2	24.1	48.2	24.8
Serious Psychological Distress	2.5	49.3	2.2	44.2	2.8	53.7	2.3	45.6	2.7	52.3
Obesity	26.9	30.1	27.9	27.7	25.7	33.0	23.3	27.2	30.5	32.3

Table 2

Description of Study Population and Low Back Pain by Sex and Age Group, Work-Related Factors, Weighted Percent

Source: Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey.

	All Workers		Male Workers		Female Workers		Younger Workers (18–40)		Older Workers (41–64)	
Variable	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Low back pain**	**% in Pop**	**% Lowe back pain**
Workplace psychological factors										
Work-family imbalance	16.8	32.3	16.2	29.6	17.5	35.1	17.0	30.2	16.6	34.4
Exposure to hostile work environment	7.6	35.7	6.3	33.0	9.0	37.9	7.4	28.2	7.8	35.3
Job insecurity	32.4	30.2	33.7	28.6	30.9	32.1	30.5	36.2	34.3	32.0
Work organization characteristics										
Non-standard work arrangement	28.2	15.8	17.7	27.4	13.6	29.4	14.2	23.1	17.3	32.4
Alternative shifts	27.1	24.4	28.2	21.5	25.8	28.0	33.0	22.5	21.2	27.4
Work hours per week										
8 to 39	28.4	27.4	20.3	25.4	37.6	28.7	32.7	25.4	24.2	30.2
40 hours	43.5	23.3	44.5	22.5	42.3	24.2	41.8	20.8	45.1	25.6
41 to 45	6.8	28.5	7.9	25.6	5.6	33.1	6.7	25.2	7.0	31.6
46 to 59	13.3	26.9	16.8	26.2	9.2	28.3	11.8	26.2	14.7	27.5
60 hours and over	8.0	28.2	10.4	27.0	5.3	30.7	6.9	28.1	9.1	28.3
Occupation										
Management	9.5	24.8	11.4	23.8	7.3	26.5	7.4	25.8	11.6	24.1
Business and financial operations	4.7	21.5	3.9	17.5	5.5	24.7	4.3	19.1	5.0	23.7
Computer and mathematical	3.1	19.4	4.3	19.4	1.7	19.1	3.5	17.8	2.7	21.5
Architecture and engineering	2.1	20.1	3.2	19.6	0.7	22.4	1.8	15.2	2.4	23.8
Life, physical, and social science	1.2	22.9	1.1	23.3	1.3	22.6	1.0	23.8	1.3	22.2
Community and social services	1.8	26.3	1.3	29.5	2.3	24.2	1.9	22.9	1.7	30.0
Legal occupations	1.3	25.7	1.3	20.7	1.4	30.9	1.3	26.4	1.3	25.0
Education, training and library	6.5	23.1	3.6	24.7	9.8	22.4	6.2	21.2	6.9	24.8
Arts, design, entertainment, sports and media	1.9	26.2	2.1	24.9	1.8	28.0	2.0	27.2	1.9	25.2
Healthcare practitioners and technical	5.2	26.3	2.4	28.1	8.4	25.7	4.8	22.2	5.6	29.8
Healthcare support	2.6	33.6	0.6	25.9	4.9	34.7	2.7	31.2	2.5	36.2
Protective service	2.1	23.5	3.1	23.8	0.9	22.2	2.1	23.9	2.1	23.0
Food preparation and serving related	5.7	27.2	4.9	20.6	6.6	32.6	8.0	25.2	3.4	32.0
Building and grounds cleaning and maintenance	4.0	25.3	4.5	22.7	3.4	29.1	3.9	20.9	4.1	29.5
Personal care and service	3.5	28.5	1.4	24.4	5.9	29.6	3.9	23.6	3.2	34.4
Sales and related	10.5	26.9	10.2	25.7	10.7	28.2	11.9	25.6	9.0	28.6
Office and administrative support	13.2	25.1	6.5	21.9	20.7	26.2	13.1	23.2	13.2	27.0
Farming, fishing, and forestry	0.6	28.0	0.9	21.0	0.3	49.5	0.7	15.9	0.5	46.5
Construction and extraction	5.2	31.8	9.7	31.8	0.2	30.8	5.2	29.3	5.2	34.2
Installation, maintenance and repair	3.7	27.9	6.8	28.1	0.3	24.0	3.4	25.7	4.1	29.8

Table 3

Logistic Regression of Work-Related Factors for Low Back Pain, Sex and Age Stratified Analysis

Source: Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey.

	All Workers (Model A)		Male Workers (Model B)		Female Workers (Model C)		Younger Workers (18–40) (Model D)		Older Workers (41–64) (Model E)	
Factors	**OR**	**95%Conf**	**OR**	**95%Conf**	**OR**	**95%Conf**	**OR**	**95%Conf**	**OR**	**95%Conf**
Workplace psychological factors										
Work family imbalance	1.27*	1.15–1.41	1.30*	1.10–1.54	1.49*	1.27–1.75	1.39*	1.17–1.66	1.42*	1.21–1.66
Exposure to hostile work environment	1.39*	1.25–1.55	1.41*	1.10–1.80	1.47*	1.20–1.80	1.58*	1.27–1.97	1.29*	1.04–1.60
Job insecurity	1.44*	1.24–1.67	1.28*	1.11–1.47	1.26*	1.1–1.44	1.33*	1.15–1.55	1.23*	1.08–1.40
Work organization characteristics										
Non-standard work arrangement	1.09	0.95–1.25	1.10	0.92–1.32	1.08	0.88–1.33	0.92	0.76–1.11	1.27*	1.06–1.52
Alternative shifts	0.81*	0.73–0.90	0.74*	0.64–0.86	0.90	0.76–1.07	0.77*	0.66–0.90	0.85*	0.72–0.99
Work hours per week										
40 hours	1.00		1.00		1.00		1.00		1.00	
8 to 39	1.11	0.97–1.26	1.09	0.90–1.32	1.14	0.97–1.34	1.16	0.96–1.41	1.06	0.9–1.25
41 to 45	1.25*	1.03–1.52	1.09	0.82–1.44	1.52*	1.16–1.98	1.20	0.92–1.58	1.30	0.98–1.72
46 to 59	1.15	0.99–1.34	1.15	0.95–1.39	1.12	0.88–1.42	1.25	0.99–1.56	1.08	0.89–1.32
60 hours and over	1.19	0.99–1.43	1.16	0.90–1.49	1.27	0.96–1.67	1.38*	1.07–1.78	1.07	0.83–1.37
Occupation										
Computer and mathematical	1.00		1.00		1.00		1.00		1.00	
Management	1.13	0.82–1.55	1.05	0.73–1.51	1.33	0.78–2.25	1.41	0.94–2.1	0.95	0.60–1.50
Business and financial operations	0.98	0.68–1.43	0.78	0.46–1.33	1.30	0.73–2.33	0.99	0.59–1.67	0.95	0.58–1.56
Architecture and engineering	0.94	0.62–1.42	0.86	0.53–1.41	1.33	0.58–3.05	0.81	0.41–1.63	0.96	0.53–1.71
Life- physical- and social science	1.25	0.75–2.07	1.18	0.60–2.32	1.43	0.68–3.04	1.47	0.70–3.08	1.06	0.54–2.09
Community and social services	1.35	0.86–2.12	1.74	0.89–3.43	1.29	0.69–2.41	1.26	0.64–2.46	1.42	0.74–2.75
Legal occupations	1.23	0.75–2.02	0.98	0.46–2.10	1.70	0.83–3.52	1.42	0.72–2.79	1.04	0.50–2.15
Education- training and library	1.10	0.78–1.53	1.24	0.74–2.08	1.22	0.73–2.05	1.07	0.66–1.73	1.11	0.68–1.81
Arts- design- entertainment- sports and media	1.29	0.84–2.00	1.22	0.69–2.15	1.53	0.75–3.12	1.50	0.86–2.62	1.09	0.57–2.07
Healthcare practitioners and technical	1.30	0.92–1.83	1.71*	1.01–2.88	1.35	0.80–2.29	1.14	0.72–1.82	1.38	0.84–2.26
Healthcare support	1.51*	1.05–2.18	1.26	0.49–3.20	1.71*	1.02–2.86	1.60*	1.00–2.56	1.44	0.81–2.57
Protective service	1.01	0.67–1.52	1.06	0.65–1.72	0.93	0.42–2.07	1.23	0.68–2.19	0.82	0.47–1.43
Food preparation and serving related	1.31	0.94–1.84	1.10	0.67–1.81	1.63	0.99–2.7	1.39	0.89–2.16	1.23	0.74–2.04
Building and grounds cleaning and maintenance	1.15	0.79–1.68	1.13	0.69–1.85	1.30	0.74–2.29	1.07	0.65–1.77	1.21	0.73–2.01
Personal care and service	1.27	0.85–1.89	1.20	0.55–2.59	1.44	0.83–2.5	1.13	0.65–1.96	1.43	0.85–2.41
Sales and related	1.25	0.90–1.73	1.23	0.81–1.86	1.34	0.82–2.2	1.37	0.91–2.05	1.11	0.7–1.77
Office and administrative support	1.04	0.76–1.42	0.99	0.63–1.56	1.21	0.75–1.94	1.10	0.73–1.65	0.98	0.63–1.53
Farming- fishing- and forestry	1.25	0.69–2.25	0.86	0.38–1.92	2.66*	1.05–6.79	0.80	0.33–1.96	2.11	0.82–5.42
Construction and extraction	1.41*	1.00–2.01	1.31	0.88–1.96	1.61	0.37–6.98	1.56	0.98–2.49	1.27	0.77–2.1
Installation- maintenance and repair	1.21	0.83–1.78	1.15	0.74–1.78	1.15	0.38–3.46	1.31	0.79–2.17	1.13	0.67–1.9
Production	1.20	0.86–1.69	1.06	0.69–1.63	1.61	0.94–2.77	1.21	0.76–1.94	1.17	0.72–1.9